The world inside your head

From structuring to representations to language

A.E.A. Goosen

Radboud University Nijmegen

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

THESIS IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN ARTIFICIAL INTELLIGENCE

The world inside your head

From structuring to representations to language

Author:
A.E.A. Goosen (s0322253)
t.goosen@student.ru.nl

Supervisors:
Dr. W.F.G. Haselager
Dr. I.G. Sprinkhuizen-Kuyper

November 27, 2008

Abstract

Much of both human and animal behavior can be understood in terms of reactive processes. However, some aspects of behavior seem to go beyond reactiveness, as they appear to involve *internal representations* of the world: internal states that stand in for aspects of the world, such that they can guide an agent's behavior even in situations in which the agent is completely decoupled from the corresponding aspects. Chandrasekharan and Stewart (2004, 2007) argue that a special kind of *epistemic structuring*, active adaptation of some structure for cognitive benefit, can generat internal traces of the world with a representational character. Their model would account for both epistemic environment adaptation and the creation of these internal traces through a single reactive mechanism. Although Chandrasekharan and Stewart demonstrate the workings of this mechanism through a set of experiments, the claim for a representational nature of the resulting internal structures has not been validated empirically.

This thesis aims to further investigate this claim on empirical and theoretical grounds. Two subsequent experiments were carried out to validate two respective hypotheses; the first experiment was designed to test whether internal epistemic structuring can facilitate the forming and use of internal *presentations*, a non-decoupled, hence weaker kind of internal states than representations; in the second experiment, an embodied, embedded agent simulation was carried out to investigate the relation between representational demand and the development of epistemic structuring capacities. The experiments provide evidence that epistemic structuring can be used to form, maintain and use both internal presentations representations. Taking into account these results, it is discussed how the epistemic structuring model might account for the nature and origin of internal (re)presentation, and how it relates to the *extended mind thesis*. Finally, the model is placed in the context of language evolution; it is speculated to play an explanatory role with respect to the nature, origin and cognitive role of language.

Acknowledgements

A lot of the writing, and most of the research that led to this thesis was done at the AI department of the Radboud University in Nijmegen, a place that have over the years has begun to feel like home. There are many people, there and elsewhere, that I would like to thank for their help and support during the course of completing this thesis. First of all my supervisors, Pim Haselager and Ida Sprinkhuizen-Kuyper, who patiently accepted and reviewed the many proposals and drafts that preceded this final version, provided me with helpful comments and suggestions, and helped me find structure in the occasional fuzzy intuition. Then there are numerous people who were so very kind to not only show interest in my work in progress, but also took the effort to think along, ask stimulating questions and provide useful comments: Joris Janssen, who reviewed almost an entire semi-finished draft, leading to many improvements in structure and content; Eelke Spaak, Iris van Rooij, Franc Grootjen, and Louis Vuurpijl whose remarks at various stages improved my insight in my own work; Annemarie Melisse, Charlotte Munnik, Jaap Robben, Jelmer Wolterink, Louis Dijkstra, Tom Schut, Wilmar van Beusekom, and many other people who made the entire process significantly more enjoyable but also motivated me to stay on track and keep going; also the people in the TK from whom I frequently stole precious computational resources, but who somehow kept appreciating my presence nevertheless. Finally, huge thanks go to my parents, Louis and Annelies, to Lijsje and to Brenda, who more than once were faced with my thesis-related mood changes and wandering thoughts, but kept supporting and encouraging me throughout.

Contents

1 Introduction **1**

1.1 Representations 2

1.1.1 Intelligence without (internal) representations? 4

1.2 Epistemic structuring 5

1.3 Investigations in this thesis 7

2 Epistemic structuring **9**

2.1 Experiments of Chandrasekharan and Stewart 10

2.1.1 Q-learning 10

2.1.2 Connectionist implementation 11

2.1.3 Environment structuring experiment 13

2.1.4 Internal structuring experiment 15

2.2 Epistemic structuring and representations 19

2.2.1 Presentations versus representations 22

3 A reversal learning experiment **25**

3.1 Introduction 25

3.1.1 Reversal learning 26

3.2 The model 27

3.3 Experiment 30

3.3.1 Agents 30

3.3.2 Procedure 30

3.4 Results 32

3.4.1 ANOVA results 34

3.5 Conclusions 36

3.6 Further analyses 38
3.6.1 Behavior analysis 38
3.6.2 Internal dynamics analysis 41
3.7 Discussion 51

4 A situated agent simulation experiment 53
4.1 Introduction 53
4.2 The simulation 54
4.2.1 Environment 54
4.2.2 Task 56
4.2.3 The model 57
4.3 Experiment 58
4.3.1 Agents 58
4.3.2 Procedure 58
4.4 Results 60
4.4.1 ANOVA results 63
4.5 Conclusions 63
4.6 Discussion 64

5 Conclusions and discussion 67
5.1 The forming of internal presentations 67
5.1.1 Internal structuring and internal presentations 67
5.2 From presentations to representations 68
5.3 Reactive and representational processing 70
5.4 External and internal structuring 70
5.5 Summary and conclusions 71

6 Afterword: from labels to language 73
6.1 Introduction 73
6.2 Epistemic structuring and language 73
6.2.1 Similarities between language and epistemic structuring . 74
6.2.2 Differences between language and epistemic structuring . 77
6.2.3 Language: communication or representation? 78
6.3 From Epistemic structuring to language 79

6.3.1 Beyond reactiveness 80
6.3.2 The internalization of epistemic structuring 81
6.3.3 From structuring to protolanguage 82
6.4 Language, epistemic structuring, and cognition 83
6.5 Conclusions 84

References **87**

A Additional figures for the reversal learning analyses **91**
A.1 Action selection 91
A.2 Q-learning network weights 94
A.3 Internal environment network weights 96
A.4 Q-learning activation patterns 98

B Additional figures for the agent simulation analysis **101**
B.1 Boxplots 101

List of Figures

1.1 The road sign problem. 4

2.1 Overview of the architectures of the experiments of C&S 16

3.1 Schematic overview of the internal structuring model, as used in the reversal learning experiment 28
3.2 Stimulus sets of the reversal learning experiment. 31
3.3 Performance of the agents in the reversal learning experiment . . . 33
3.4 Scatter plots of the scores of the second and third rounds of the reversal learning experiment. 37
3.5 Action plots for the reversal learning agents 39
3.6 Performance of reversal learning agents 40
3.7 Weight development of the Q-learning network in reversal learning agents 43
3.8 Weight development of the IE network in reversal learning agents 45
3.9 Cross-correlations of class and activation of the units of the Q-learning network of an agent with 0 hiddens and 6 outputs 47
3.10 Cross-correlations of class and activation of the units of the Q-learning network of an agent with 6 hiddens and 6 outputs 48
3.11 Cross-correlations of class and activation of the units of the IE network of an agent with 6 hiddens and 6 outputs 50

4.1 The environment of the multi agent simulation 55
4.2 Schematic overview of the internal structuring model, as used in the agent simulation 56

4.3 Contour plots of performance in the agent simulation experiment . 61
4.4 Boxplots of performance in the agent simulation experiment . . . 62

6.1 Proposed schema of the origin of language 80

A.1 Action plot of an IE with 0 hidden units and 6 outputs 92
A.2 Action plot of an IE with 2 hidden units and 6 outputs 92
A.3 Action plot of an IE with 6 hidden units and 6 outputs 93
A.4 Action plot of an IE with 12 hidden units and 6 outputs 93
A.5 Q-weights in an agent with 0 IE hiddens and 6 IE outputs 94
A.6 Q-weights in an agent with 2 IE hiddens and 6 IE outputs 94
A.7 Q-weights in an agent with 6 IE hiddens and 6 IE outputs 95
A.8 Q-weights in an agent with 12 IE hiddens and 6 IE outputs 95
A.9 IE-weights in an agent with 1 IE hidden and 6 IE outputs 96
A.10 IE-weights in an agent with 2 IE hiddens and 6 IE outputs 96
A.11 IE-weights in an agent with 6 IE hiddens and 6 IE outputs 97
A.12 IE-weights in an agent with 12 IE hiddens and 6 IE outputs 97
A.13 Q-learning activation patterns in an agent with 0 IE hiddens and 6 IE outputs 99
A.14 Q-learning activation patterns in an agent with 6 IE hiddens and 6 IE outputs 100

B.1 Boxplots of performance in the conditions with 1 target 102
B.2 Boxplots of performance in the conditions with 2 targets 103
B.3 Boxplots of performance in the conditions with 3 targets 104

Chapter 1

Introduction

Human behavior is deeply rooted in its evolutionary heritage. However, many of us may not feel to share much of our cognitive abilities with animals that have been around much longer than the couple of hundreds of thousands years that we have. In our daily lives, we are constantly interpreting the world around us, reasoning about the many things inside it, planning ahead, engaging in social interaction and so on. Simpler animals like insects, fish or even other mammals, such as rodents, on the other hand seem to act in a mostly reactive manner, responding directly to stimuli with little internal processing. Indeed, reactiveness as a basis for natural behavior has since long been acknowledged (Balkenius, 1995), both in traditional psychology (cf. Lewin, 1936) and in cognitive modelling (e.g. Braitenberg, 1984; Arkin, 1990; Brooks, 1991).

However, it does not seem feasible to explain the entire spectrum of the cognitive abilities of humans in terms of reactive processing. Reasoning about both concrete and abstract concepts, making predictions, planning ahead and engaging in conversations are just a few of the many possible activities that appear to rely on complex inner processing rather than resulting from one-way stimulus-response couplings. In contrast to purely reactive behavior, such advanced cognition often is hard – if not impossible – to understand without assuming *internal representations*; reasoning about something in the absence of that something requires the manipulation of something that stands in for it. While language, whether formal or natural, provides symbols that can fulfill this role of standing in – the word

'couch' refers to the thing at home you may plan to spend the evening on – the way the human mind deals with such issues is far from clear. However, evolution does dictate that human cognition has, through a gradual process, arisen from more simple systems – ultimately from the earliest reactive creatures that had no more than a rudimental stimulus-response system. Hence, cognitive science faces two questions concerning internal representations: What is their nature? and What is their origin? The approach taken in this thesis is to take the little we know with respect to the latter issue – cognition is rooted in reactive behavior – to add to the, arguably, even fewer knowledge that currently exists concerning the former. Before proceeding however, it is essential to get a clearer view of what representations are considered to be, and of some of the issues concerning them.

1.1 Representations

A natural first step in introducing any concept is providing a definition. As is often the case, a multitude of interpretations of representations are available, and agreement among them is lower than desired. One interpretation however does appear to be quite popular (cf. Clark, 1997; Haselager, Bongers, & van Rooij, 2003), presumably because of its clarity and broadness, but it is also fairly agnostic to the *nature* of representations. This interpretation is the one by Haugeland (1991); here is how it is cited by Haselager et al. (2003):

> A sophisticated system (organism) designed (evolved) to maximize some end (e.g., survival) must in general adjust its behavior to specific features, structures, or configurations of its environment in ways that could not have been fully prearranged in its design. [...] But if the relevant features are not always present (detectable), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*; that which it stands for is its content; and its standing in for that content is *representing* it. (Haugeland, 1991, p. 62)

The something-standing-in-for-something part of this interpretation is intuitively essential to the concept of representation. Interesting is that Haugeland places this standing-in into a context of meaningful *behavior*. One way to interpret this is that representations necessarily *underlie* meaningful cognitive behavior, which pretty much complies with the view of traditional AI. Another interpretation would take representations as *supporting* such behavior, but not necessarily forming the basis for it.

Given a system displaying meaningful cognitive behavior, how can we find out whether it has internal representations? In the case of symbolic AI programs, this is easy. We can inspect their workings at a very detailed level and still find linguistic or semi-linguistic constituents like variables or propositions with explicit semantic reference. Systems less tailored, such as animals or adaptive sub-symbolic artificial control mechanisms like neural networks, provide a different case. Especially in cases where such systems operate in an embedded, embodied context and where their behavior is the result of complex interactions between brain, body and environment, looking for individual content-bearing units will prove fruitless. Concluding, the absence of representation in any sense might then be attractive, but it neglects the strong suggestion provided by both introspection and empirical findings (e.g. Shepard & Metzler, 1971) that some cognizers (humans, and probably other) do form, keep and use representations. It also leaves a large explanatory gap with respect to behavior that seems to require reasoning and planning.

Ways to assess the presence of representation in systems of the hard-to-analyze kind have been suggested. For example, Clark and Grush (1999) define "minimal robust representationalism", for which the following criteria are provided:

1. representations would be inner states whose adaptive functional role is to stand in for extra-neural states;
2. the states with representational roles should be precisely identifiable;
3. the representations should enhance real-time action. (Chandrasekharan & Stewart, 2007, p. 343)

These criteria, although themselves of course open to debate, provide a quite concrete schema to test a non-symbolic cognitive system against. Near the end of this

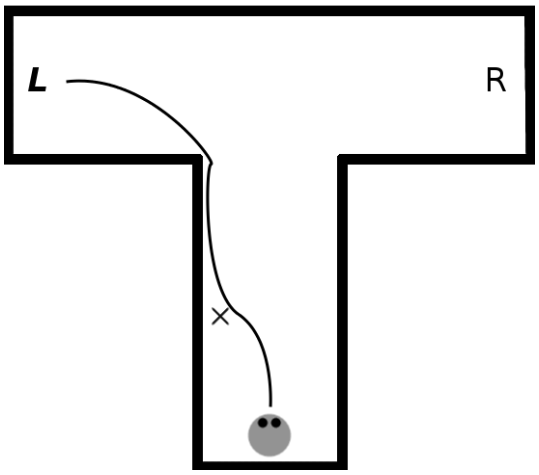


Figure 1.1: The road sign problem. The stimulus (cross) is placed at the left side, which indicates that the reactive robot should go left at the junction. It manages to do so by moving towards the stimulus and following the wall. Adapted from Rylatt and Czarnecki (2000).

chapter, in Section 2.2, the application of these criteria to a concrete model, and subsequently a number of contrasting views on representations will be discussed.

1.1.1 Intelligence without (internal) representations?

Discontent with the traditional view of representation has been growing since the mid-1980's and the rise of *situated cognition* (Ziemke, Bergfeldt, Buason, Susi, & Svensson, 2004). While Brooks (1991) famously argued for intelligence *without representations*, more subtly formulated suggestions have been made and backed with experimental results. An agent may 'outsource' its representational needs to its body, environment or distribute it over all of these. An example can be found in solving the so called 'road sign problem' (Rylatt & Czarnecki, 2000), a *delayed response task* that requires a robot to decide whether to take the left or right branch of a T-maze on basis of the position of a visual stimulus presented earlier. The stimulus is placed to the left or to the right, indicating that the robot should take the corresponding branch. An illustration is given in Figure 1.1. Thieme and Ziemke (2002) showed that a purely reactive solution exists to this problem. Reactive agents evolved a strategy in which they moved to the side where the

stimulus was placed and followed the wall to stay on that side and move towards the junction, at which point taking the shortest angle results in 'choosing' the correct direction. The robot in this case can be considered to effectively use its position with respect to the wall rather than some internal structure as a memory or representation (Ziemke et al., 2004).

The general remark that can be made is that a lot of behavior seemingly or expectedly incorporating internal representation can in fact be brought about reactively. Clever usage of external structures, whether pre-existent or established by the agent itself, often leads to fast and reliable solutions that do not rely on internal capacities that are expensive to use in terms of energy or may not even be available. By its opportunistic nature (e.g. Ayala, n.d.), evolution is likely to prefer, given similar profit and reliability, reactive task solutions to those involving internal processing, which are likely to be slower, more vulnerable and more expensive in terms of energy than their reactive counterparts. Although there is no doubt that human behavior is beyond reactiveness, evolutionary heritage should not be neglected. Hence, from an evolutionary perspective there is much value in the adage put forward by Haselager et al. (2003): "Don't use representations in explanation and modeling unless it is absolutely necessary." While this is a healthy advise, it does not provide any clues on how to deal with representations in those cases where we do not know how to avoid them, if they can be avoided at all. In this thesis, a model will be discussed that provides an explanation for representational processing in agents that however retain their reactive mode, hence accounting for the mixture of reactive and representational processing that is found in natural cognitive agents. A short introduction to this model will be presented below. Chapter 2 will be dedicated to an in-depth discussion of its workings, backgrounds and implications.

1.2 Epistemic structuring

Chandrasekharan and Stewart (2004) introduced the concept of *epistemic structuring* and provided the basis for the model that will be central to the present investigations. This section provides a short introduction; Chapter 2 is dedicated to a more in-depth discussion.

Epistemic structuring as a concept is rooted mostly in the work of David Kirsh, who posed that agents can, and in fact do, gain cognitive leverage by adhering to the motto "changing the world instead of oneself" (Kirsh, 1996). In Kirsh's model, agents add structure to their environment through *task-external actions*: actions that themselves are not in the repertoire of actions required to physically complete a task. Kirsh (1994) makes the distinction between *pragmatic* and *epistemic* action. Pragmatic actions are those that are "performed to bring one physically close to a goal", while epistemic actions are "actions performed to uncover information that is hidden or hard to compute mentally". Chandrasekharan and Stewart (C&S) apply this dichotomoy to physical structures that agents can generate in their environment. Epistemic structures are those structures that reduce *cognitive* complexity in the context of a task. By making use of task-external, epistemic actions, task-relevant paths through state space can get shortened, lowering cognitive complexity (Chandrasekharan & Stewart, 2004).

C&S pose the question how such epistemic structures might be generated. They suggest that systematic epistemic structuring by agents can emerge in the context of a task and two biologically plausible conditions:

1. agents create random structures, which do not necessarily serve an epistemic goal, and

2. agents get tired, can track their tiredness and tend to reduce it.

Once every while randomly generated structures (condition 1) have the unforeseen effect of reducing an agent's effort (condition 2) required to execute its task. If this happens, the agent will associate that structuring behavior with reduced tiredness and hence adopt this structuring pattern into its behavioral strategy. By doing so, the agent has, in the vocabulary of Kirsh, shortened a path in its state space and is likely to discover that by following and reinforcing this path it can achieve a systematic tiredness reduction. In a multi-agent scenario, collective structures can emerge that are both used and reinforced by all members of the population.

In two respective experiments, C&S (2004, 2007) investigate two modes of epistemic structuring: external and internal. The notion of *external structuring* refers to an agent's physical structuring of its environment to make it more cognitively hospitable, whereas *internal structuring* denotes applying the epistemic

structuring mechanism to a structurable internal module. C&S (2007) claim that internal structuring provides a means for agents to engage in representational processing. Thus far however, too few empirical findings have been reported to gain sufficient insight into the workings of epistemic structuring, or to come to any fundamental conclusions about their relation to internal representations. It is the aim of this thesis to provide further empirical investigation as well as theoretical embedding.

1.3 Investigations in this thesis

In the following chapters, C&S's (2007) claim of representational processing through epistemic structuring will be further validated. First, Grush's (1997) distinction between presentations and representations (to be discussed in Section 2.2.1) will be employed to, incrementally, empirically investigate the representational capabilities of the epistemic structuring model. This is done in two subsequent simulation experiments aimed at testing the following respective hypotheses:

1a. Epistemic structuring applied to internal environments provides an agent with the ability to form *internal presentations* and use these to guide its behavior.
This hypothesis will be tested through a simulation that provides a non-embodied, non-embedded context, and a task in which presentational processing, rather than the use of direct sensor-motor couplings, leads to increased performance, but that demands no guidance by counterfactual (non-perceivable) information.

1b. The same mechanism can be used to keep counterfactual variants of these presentations: *internal representations*.
This second hypothesis builds on the first in that presentations can be considered a more basic kind of internal states on which representations rely. It is tested in a multi-agent simulation with an embodied, embedded context that provides a more realistic scenario and places representational demands on the agents.

A second goal of this thesis is to gain insight into the dynamics of the processes that underlie the model's hypothesized capacities. Therefore, two kinds of qualitative analyses were carried out on the results of the first of the above mentioned experiments:

2a. A *behavior analysis*, in order to determine in what way internal and external actions make up the pattern of internal (re)presentation through epistemic structuring.

2b. An *internal dynamics analysis*, to explore the processes that may underlie internal (re)presentation and the interactions between the components of the model.

A final goal is to provide further theoretical context for the epistemic structuring model. The final chapters of this thesis speculate about the explanatory power of epistemic structuring from two, potentially related, perspectives:

3a. *Embedded, embodied cognition*: can epistemic structuring account for a broad range of cognitive phenomena even though it is firmly rooted in reactive behavior and environment interaction?
Representational capacities are associated with high level cognition. Yet, epistemic structuring has reactiveness in its core, and therefore potentially has a great explanatory scope, bridging the gap between reactiveness and higher cognition.

3b. *Language evolution*: does epistemic structuring provide a viable starting point for a theory of language evolution?
Epistemic structuring will be shown to share properties with language, and be compatible with established views of the nature and cognitive role of language. An evolutionary development of language rooted in epistemic structuring will be sketched.

Prior to the presentation of these investigations, comes an in-depth discussion of the epistemic structuring model of C&S (2004, 2007), the mechanisms that underlie it, and its hypothesized relation to presentations and representations.

Chapter 2

Epistemic structuring

In the previous chapter, epistemic structuring was briefly introduced. In this Chapter it will be more closely examined by means of a description of the work of C&S (2004, 2007), who introduced the concept and provided a model for epistemic structuring in reactive agents.

Epistemic structuring is the generation and reinforcement of *epistemic structures*, defined by C&S (2007) as "stable organism-generated (...) structures that lower cognitive load" (p. 330). C&S initially require these structures to be external to the agent (i.e., exist in the environment), but subsequently propose an *internal* modality of epistemic structuring. Hence, a distinction can be made between two modes of epistemic structuring: external, by means of adapting the environment, and internal, through a special kind of epistemic actions[1] that cause restructuring of an internal module. C&S claim that by the latter process "internal traces of the world could originate in reactive agents within lifetime" (p. 330); these traces are argued to have a representational character.

The following sections describe the experiments of C&S (2004, 2007) and the composition and workings of their model. After a description of the mechanism underlying both external and internal structuring, both modes are discussed subsequently. The proposed relation between internal structuring and representations is discussed at the end of this chapter.

[1] the distinction between pragmatic and epistemic actions is discussed in Section 1.2

2.1 The experiments of Chandrasekharan and Stewart

In the experiments of C&S (2004, 2007), embodied, embedded agents and their environment are simulated. The agents possess a set of relatively high level, but strictly local sensors and a number of task-specific and task-external actions, one of which is selected at each time step. A mapping between these input states and actions is learned by the agents on basis of feedback, by means of a reinforcement learning mechanism that constitutes the control structure of the agents. Before proceeding with a more detailed description of the agent simulation, an introduction of the control structure will be presented, as it plays a central role in understanding the dynamics of epistemic structuring as introduced by C&S.

2.1.1 Q-learning

C&S (2004, 2007) chose to base the control structure for their agents on a reinforcement learning algorithm called Q-learning (Watkins, 1989). One of the motivations for using Q-learning is that it can, in a straightforward fashion, model a creature's tendency to avoid unfruitful effort and thus unnecessary tiredness. Q-learning maps input states to actions and adjusts this mapping on the basis of quantitative feedback it receives as a consequence of selecting a specific action. In the model of C&S, this feedback consists of 'tiredness feedbacks' of -1 at each time step and a reward of $+10$ upon completing a trip.

Rummery and Niranjan (1994) give an apprehensive description of the workings of Q-learning. The mechanism revolves around the *Q-function*, which defines an estimated goodness of an action in the context of a given input state, and is learned on basis of rewards. In its most simple form, the learning of the Q-function takes place by the following update rule after an action a_t has been selected given an input $\mathbf{x}_t$:

$$Q(\mathbf{x}_t, a_t) \leftarrow r_t + \gamma V(\mathbf{x}_{t+1}) \tag{2.1}$$

where r_t is the feedback received after choosing the action, γ is a fixed *discount*

factor and $V(\mathbf{x}_t)$ is the *value function*, which gives a prediction of the feedback for the given input state. As the algorithm in principle selects the actions that yields the highest expected result, the function can be written as:

$$Q(\mathbf{x}_t, a_t) \leftarrow r_t + \gamma \max_{a \in A} Q(\mathbf{x}_{t+1}, a) \tag{2.2}$$

From this update rule it follows that not only the immediate feedback value guides the learning of the Q-function, but also subsequent feedback has its influence, the strength of which is governed through the γ parameter.

An additional parameter ε determines the chance that instead of the action likely to optimize the reward based on current information, a random action is chosen. This leads to a form of *exploration*, allowing the agent to find out (and learn about) the implications of certain actions in certain contexts and to deal with potentially dynamic aspects of the environment.

2.1.2 Connectionist implementation

In Q-learning, a mapping between input states and actions goodness is learned. A straightforward way of storing this mapping is by listing separately the Q values for all combinations of input states and actions in some sort of lookup-table. This approach is used in the experiments (2004, 2007) of C&S. There are several downsides to this approach however. In practice such lookup-tables easily become enormous depending on the number of states and actions one would like to be able to discern. In somewhat complex situations, accessing and updating the Q values may involve unaffordable computational overhead. Apart from this practical objection, the lookup-table approach requires explicit discretization of input states. Two downsides of this are the arbitrariness of the level of discretization and the inherent inability for the algorithm to generalize over input states. To exemplify the latter issue, suppose a system that has learned to associate input values $1 \ldots 49$ and $51 \ldots 99$ with action A and input values $100 \ldots 199$ with action B. Now if it encounters the unseen before input value of 50, would it not be desirable for the system to select action A? A lookup table however does not establish such behavior as all associations are independent.

A solution to these problems (Rummery & Niranjan, 1994) is to use feed-forward neural networks (FFNN's) to *approximate* the lookup-table based Q-function. FFNN's are known to be capable of classifying over continuous inputs and to scale well to large input and state-spaces (Rummery & Niranjan, 1994, p. 5). Rummery and Niranjan and Kuzmin (2002) describe and compare several methods implementing connectionst Q-learning. For the experiments that will described further on, an implementation (QCON: Kapusta, 2008) of connectionst Q-learning based on the findings by Kuzmin (2002) was used.

The variant of connectionist Q-learning used in this framework, and hence in my experiments (Chapters 3 and 4), is called Modified Connectionist Q-learning (MCQ-L). It will be described here shortly. For a detailed account of this algorithm and several variants, I refer to Rummery and Niranjan (1994).

Action selection is straightforward. The inputs are set according to the current input state of the agent and the network is activated. There are as many output units as there are possible actions, and the activation value of each output unit is interpreted as the estimated goodness of the corresponding action. Once an action is selected and feedback is received, this needs to be reflected in the Q-function.

Rummery and Niranjan (1994) describe how the network can be trained, i.e. how the weights can be adjusted, using an on-line version of temporal difference learning (TD-learning), which builds on the work of Watkins (1989) and Sutton (1989). This kind of learning depends on the storage of a so called *eligibility traces* $\mathbf{e}$ for each weight of the network. It keeps track of preceding error gradients and gets updated at each time step. This update happens on basis of the *error gradient* that is provided by the backpropagation algorithm for the current state of the network and the output activation set such that the selected action is activated (e.g. a positive activation of 1 with the the other actions at 0). This error gradient is added to the previous eligibility trace, which is discounted by a factor λ:

$$\mathbf{e}_t = \nabla_w Q_t + \gamma\lambda\,\mathbf{e}_{t-1} \tag{2.3}$$

(Rummery & Niranjan, 1994) where $\nabla_w Q_t$ is the error gradient and γ is Q-learning discount factor mentioned earlier. When $\mathbf{e}$ has been updated, the action gets executed and in the next time step an action is selected on basis of the input and the

current state of the network. Then the just calculated eligibility traces are used to update the network:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \alpha(r_{t-1} + \gamma Q_t - Q_{t-1})\mathbf{e}_{t-1} \tag{2.4}$$

(Rummery & Niranjan, 1994).

As can be seen, the *difference* between the successive Q-values [2] is used, and no future Q-values need to be consulted. Hence, this algorithm constitutes on-line temporal difference Q-learning.

2.1.3 Environment structuring experiment

In the first simulation of C&S (2004, 2007), 10 agents were placed in a 30×30 grid world containing two 3×3 patches designated a home location and target location. The agents are considered to be successful to the degree that they manage to move back and forth between the home and target locations within a limited time frame. This can be thought of as a foraging task, in which the agents gather food from a single source and bring it home unit-by-unit.

Agents

The architecture of the agents of this experiment is shown schematically in Figure 2.1(a). The agents are controlled through the reinforcement learning mechanism Q-learning, the workings of which were described in detail in Section 2.1.1. [3] Recall that it selects one (unparameterized) action out of a fixed set at each time step, the selection being driven by a goodness estimation based on the current input and feedback values it receives after the execution of each action.

Actions C&S provided their agents with five possible actions: moving into a random direction, moving into a 'home like' direction, moving into a 'target-like' direction and finally dropping two kinds of pheromones (two separate actions). The two kinds of pheromones are 'home-like' and 'target-like' respectively, akin

[2] Q_t is the Q-value associated with the selected action, short for $Q(\mathbf{x}_t, a_t)$

[3] Chandrasekharan and Stewart also carried out their experiment with a genetic algorithm *instead* of Q-learning, but I will not discuss that here as it is of little interest for the current purposes.

to pheromone systems found in ants (C&S, 2004). This means that execution of the 'move towards home-like' actions brings an agent to the home zone if it is within reach or otherwise moves it into the direction with the highest level of 'home' pheromone. Dropping pheromones of either kind increases the amount of pheromone on the agent's current location. This amount is subject to decay (its level decreases over time) and dispersion (a cell receives small amounts of pheromones from its neighboring cells). The levels of home pheromones $PH_{c,t}$ and target pheromones $PT_{c,t}$ on cell c at time step t are given by:

$$PH_{c,t} = e\left(PH_{c,t-1} + d\left(\frac{1}{8}\sum_{a=1}^{8} PH_{s_{c,a},t-1} - PH_{c,t-1}\right)\right) \tag{2.5a}$$

$$PT_{c,t} = e\left(PT_{c,t-1} + d\left(\frac{1}{8}\sum_{a=1}^{8} PT_{s_{c,a},t-1} - PT_{c,t-1}\right)\right) \tag{2.5b}$$

with e being the evaporation rate, set to .99, d the dispersion rate, set to .04 and $s_{c,a}$ the ath of the 8 cells surrounding cell c. Initially, values of PH and PT are set at 0 for all cells of the environment.

Perception The sensory capacities of the agents are few but very high-level. There are four input values to the control system: a binary value that tells whether the agent has visited the target zone ('is carrying food'), two more values that represent the amount of home-like pheromones and target-like pheromones at the current location respectively, and a final value that represents the time that has passed since the last time the agent dropped pheromones.

Adaptation The feedback schema that drives the Q-learning algorithm was as follows: a penalty of -1 is given for each executed action (i.e. at each time step as exactly one action has to be chosen), and $+10$ for completing a 'trip' which is defined as visiting the home location after having visited the target location at least once since the previous trip (or since the beginning of the experiment in case of the first trip). Notice that all actions are equally expensive and if an agent chooses a structuring action, in can be considered to do so *instead* of a movement. This makes the structuring actions 'task-external' in the terminology of Kirsh (1996).

According to C&S (2004),

> The best way to envisage this is to think of an action that a creature might do which inadvertently modifies its environment in some way. Examples include standing in one spot and perspiring, or urinating, or rubbing up against a tree. These are all actions which modify the environment in ways that might have some future effect, but do not provide any sort of immediate reward for the agent. (p. 3)

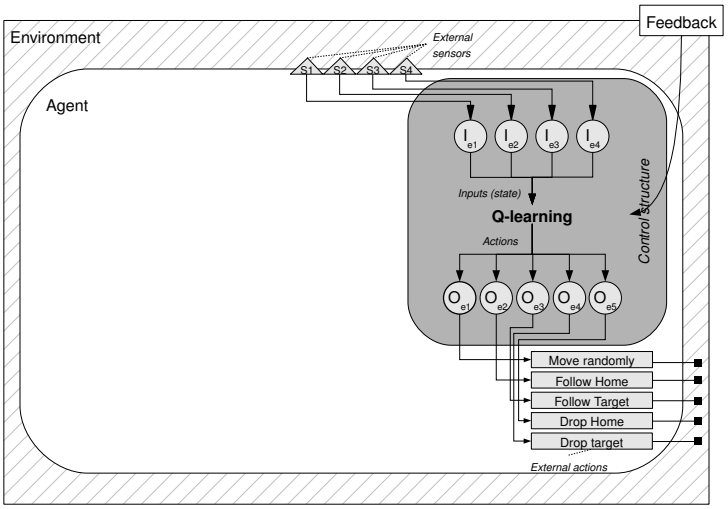
Experimental results

C&S (2004, 2007) report that the agents in the experiment as described above learn to improve their performance by enhancing the environment through environment structuring. A comparison to an alternate condition of the experiment with identical configuration except for the absence of structuring actions (drop pheromones) showed that the agents were still able to improve their performance slightly over time but not as much as in the condition with structuring. Unfortunately, C&S did not investigate whether a significant effect of the ability to use structure actions was present. However they did do a comparative behavior analysis which showed that agents with structure generation spend 58% of their time generating structures (and therefore over half of their time *not* moving). Agents without structure generation showed a higher fraction of random movement to directed movement than agents with structure generation.

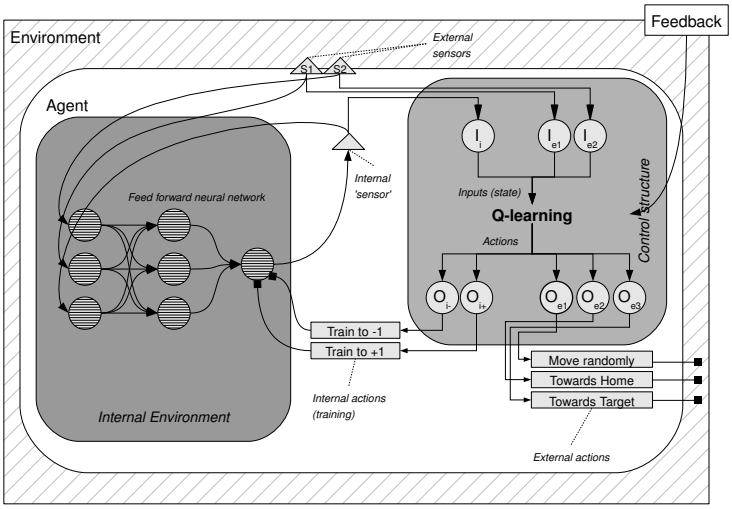
With their experiment, C&S (2004, 2007) have shown that reactive agents (i.e. non-symbolic, non-planning agents with no internal data storage or recursion) can learn, during lifetime, to utilize task-external actions to modify their task environment and increase its "cognitive congeniality" (Kirsh, 1996). These findings formed the basis for an extension of their model, which lifts these structuring, yet still principally reactive agents to a higher cognitive level.

2.1.4 Internal structuring experiment

Chandrasekharan and Stewart's (2004, 2007) external structuring experiment showed that agents can learn how to add structures to their external surroundings that



(a) Experiment 1: Environment structuring



(b) Experiment 2: Internal tracing

Figure 2.1: An overview of the architectures of the experiments of Chandrasekharan and Stewart (2004, 2007). The first schema shows an agent with no internal environment, capable of dropping and following pheromones as used in the first (environment structuring) experiment. The second schema shows an agent with an internal environment, as used in the second (internal tracing) experiment. Lines with arrows are connections, a black square indicates the object of an action. Triangles are sensors, circles are units, either of the IE neural network (horizontal stripes) or of the Q-learning mechanism (solid gray).

shorten paths in their state space and thus lower the burden on their internal resources. In a follow-up of their 2004 paper, C&S (2007) remark:

> This within-lifetime learning model raises an interesting question: can similar within lifetime learning lead to the generation of novel structures in the agents mind, rather than in the agents environment? This seems to be both a natural extension of our work on external structures, and, more importantly, a novel way to model the origin of internal representations in rudimentary agents within their lifetime. If an agent can learn this strategy of generating internal structures to lower tiredness, then it can choose to remember particular things in particular ways to benefit it in the long term, just as our earlier experiments showed that it was possible to choose to drop pheromones in useful ways. (p. 338)

The insight that the environment structuring framework could be extended to make possible a kind of *internal structuring* set the stage for a new simulation experiment. The context of the experiment was adopted from the prior simulation (see previous section), with a few changes made to allow the agent to engage in internal structuring rather than external structuring. The task again was a foraging task, that required the agent to move back and forth between a home location and a target location as often as possible within limited time. The feedback scheme of a $+10$ reward for finishing a 'trip' and -1 penalties for every action was kept, and so was the Q-learning based control structure.

A number of things were altered, both in the setup of the experiment and in the agent's control structure, as can be seen in Figure 2.1(b). First of all, a single agent was used instead of multiple agents. As the agents cannot directly sense each other and in this case there is no indirect sensing or influencing through external structures, it makes no difference how many agents operate simultaneously. Second, a new sets of input variables and possible actions were introduced. The pheromone related actions and sensors were taken away. The only external sensors the agents now had were a 'home detector' and a 'target detector' which read 0 if the agent is not on the specified location, and 1 if it is. Three physical actions were at the agent's disposal: moving random, moving into the direction of the

home location, and moving into the direction of the target location. The latter actions always and reliably move the agent towards the respective locations. At first glance this appears to turn the foraging task into one highly trivial, but notice that the agent's challenge has shifted radically: it can no longer sense whether it has visited the target location, as it could in the earlier simulation described above. It now somehow has to keep track of where it is going, as it cannot rely on any stimuli to determine its heading.

The internal environment

To enable the agent to determine its heading, C&S (2007) provided the agent with what they call an *internal environment*. Formed by a multilayered feedforward neural network (FFNN), trainable through backpropagation (Rumelhart & McClelland, 1986), this internal environment provides the agent with a target for epistemic structuring actions, in that respect replacing the *physical environment* of the environment structuring experiment. The FFNN has as many input units as there are inputs to the Q-learning mechanism, and a single input.

Structuring of the internal environment is defined by C&S as training the FFNN with the current input state as an input pattern and one of two output activations (-1 and $+1$) as a target. This adds two actions to the action set of the control mechanism, one for each target output. One execution of the training action causes ten successive training cycles to be executed. This means that ten times in a row, the network gets activated, an error score at the output unit gets calculated and the weights of the network get adjusted according to this error. The training actions are considered *epistemic structuring actions* by Chandrasekharan and Stewart (2007), in the same sense as tagging physical objects or marking a path in the environment.

At each time step, the FFNN is activated according to the current input and the acquired weight setting. Because of the recurrent connection from the output unit to the input layer of the network, the network has a certain dynamics: activating the network successively with the same external input can lead to different output values due to a changing recurrent input value. To deal with this, the network gets

activated 100 times in a row. [4] The eventually resulting output activation is used as an input to the Q-learning algorithm, in addition to the external sensors. As the entire input array is shared between the FFNN and the control structure, the FFNN also recursively receives its own output activation as an input value.

Experimental results

To inspect the effectiveness of internal structuring, C&S (2007) compared the performance of agents with internal structuring, as defined above, to that of agents without any structuring mechanism. They found that the agents with internal structuring outperform those without, and additionally that this advantage increases somewhat along with the distance between the target and home locations. Only in the highly simplistic situation where home and target are located directly next to each other, the ability to form internal epistemic structures decreases performance.

According to the authors, this shows that the same mechanism that allows agents to learn how to add epistemic structures to the world, can account for the creation of *internal traces* of the world, which they argue the resulting internal structures are. The theoretical consequences of this claim are discussed in the following section.

2.2 Epistemic structuring and representations

Assuming that the internal structures generated by the agents of the experiments of C&S (2007) can indeed be considered traces of the world, the question arises whether these traces are used by the agents to *represent* the world. To answer that question, C&S turn to the criteria for 'minimal robust representationalism' provided by Clark and Grush (1999): "(i) representations would be inner states whose adaptive functional role is to stand in for extra-neural states; (ii) the states with representational roles should be precisely identifiable; (iii) the representa-

[4] Although this aspect is not explicitly mentioned or motivated by C&S (2007), it is a feature of their simulation (Stewart, 2006). It has a clear rationale: after about 100 updates, the output activation can be expected to have converged unless it oscillates – in which case further updating makes no sense.

tions should enhance real-time action" (C&S, 2007, p. 343). According to C&S, by these criteria it is justified to consider the internal traces *proto-representations*: they are inner states that stand in for something specific in the world, and are useful because of their aboutness. "However," C&S remark, "these internal traces are not full-bodied representations, (...) because our agents do not use the internal traces as surrogates to model the world when the actual structures do no exist in the world." Two additional reasons for not considering internal traces as "full-bodied representations" are mentioned: internal traces cannot be fully decoupled from ongoing environmental input, and the selectiveness of the agents' representation of the world, it being "highly constrained by the biological niches within which the organisms evolved" (p. 343).

Rebuttals to these objections come from C&S themselves, but arguments for discarding some of the requirements as needlessly strong can be drawn from other sources (e.g. Clark, 1997). To begin, C&S (2007) put in contrast to the classic notion of representations as static structures what they call the *distributed origin* thesis of representation. This thesis describes the forming of representations as a result of

> an incremental process based on feedback of cognitive load [in which initially random elements] gradually become systematically stored and acquire a representational nature. Such an internal representation is not a single well-defined structure that reflects the world mirror-like, but a systematic coagulation of contexts and associated actions, spread over a network... Metaphorically, such an internal representation resembles the core of an active bee swarm, rather than static symbolic entities, such as words or pictures. (p. 344)

In this model, reference relations between structures internal to the agents and elements of the environment emerge if they lower cognitive load. In contrast, traditional systems of representation, whether of symbolic or distributed nature, typically bear *a priori* reference relations. Admittedly, such models provide much more insight into the role and nature of the representations, but, as C&S remark, fail to explain why representations arise. The importance of the role that representation play in a system is stressed by Clark (1997):

> The status of an inner state as a representation (...) depends not so much on its detailed nature (...) as on the role it plays within the system... What counts is that it is *supposed to* carry a certain type of information and that its role relative to other inner systems and relative to the production of behavior is precisely to bear such information. (p. 146)

Subsequently, Clark sketches a continuum of representational possibilities from "mere causal correlations" to "Haugeland's creatures that can deploy the inner codes in the total absence of their target environmental features." Between these extremes lies a range of cases Clark terms *adaptive hookup*. As a very simple example of such a hookup, a sunflower directing itself towards the sun and light-seeking robots are mentioned. A level at which speaking of representations starts to make sense, according to Clark, is reached "when we confront inner states that (...) exhibit a systematic kind of coordination with a whole space of environmental contingencies." (p. 147)

Considering an agent's representational capacities as a property that can be defined within a continuum, rather than being an all-or-nothing issue makes sense from an evolutionary perspective. It seems reasonable to assume that more complex ways of dealing with the dynamics of the world build on the simple ones. There probably are some important qualitative differences between the systems at the lower ends of Clark's continuum and the more complex adaptive hooks to take into account – for example, turning towards light can be done through feed-forward processing, while accessing internal models requires some degree of recursiveness. For the most, however, differences can likely be accounted for in terms of gradual improvements.

Coming back to epistemic structuring, C&S's (2007) model of internal traces seems a strikingly fit candidate for providing adaptive hookup, covering a large portion of the just described continuum. To recapture, the model is compatible with a view of representations as internal structures that gradually emerge as a result of interaction with the environment. Therefore, in the context of agents of a complexity comparable to that of insects given their task and environment, these representations will be context-sensitive (i.e. not decoupled from environmental input) and action-oriented, rather than objective and action-independent (Clark,

1997). Consequently, internal traces faced by C&S are not in the upper range of the representational continuum. However, there is no reason to rule out (or, as yet, to assume) that epistemic structuring, in principle, has the potential of providing representations of the kind closer to the interpretations of Haugeland (1991) or Clark and Grush (1999). Further investigations will have to show the extent of epistemic structuring.

2.2.1 Presentations versus representations

Before commencing such investigations, a final distinction has to be introduced. Grush (1997) clears up some of the fog traditionally surrounding the topic of representations by making quite a clean cut between representations and what he terms *presentations*:

> ...what distinguishes presentations from representations is the use they are put to. A presentation is used to provide information about some other, probably external in some sense, state of affairs. It can be used in this way because the presentation is typically causally or informationally linked to the target in some way. The representation's use is quite different: it is used as a counterfactual presentation. (p. 5)

To put it shortly: presentations are about the actual perceived state of affairs, while representations can be used to stand in for things not available to the senses.

As indicated by the final sentence of the above quotation, a hierarchical relation can be outlined: representations are like presentations, but do not depend on the environmental state. Additionally, and departing somewhat from Grush's elaborations, presentations should be distinguished from mere sensations. Consider a Braitenberg vehicle. Few would oppose to attributing sensory abilities to such a vehicle. However, there appears to be quite a difference between its direct sensor-motor coupling and what goes on inside a creature that might follow a strict 'out of sight is out of mind' schema, but however does seems to have some degree of understanding of what it perceives. Take for example a dog recognizing its owner out of a group of people. Assuming Molly does not think of her loving owner when he is not around, but does recognize him from a broad range of view-

points, and regardless of the cloths he his wearing today[5], her behavior is neither fully reactive, nor the result of representational processing. The internal state that does cause her specific reaction, a state an external observer would label OWNER, should be considered an *internal presentation*.

Representations are like presentations, and can guide behavior in a similar way, yet are counterfactual with respect to the state of the world from an agent's perspective. To apply this to the dog's presentation OWNER and a potential representation of this owner: the latter can be used to miss *him* – rather than just long for the sound of him opening a can, or simply his smell – or imagine what he might be doing while he is not around. This distinction will be used in the following chapters to incrementally examine the representational nature of the internal traces resulting from epistemic structuring. In a first experiment (Chapter 3, it will be investigated whether an internal environment can establish internal representations. The next step (Chapter 4) is to apply the model in a situations that require the use of representations, or counterfactual presentations.

[5]and not by his smell either, for the sake of simplicity

Chapter 3

A reversal learning experiment

3.1 Introduction

Chandrasekharan and Stewart (2007) argue that their model of internal epistemic structure (for a description, see Chapter 2) allows an agent to engage in representational processing. Although the internal traces that underlie this processing are described carefully as *proto-representations* – as opposed to complete, context independent substitutes of the world that representations often are taken to be – it is quite a bold statement to make and thus demands substantial empirical backing. In this chapter and the one following it, two experiments will be presented as an attempt to provide evidence for the claims of C&S.

In the experiment described in this chapter, a reversal learning (RL) experiment was simulated with agents of variously configured internal environments as subjects. The goal of the experiment was to discover whether agents, in a clear, neither embodied nor embedded setting, can learn to make use of their internal environment to improve their performance on the task. If more substantial internal environments would lead to increased performance, this would, due to the nature of the RL task, provide evidence for the agents' ability to actively form internal presentations Besides a performance analysis, detailed inspection of the agents' behavior and internal dynamics may provide additional insight into the workings of the internal environment and whether it affords (re)presentational processing.

3.1.1 Reversal learning

RL is an experimental paradigm that has been used to investigate *conceptualization*[1] in animals (Hurford, 2007). The paradigm requires a subject to choose between two stimuli, from separate classes, one of which leads the subject to be rewarded, while the other does not. The subject has to learn this relation, starting out with no knowledge about which stimulus should be associated with a reward. The essential aspect of the paradigm is that after a number of trials, or after a predetermined level of success has been achieved, the stimulus-reward relation gets *reversed*. So, after this reversal, the stimulus previously associated with a reward leads to no reward and vice versa. To keep being rewarded after the reversal, the subject will have to somehow unlearn the relation it just mastered and teach itself the opposite pattern.

Reversal learning and internal presentation

How can a reversal learning experiment show whether a subject has internal presentations rather than a fully reactive mode underlying its behavior? Recall from Section 2.2.1 that internal presentations, at least as interpreted here, are internal states that arise from, but go beyond sensory input. Fully reactive systems have a direct mapping between sensory input and output. The state of this mapping do not constitute internal presentations; for an internal state to be considered an internal presentation, it has to be a potential object of manipulation *itself*. This notion can be illustrated as follows: a vehicle wired, Braitenberg style, to be attracted by light sources has no internal presentation of the light source it is moving towards. In contrast, a human instructed to approach a lamp turned on at the other side of the room will be guided by the perception of the *lamp*, not by the sensation of the light it emits. The presentation itself might be based mainly on this sensation, but it seems at least awkward to skip the intermediate level of internal presentation.

An important advantage of internal presentations is that they allow for generalization over sensory states. Learning to recognize, say, chairs, essentially comes

[1] Hurford (2007) uses the term *concept* rather loosely, it seemingly covering both what we have called presentations and representations. The RL experiment however appears to require little *re*presentation in the sense used here.

down to defining one's internal presentation (at an abstract level) of a chair. Once properly defined, one perceives a 'chair' rather than 'an arrangement of horizontal and vertical surfaces supported by four rather thin columns'. If one then encounters some hard to identify object, and subsequently is informed that it is actually some new kind of chair (hooray for modern design!), one can somehow retune the mechanism that delivers the presentation CHAIR. A creature without the ability to form internal presentations can of course learn for all kinds of objects that they afford sitting, but will lack a general notion that unifies the set. It would for example be rather hard to explain to this creature the game of 'musical chairs'[2] unless perhaps all chairs are of exactly the same type.

This should also make clear the relation between reversal learning and internal presentations: agents capable of forming and using presentations are capable of applying an internal reversion to an entire class of sensory states through an operation on their presentation, rather than having to completely rewire their input-output mapping. This difference, as pointed out by Hurford (2007, p. 25), can be framed, in the terminology of Deacon (1997), as learning an *indexical* connection (one in each condition) versus learning a *symbolic* connection to a 'previously acquired inner representation'. Negating one's internal presentation is symbolic in the sense that an operation (or 'computation') is executed on an entity, specifically negation on a presentation. This could be expressed in a symbolic fashion, for example: STIMULUS S → REWARD becomes STIMULUS S → NOT(REWARD).

3.2 The model

The model, viz. the agents' control structure including internal environment, is an extension of that of the original internal structuring model (C&S, 2007), described in Chapter 2, and depicted schematically in Figure 2.1(b). An overview of the extended model is given in Figure 3.1 on page 28. Like the model of C&S, it consists of two modules: a Q-learning control structure (CS), and a feed forward neural network that functions as the internal environment (IE).

The CS is based on the QCON platform (Kapusta, 2008), a connectionst im-

[2] Stoelendans

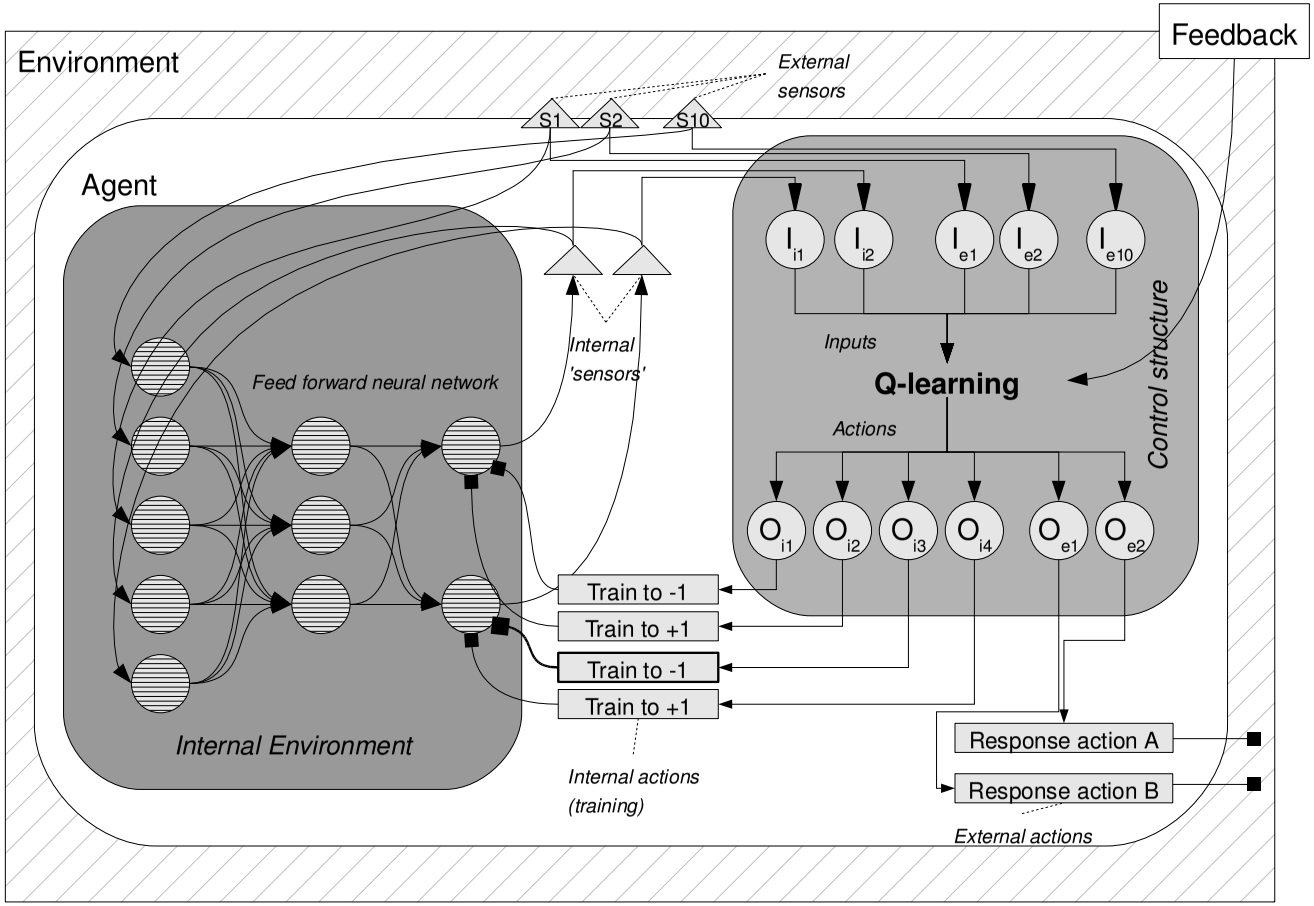


Figure 3.1: A schematic overview of the internal structuring (IS) model as used in both the reversal learning experiment (Chapter 3) and the multi agent simulation (Chapter 4). The model is based on the one used by Chandrasekharan and Stewart (2007). This figure shows inputs and actions for the reversal learning experiment. Lines with arrows are connections, a black square indicates the object of an action. Triangles are sensors, circles are units, either of the IE neural network (horizontal stripes) or of the Q-learning mechanism (solid gray).

plementation of the Q-learning algorithm (see Section 2.1.2). Sensory values are fed into its input layer, which directs connectly to an output layer with one unit per possible action. The input pattern is a combination of values from the external sensors, which in the reversal learning experiment reflect the current stimulus, and values coming from the IE. The actions can also be grouped in two categories: external and internal actions. In this case, there are two external actions that correspond to the two possible responses in the reversal learning task (which will be explained below). The internal actions cause a restructuring of the network by means of backpropagation, as is the case in the original model. However, the extension differs to the extent that the IE can have multiple output units – and thus, multiple *pairs* of training actions: per IE output, one that targets it to $+1$, and one that targets it to -1. The IE poses no difference from the IE of the original model, other than the number of output units of its network being variable. It takes the same input pattern as the CS and propagates its real-valued output activations to this same, shared input array.

The dynamics of the entire system (i.e. information flow, learning of the Q-function, action selection, training of the IE) are as described in Chapter 2, so I will not cover all of that here. One aspect that is different from the model of Chandrasekharan and Stewart (2007) and needs some explanation is the training of the network with respect to the variable number of outputs in the IE. For every IE output there are two training actions available to the CS: one that sets a positive target value and one that sets a negative value. Upon execution of one of these actions, the network gets trained through standard backpropagation (Rumelhart & McClelland, 1986) with the current input state as an input pattern and *ignoring* the activations of the output units not corresponding to the selected action. That is, all errors at the output layer are considered to be 0, except for the unit associated with the action chosen by the CS.

In the experiment, the agents are subjected to two simple stimuli consisting of three boolean (either -1 or $+1$) values each. They can externally respond to these stimuli through two respond actions. Except for these and additional training actions, there are no actions to chose from. Details about the agents' perception and action are given below.

3.3 Experiment

3.3.1 Agents

The subjects in this experiment are simulated agents, instantiations of the above described model. The experiment was run with a range of agent configurations. Each configuration is defined by the number of hidden units (0,1,2,3,6 or 12) and the number of output units (0,1,2,3,6 or 12) in the network of its IE. Agents with no output units effectively have no IE, as there are no input units to the Q-learning that come from the IE, nor are there any training units. Agents with no hidden units, but with one or more output units do have an internal environment, although it is irresponsive as there is no coupling between its input layer and its output layer. However, the number of actions and inputs to the control structure is dependent on the number of outputs of the IE. These conditions were included because it cannot be ruled out on forehand that these dimensions have an effect on an agent's performance.

The Q-learning based control structure was equally configured for all types of agents: $\alpha = 0.2$, $\gamma = 0.3$, $\lambda = 0.3$ and $\varepsilon = .1$. The learning parameter of the neural network of the IE was set fixed to $\eta = .2$, and no momentum (see Rumelhart & McClelland, 1986) was used. These values were chosen such that an agent of either type is capable of learning the first (pre-reversal) round of trials effectively.

The neural network of the Q-learning control structure was a feed forward neural network with no hidden layer. This means that it has one matrix of weights: those between the input values and the action units (see Section 2.1.2).

3.3.2 Procedure

A run of the experiment consists of four consecutive rounds, each of which contains 10,000 trials. During a trial two five bit stimuli are presented to the agent as a ten bit input vector. The two stimuli are selected randomly, one out of each of two exclusive stimulus sets, which can be seen in Figure 3.2. The two selected stimuli are concatenated in random order; either the five bits of the stimulus out of the first set are shown before those of the stimulus out of the second set or the other way around. The entire string of 10 bits is fed to the agent (both IE and CS),

+	+	-	+	+	+	+	+	+	+
-	+	-	-	+	-	+	+	-	+
-	-	-	-	-	-	-	+	-	-
+	-	-	-	+	+	-	+	-	+

(a) Stimulus set 1 (b) Stimulus set 2

Figure 3.2: The two stimulus sets of the reversal learning experiment.

with $-$ stimulus bits as an input value of -1, and $+$ stimulus bits as $+1$. As an example, a selection of the stimuli from the bottom row of the respective sets will thus lead to an external input of $[+1\ -1\ -1\ -1\ +1\ +1\ -1\ +1\ -1\ +1]$.

There are two external actions, which can be thought of as buttons corresponding to the respective stimuli. If an agent selects the first action, it chooses the first stimulus, gets feedback, and moves on to the next trial. The second action similarly corresponds with the selection of the second stimulus.

The feedback may be a reward, in which case a score of $+10$ is given to the Q-learning algorithm, or it can be a penalty, which is given through a negative feedback score of -10. For each executed response or training action, a feedback score of -1 is given to the agent, thus introducing a form of *effort penalty* as not to make internal structuring a 'free' operation, leading to 'chicken' behavior and getting stuck in local optima. The execution of a training action does not end the present trial.

As an example, the course of a trial could be as follows: stimuli are presented in the order [SET 2 SET 1] – that is, the first five inputs are taken from a random stimulus from set number two, and the remaining from the set number one. Assume that in this round, stimulus set 1 is associated with the reward. The agent might first select a training action, leading to backpropagation of the network of the IE with the output unit corresponding to the selected action targeted to the value associated with the action, either -1 or $+1$. The Q-learning mechanism receives a feedback of -1 for the execution of an action. Then the agent might select one or more training actions, and eventually choose to execute one of the response actions. If it selects response action A, it will receive an negative feedback of $-1 + -10 = -11$ for having selected an action and responding to the stimulus not associated with the reward. On the other hand, selecting response action B

will lead to receiving a feedback of $-1+10=+9$, as positive feedback is given for selecting the action associated with the reward. During the trial, the agent's external input remains equal. After it has executed a response action, a new trial is begun and thus new stimuli are selected and presented until the end of the trial.

At the end of a round (i.e. after each 10,000 trials), the stimulus-feedback relation is reversed, so that the stimulus leading to positive feedback in the one round, leads to negative feedback in the following round and vice versa. Each run consists of four such rounds, resulting in a total of 40,000 trials per run. At the beginning of each run, the agent has a randomly initialized control structure and internal structuring mechanism, so an agent's learned associations do not get passed from one run to the next.

The scores of all agents were recorded as a vector of binary values, one for each trial, where a 0 represents an incorrect response and a 1 represents a correct response. When averaging over the score vectors of all runs within a condition, a vector of average performance results from which slopes over specified domains can be calculated. So, for an agent a there is a score vector:

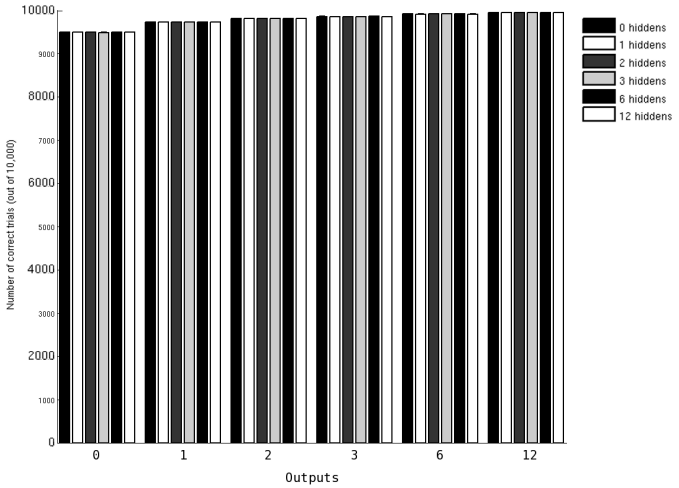
$$\mathbf{s}_{a,r} = [s_{a,r,1} \ \ldots \ s_{a,r,T}] \tag{3.1}$$

for each run r, consisting of $T = 40,000$ trials.

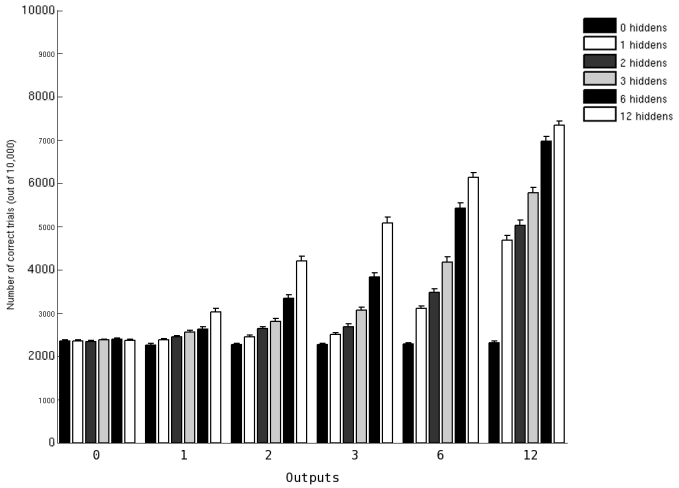
The experiment was run a hundred times for each condition. Thus, 100 observations were obtained for each combination of Outputs $\in \{0,1,2,3,6,12\}$ and Hiddens $\in \{0,1,2,3,6,12\}$.

3.4 Results

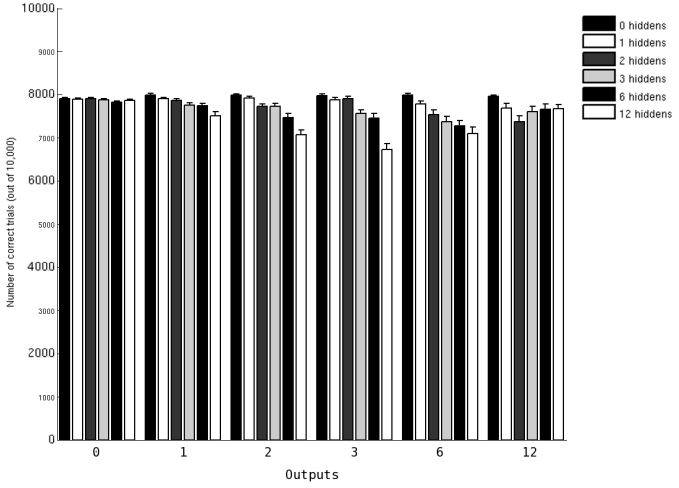
The effect of the number of hidden and output units on reversal learning performance was assessed by means of an analysis of variance (ANOVA). The numbers of hidden and output units served as between-subject factors: Hiddens (six levels: 0,1,2,3,6,12) and Outputs (0,1,2,3,6,12). Separate analyses were carried out for each of the four rounds of the experiment. The results are depicted in Figure 3.3. These graphs clearly suggest different effects between the rounds. In the first round (Figure 3.3(a)), all conditions show roughly equal, high levels of



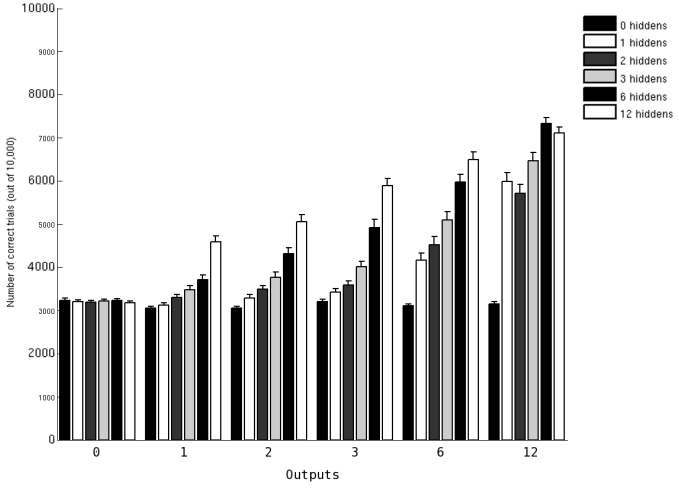
(a) Round one



(b) Round two



(c) Round three



(d) Round four

Figure 3.3: Performance of the agents in the reversal learning experiment

performance, which indicates that all conditions are fit to learn a sound stimulus-response pattern. In the following rounds, effects of the numbers of hidden and output units are visible in the graphs. The second and fourth rounds shows a general pattern of a rise in performance along with both the number of hidden units and the number of output units. The effects in the third round seem a bit less straightforward – the overall pattern shows a decrease with the number of hidden units, but in the higher levels of Outputs, this pattern starts to break down. After an overview of the results of the statistical analysis, these results will be examined more closely and conclusions will be drawn.

3.4.1 ANOVA results

For the first round, a significant main effect of Outputs was found ($F(5,3564) = 84362$, $p < .001$, $\eta_p^2 = .992$)[3], but not of Hiddens ($F(5,3564) = 322.7$, $p = .140$). A significant, but very weak interaction effect Hiddens $\times$ Outputs was with $F(25,3564) = 377.1$, $p = .003$ and $\eta_p^2 = .013$. To further investigate the effect of Outputs, pairwise comparisons with Bonferroni correction were carried out, which showed that the scores for each level of Outputs were significantly lower (all $p < .001$) than the scores of higher levels.

Analysis of the second round yielded significant main effects of both Outputs ($F(5,3564) = 1314$, $p < .001$, $\eta_p^2 = .648$) and Hiddens ($F(5,3564) = 771.2$, $p < .001$, $\eta_p^2 = .520$) as well as a significant Hiddens $\times$ Outputs interaction effect ($F(25,3564) = 92.392$, $p < .001$, $\eta_p^2 = .393$). Pairwise comparisons (Bonferroni corrected) showed that for all levels of Outputs, scores at lower levels were significantly lower (all $p \leq .001$) than scores at higher levels. The same relation was found for Hiddens: agents with fewer hidden units gained significantly lower scores (all $p < .001$). In Figure 3.3(b), the interaction effect can be seen clearly: agents with no output units or no hidden units perform equally well at each level of respectively Hiddens and Outputs, whereas agents with one or more output units show an increase of performance along with the increase of the number of hidden units. Additional ANOVA's over this round with one of both factors fixed at 0

[3] η_p^2 denotes *partial eta squared*, a measure of effect size, and is not related to the earlier mentioned neural network parameter η.

and the other as the only independent variable confirmed this. Within the group of 0 output units, no significant effect of Hiddens was found ($F(5,594) = .470$, $p = .799$). Within the group of 0 hidden units, no significant effect of Outputs was found ($F(5,594) = 1.347$, $p = .243$).

In the third round, significant effects of Outputs ($F(5,3564) = 16.036$, $p < .001$, $\eta_p^2 = .022$), Hiddens ($F(5,3564) = 43.816$, $p < .001$, $\eta_p^2 = .058$) and Hiddens $\times$ Output ($F(25,3564) = 5.957$, $p < .001$, $\eta_p^2 = .040$) were found, too. However, Bonferroni corrected pairwise comparisons showed a more complex pattern than in the preceding rounds. To start with Outputs, 0 outputs yields significantly higher scores than 2,3,6 and 12 outputs (all $p < .001$). Agents with 1 output perform better than agents with 2 ($p = .039$), 3 or 6 outputs (both $p < .001$). Configurations with 2 outputs perform better than 6 outputs ($p = .048$), and 6 outputs yield lesser scores than 12 outputs ($p = .025$). The general pattern is that agents with fewer output units perform better than or equal to agents with more output units, with the exception of 6 outputs, which performs worse than 12. For Hiddens, 0 performs better than 2, 3, 6 and 12 hidden units (all $p < .001$); 1 has higher scores than 3, 6 and 12 (all $p \leq .001$); 2 performs better than 6 ($p = .035$) and 12 ($p < .001$); and 6 performs better than 12 ($p < .001$). The interaction effect is quite clear from Figure 3.3(c). Again, no outputs shows no difference over Hiddens, in contrast to 2, 3, 6 or 12 outputs. The former three of those show a monotonic decrease over Hiddens, whereas the latter has its depth in the middle.

Round four, finally, shows results rather similar to those of the second round. Main effects of Outputs ($F(5,3564) = 390.9$, $p < .001$, $\eta_p^2 = .354$), Hiddens ($F(5,3564) = 245.6$, $p < .001$, $\eta_p^2 = .256$) and a significant interaction effect Hiddens $\times$ Output ($F(25,3564) = 24.30$, $p < .001$, $\eta_p^2 = .146$) were found. Pairwise comparisons (Bonferroni corrected) showed that on Outputs, lower levels yielded scores lower than those of higher levels (all $p \leq .001$). For hiddens, the same effect was found (all $p < .001$) except for levels 1 and 2, between which no significant difference was found ($p = 1.00$). The interaction effect, again, can be explained by the apparent absence (see Figure 3.3(d) of an effect of Hiddens within the 0 outputs group, in contrast to an increase in performance along with the number of hidden units in the other output groups.

3.5 Conclusions

The main conclusion that can be drawn, is that the number of hidden and output units in an agent's IE have an effect on the agent's ability to adapt to stimulus-reward relation reversals. However, the specifics of these effects differ strongly between rounds.

The first round appears to be a special case, in which performance is relatively high for all configurations. The significant increase of performance along with the number of units can be explained as an artifact caused by the exploration behavior of the Q-learning algorithm. As explained in Section 2.1.1, for values of $\varepsilon > 0$, the algorithm selects a random action rather than the action associated with the highest Q-value in a percentage of the time steps. Assuming an optimal stimulus-response association, the chance of selecting the correct response action at a given time is given by

$$p = 1 - \frac{\varepsilon}{2+2x} \tag{3.2}$$

where x is the number of output units. With $\varepsilon = .1$, as it was set in the experiment, this amounts to average scores of .950, .967, .975, .980, .993, and .996 for 0, 1, 2, 3, 6 and 12 output units respectively, which corresponds well with the actually obtained scores of respectively 9497, 9735, 9818, 9861, 9918, and 9951 over 10,000 cycles.

The results of the second round are arguably most important, as they reflect the agents' ability to adapt to a reversal. Since the performances in the first round were highly similar, the role of any potential bias effect can be considered negligible. In other words: differences in performance in the second round are most likely to be caused by *how* associations are learned and stored, not *which* associations are stored to what extent. As expected, the number of hidden units does not make a difference if there are no output units. The same goes vice versa: the configurations with no hidden units show equal performance over all levels of Output. Among the other configurations, the performance improves along with both Outputs and Hiddens. This suggests that more a efficient adaptation to the first reversal is achieved by agents with more units in their IE.

The third round shows an opposite pattern in terms of performance when compared to the second round: the configurations with no hidden units or no output

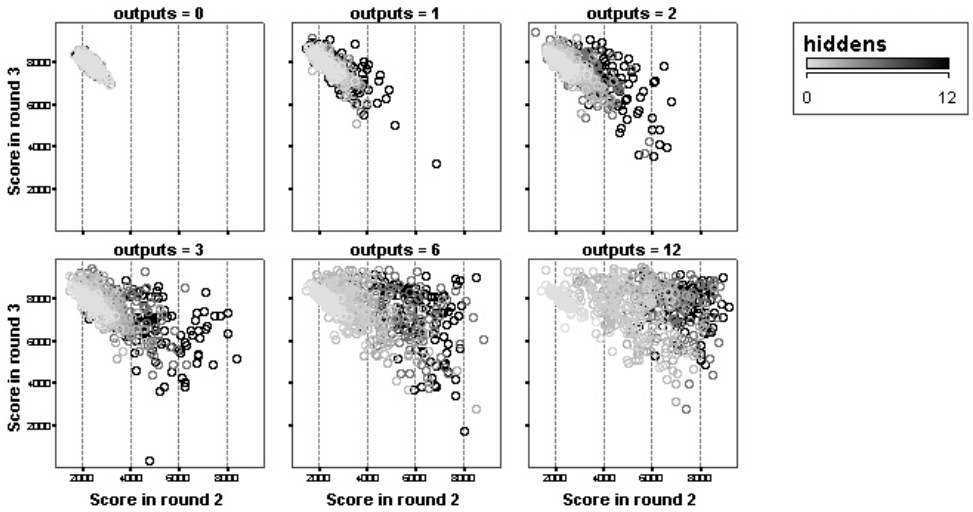


Figure 3.4: Scatter plots of the scores of the second (horizontal axes) and third rounds of the reversal learning experiment. Each plot shows a different level of Output. Levels of Hiddens are represented by brightness.

units perform best, while, with exception of Output level 12, additional units lead to decreased performance. Figure 3.4 shows an overall negative correlation between the scores on round two and round three. This can best be explained as a carry-over effect from the second round: agents that did not manage to learn the new relation in the second round, profit in the third as the not-yet unlearned relation of round one becomes correct again. However, agents with sufficient hidden and output units manage to maintain reasonable scores, with differences among configurations being much smaller than in the second round.

In the fourth round, the pattern of round two is restored globally. This confirms the robustness of configurations with substantial internal environments, in contrast to the inability to deal efficiently with reversal of agents without a substantial internal environment.

3.6 Further analyses

The above results show internal environments help agents to deal with stimulus-reward relation reversals, and therefore give us evidence for presentational abilities provided by internal environments. However, it does not tell us *how* such abilities may be provided. Insight in what actually happens during processing, and understanding the role of the internal environment in this issue is essential for a judgement of any value. There are two approaches to gaining such insight. First, the *behavior* of an agent can be analyzed. In our framework, behavior can be taken to include both external actions (the two response actions) and internal actions (IE training). A second approach is to analyze the *internal dynamics* of the agents. I will take both approaches and cover them respectively.

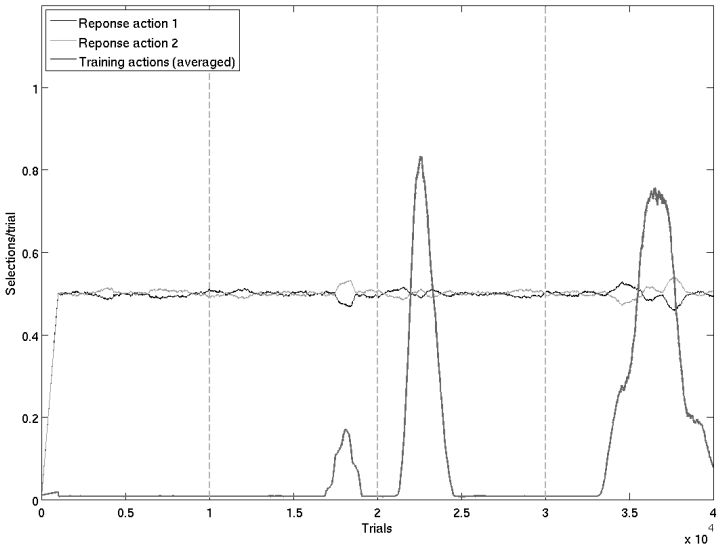
3.6.1 Behavior analysis

Figures 3.5 and 3.6 describe the behavior of agents with different IE configurations. The first of the two figures shows averaged *action plots* for agents with 6 output units and respectively 0 and 6 units[4] in their IE.[5] The action plots show how often each of the actions at the disposal of the agent's control structure is selected per trial over a certain timespan. As expected, all agents select the two response actions about equally often most of the time. Overall, the training actions are executed less often than the response actions in all configurations. The first 'bump' in the selection ratio of the training action takes place some time after the first reversal. It is followed closely by a temporary bias towards one of the training actions (resulting in characteristic 'slits' in the plot). When comparing the plots of the different agents, the most salient difference is the timing of this pattern: the more hidden units, the sooner after the first reversal it starts to emerge.

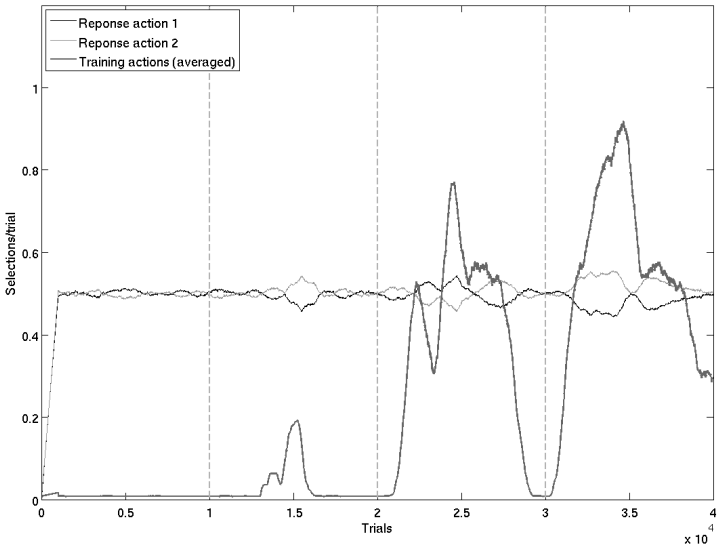
When considering the *performance* of the agents, which is plotted for agents of the same configurations in Figure 3.6, a correspondence with the action plots appears. After the first reversal (at 10,000 trials), the agents with multiple hidden

[4] Additional plots of agents of different configurations are printed in Appendix A, pp. 92–93.

[5] The agent with no hidden units has no functional internal environment, i.e. it has no input-output mapping and can not effectively be structured. In these analyses, it plays the role of agent with no internal structuring. It has been shown (see Section 3.4) to have a performance equal to configurations with no hidden units in the IE, including the configuration with no output units.



(a) IE with 0 hidden units and 6 outputs



(b) IE with 6 hidden units and 6 outputs

Figure 3.5: Action plots for agents with varying configurations. The plots show the selection ratio of each of the actions. All training actions are averaged into a single variable, which is plotted as the thickest line. The plots are of averages over ten runs per configurations, and are smoothed with a moving average filter with window size 1000. Reversal takes place every 10,000 trials.

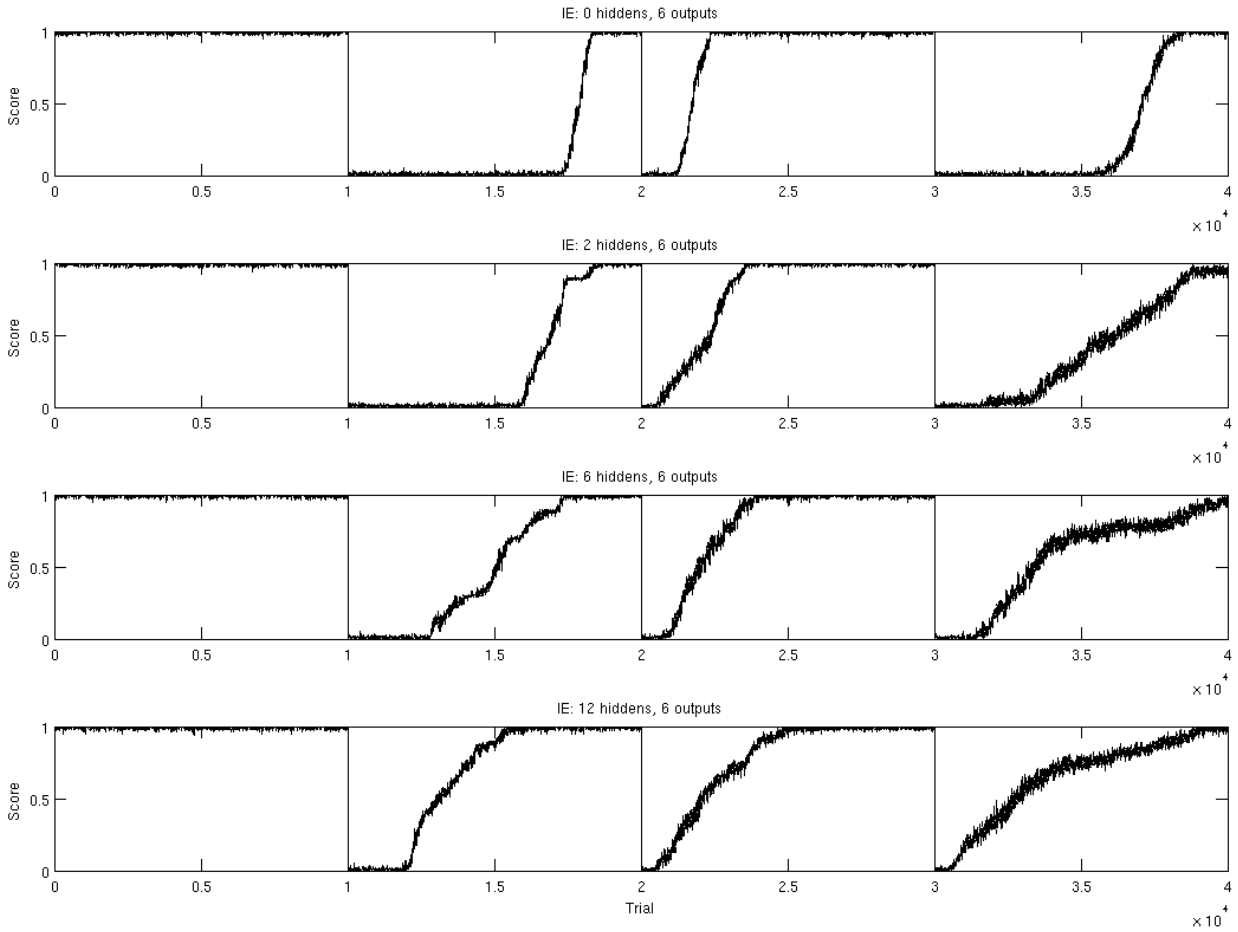


Figure 3.6: The performance of agents of varying configurations. The reversals takes place every 10,000 trials. The plots show averages over ten runs per configuration and was smoothed using a moving average filter with a window size of 100.

units, on average, regain a sound response pattern much quicker than the agent with no effective IE. In the following round, trials 20,000 – 30,000, the agent with no effective IE recovers fairly quickly, a sign that it has not completely unlearned its first association. The agent with 12 hidden units shows a similar performance in this round. Notice that these plots show averages over ten runs per condition. The fact that, in this figure, agents with IE's with more hidden units show a more *gradual* performance increase over the course of a round, can be considered an artifact of this averaging. When inspecting individual runs, it becomes clear that all conditions show similar slopes. However, considering the extremes, all agents with no IE hidden units display the switch at approximately the same point in time; the onsets of the slopes of agents with 12 IE's are distributed over a wider range within the round.

These findings are consistent with the statistical analysis of Section 3.4: agents with no functional IE perform well in the first and third rounds, but not in the second. Agents with sufficient amounts of hidden and output units perform well in all rounds. The behavior analysis makes clear that this performance is a function of the time it takes an agent to adapt to the new stimulus-reward relation. The action plots of Figure 3.5 suggest that this adaptation is related to interaction with the internal environment. To investigate this relation, we need to look into the internal dynamics of the system.

3.6.2 Internal dynamics analysis

Two elements, and their interaction, determine the internal dynamics of the agent: the neural network of the IE and the control structure's Q-function, implemented as a neural network as well.

Unfortunately, neural networks are notoriously hard to interpret. The information they carry is distributed across the entire set of units and tracking the effect of local activation on global behavior is not trivial. Somewhat easier to interpret than the actual information encoded by these networks, are the *dynamics* of their processing. There are two related aspects of the networks that can be inspected as a function of time: the weight settings and the activation levels of individual units. The weight settings define the mapping of input values to output activa-

tions. In the case of the Q-learning network, these output activations determine action selection; the output values of the internal environment serve as input to the Q-learning network. The activation levels at a given time are a function of the input state and the weight settings at that time. Therefore, the weight setting can be thought of as relatively stable *information* carried by the agent that forms the basis of the agent's behavioral pattern, while the activation levels reflect its *computational state* with respect to this information, a more fluid dimension underlying behavior. I will discuss the dynamics of both aspects subsequently in the following sections.

Weight settings

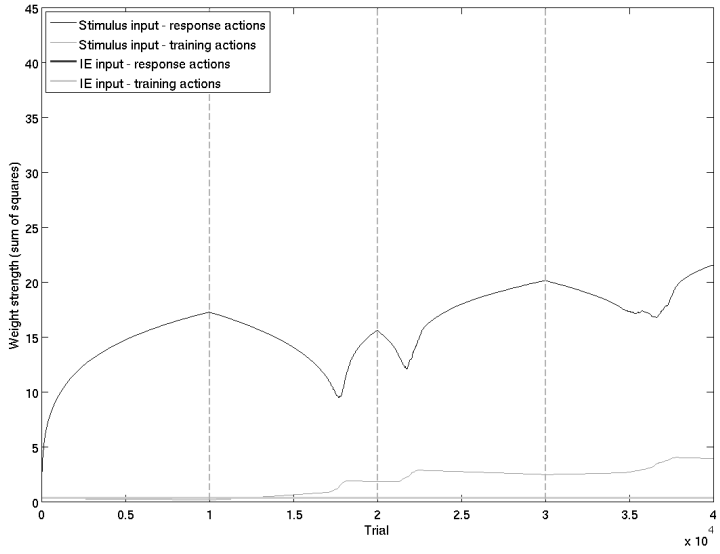
Weights of the Q-learning network The weights of the Q-learning network constitute the mapping from external (sensory) and internal (from the internal environment) input states (a vector $\mathbf{s}$) to goodness values for all actions (a vector $\mathbf{a}$). These weights are trained on basis of feedback, as described in Section 2.1.2. There is no hidden layer in this network, which means that there is one $|\mathbf{s}| \times |\mathbf{a}|$ matrix $\mathbf{W}$, of which value $\mathbf{W}_{i,j}$ represents the amount of (either positive or negative) influence that input state element $\mathbf{s}_i$ has on the goodness of action $\mathbf{a}_j$.

These values, and even more so the development of which, can tell us a lot about the processing that underlies an agent's behavior. Compare the two plots[6] of Figure 3.7 on page 43. These plots show, for each configuration, the *average squared weights* (ASW) of two subsets of the weight setting of the Q-learning network over time. The values are obtained as $ASW(\mathbf{W}_{t,s})$, at each timestep t, for four subsets $\mathbf{S}$ of the weight setting.

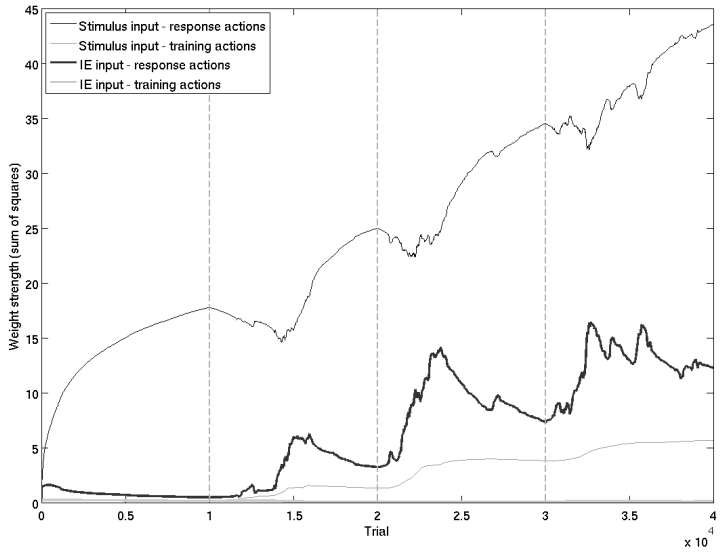
$$ASW(\mathbf{W}_{t,\mathbf{S}}) = \frac{1}{mn}\sum_{m}\sum_{n}\mathbf{W}_{t,\mathbf{S}_{m,n}}{}^{2} \tag{3.3}$$

All configurations develop increasingly stronger weights from the stimulus inputs (i.e. the first ten inputs, thinner lines) to all output units. However, the path by which this rising can be described differs strikingly between agents without an effective IE (Figure 3.7(a)) and those with multiple hidden and output units

[6] Additional plots of agents with different configurations are printed in Appendix A, pp. 94–95.



(a) IE with 0 hidden units and 6 outputs



(b) IE with 6 hidden units and 6 outputs

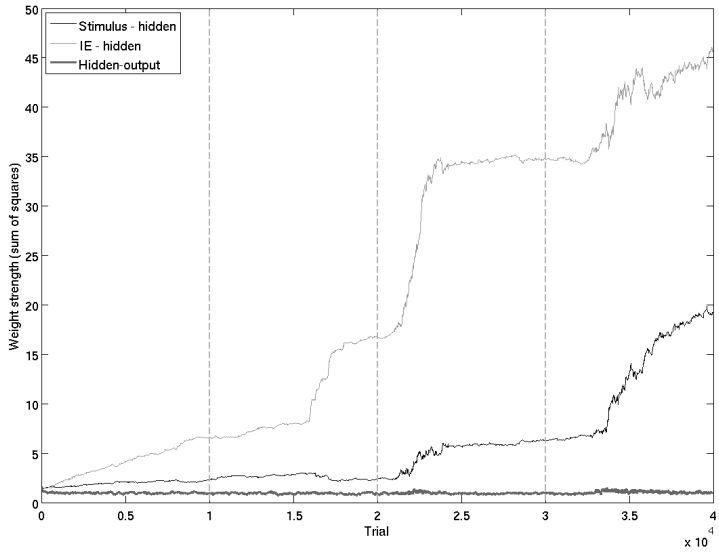
Figure 3.7: The developments of the weight settings of the Q-learning networks of agents in the reversal learning experiment. The plots show the ASW of four subsets of the weight matrix: (1) weights from the stimulus inputs (SI) to the response actions (RA), (2) weights from the SI to the training actions (TA), (3) weights from the inputs that take the IE output values as their activations (IEI) to RA, and (4) weights from IEI to TA. Averaged over ten runs of each of the configurations.

(Figure 3.7(b)). The condition with no hidden units in the IE shows nearly symmetrical 'arcs', that span successive rounds, representing the weights from the stimulus inputs to the response actions. The lowest point after a reversal, clearly visible in the plot at about 17,500 cycles, marks the beginning of re-adaptation. The conditions with more hidden units show an increasingly early interruption of this apparently *natural path* of the weight settings. In the plots, this results in the line getting 'kicked up' repeatedly, preventing it from dropping in each round. This kicking up has to be causally related to the IE, as the only parameter being varied is its number of hidden units. It also structurally coincides with peaks in the weights coming from the recursive IE inputs (the thick line). Therefore, it seems plausible to interpret this pattern as either the IE, or the joint system of IE and control structure, interferring with the adaptation process of the Q-learning mechanism.

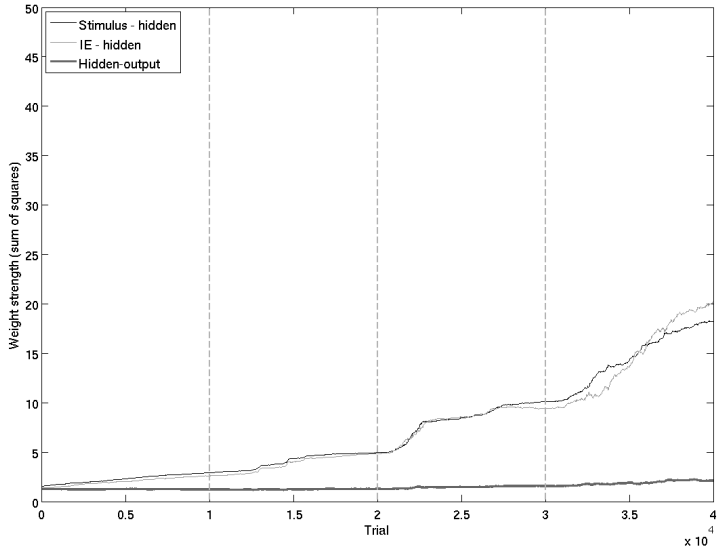
Weights of the internal environment network Figure 3.8 (page 45) shows plots of the weight development in the IE of agents with respectively 1 and 6 hidden units in the IE.[7] Each subfigure shows plots of three subsets of the weight settings of the IE network: the weights from the stimulus inputs to the hidden layer, the weights from the input units that take their values from the output layer of the IE (called 'IE' in the legend) to the hidden units, and the weights from the hidden layer to the output units. The values that are shown are *average* squared weights, so the fact that the networks of these configurations do not have equal numbers of units should be taken into account. Absolute comparison between the conditions is of little value; a better approach is to consider the weights settings of the three subsets relative to each other.

To start with the first graph, which represents an agent with only one unit in the hidden layer of the IE network, the weights coming from the IE input units are rather high in comparison to the other weights, and grow over time. The weights from the stimulus grow too, but stay relatively low. The weights from the hidden layer to the output layer remain at a level of about 1, which is not surprising for

[7] In contrast to previous comparisons, an agent with no hidden units in its IE is not analyzed here, as its weight setting would be defined by empty matrices. Additional plots of agents with different configurations are printed in Appendix A, pp. 96–97.



(a) IE with 1 hidden unit and 6 outputs



(b) IE with 6 hidden units and 6 outputs

Figure 3.8: The developments of the weight settings of the internal environment networks of agents in the reversal learning experiment. The plots show the ASW of three subsets of the weight settings: (1) the weights from the stimulus input to the hidden layer (HL), (2) the weights coming from the IE outputs to the HL, and (3) the weights from the HL to the output units. Averaged over ten runs of each of the configurations.

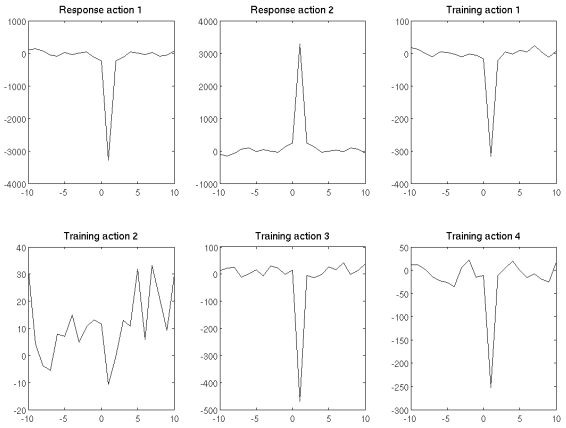
a network with a hidden layer of one unit, which leaves little room for further processing.

The other configurations show much lower average weight levels between the IE inputs and the hidden layer, and therefore weaker contrasts with the other weights. The weights from the stimulus to the hidden layer show a weak decrease as the number of hidden units increases, while the weights to the hidden layer develop a rising pattern, especially in the agents with 12 hidden units. Inspecting the fluctuations that can be seen with respect to the global trend of rising that is present in all weight groups, again a pattern of earlier response in agents with more substantial IE's can be discerned. A final remark that can be made is that the weight adaptation of the agents with more hidden units in their IE are more smooth, i.e. there are less sudden rises or strong fluctuations.

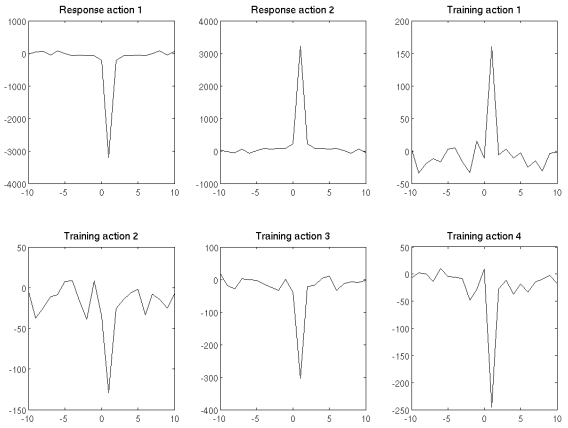
Activations patterns

A reflection of the processing in an agent carrying out the reversal learning experiment can be found in the *activation levels* of the units in the networks. In every trial, a random pattern out of a randomly selected pattern set is presented. As the task of the experiment is to respond according to the set (or class) of the pattern, a well-performing agent should display correlations between the set (labeled -1 and $+1$ respectively) and activation patterns. Calculating *cross-correlations* of stimulus class and the activation level of a units of the neural networks gives an impression of how a unit *responds* to the current and past presentations of patterns from the respective sets. Cross-correlations can be plotted, typically showing a range of *lags* on the horizontal axis, running from negative lags to positive lags. These lags represent a time difference. The value on the vertical axis represent the correlation between the two vectors that are compared, the one vector being shifted in time by the given lag with respect to the other.

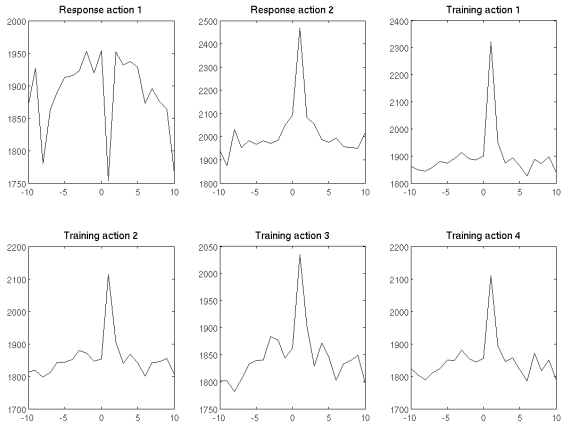
Activations in the Q-learning network Such cross-correlations plots are shown in Figures 3.9 and 3.10 (pages 47 and 48) for the Q-learning networks of agents with 6 output units and respectively 0 and 6 hidden units in their IE. Both figures show four subfigures, representing subsequent stages of the second round of the



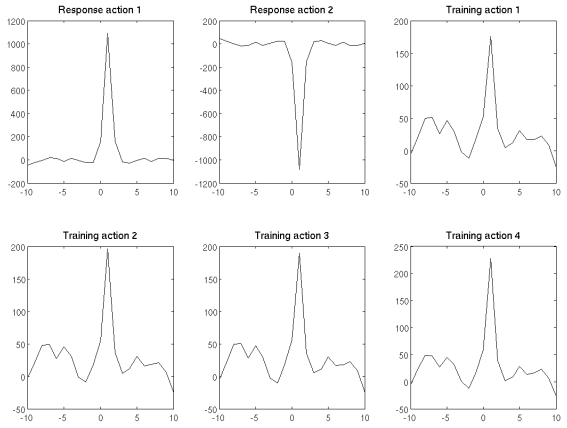
(a) 10,000 – 13,000



(b) 13,000 – 16,000

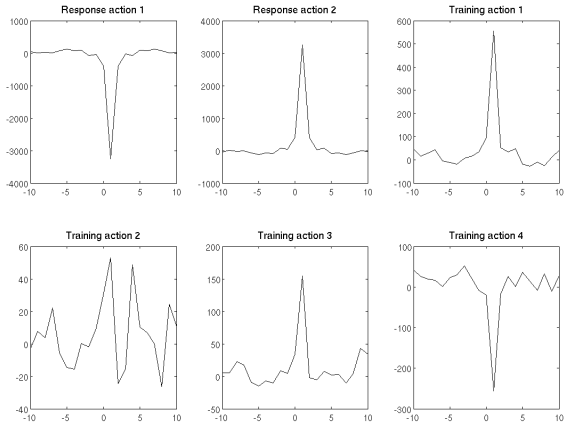


(c) 16,000 – 19,000

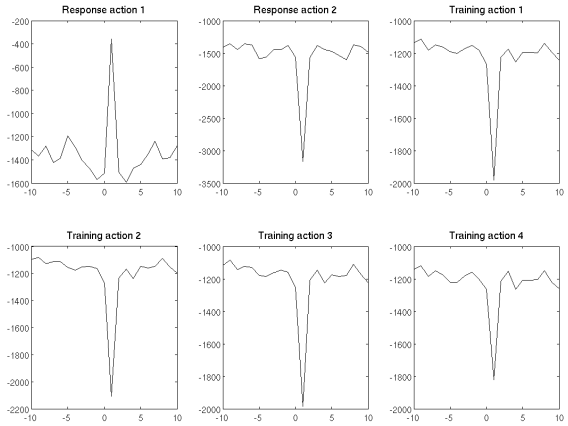


(d) 19,000 – 20,000

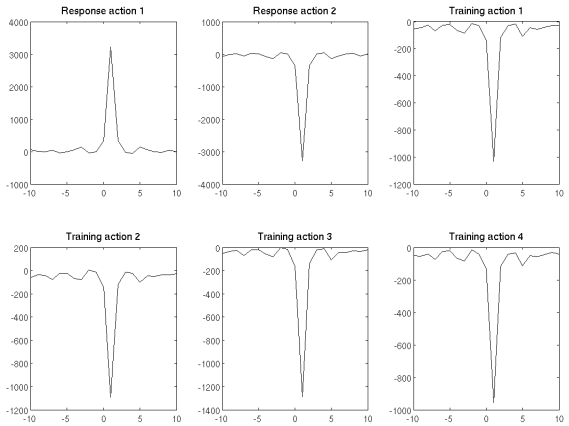
Figure 3.9: The cross-correlations of class (-1 or $+1$) and activation level of the units of the Q-learning network of an agent with 0 hidden units and 6 outputs. The subfigure captions indicate the trial numbers over which the cross-correlation was calculated. Lags are on the horizontal axes, correlation values on the vertical axes. The first two units correspond to the response action; the remaining units to the training actions. Not all of these are shown; a complete overview can be seen in Appendix A, p. 99.



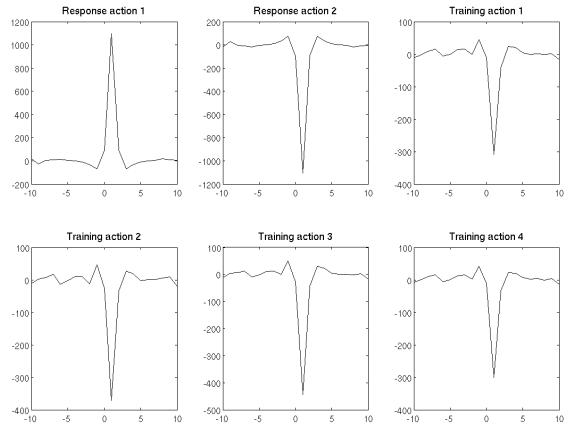
(a) 10,000 – 13,000



(b) 13,000 – 16,000



(c) 16,000 – 19,000



(d) 19,000 – 20,000

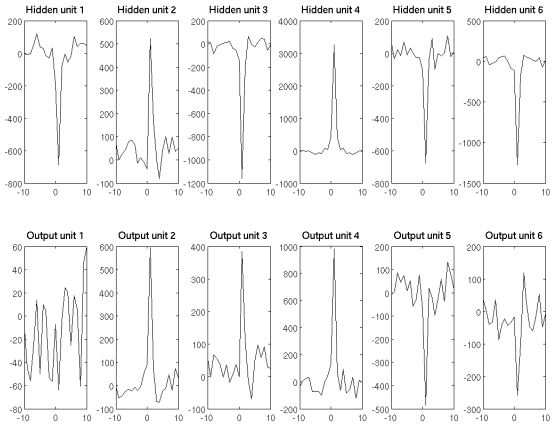
Figure 3.10: The cross-correlations of class and activation level of the units of the Q-learning network of an agent with 6 hidden units and 6 outputs. Not all training action units are shown; a complete overview can be seen in Appendix A, p. 100.

reversal learning experiment: respectively the first, second and third sequences of 3,000 trials and the final 1,000 trials of that round. In the first 3,000 trials (Figures (a)), both configurations have strong correlations between class and activation in units 1 & 2 at the time of presentation. This correlation still corresponds to the correct association of the previous rounds, and thus leads to incorrect responses.

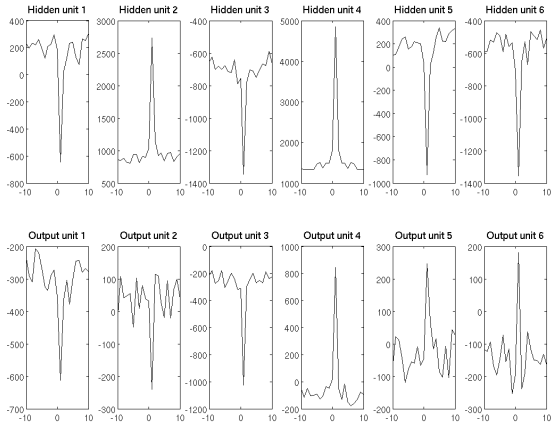
In the following plots (Figures (b) to (d)), the process of adaptation to the reversal can be seen occurring. Eventually, (Figures (d)) both configurations show strong correlations in units 1 & 2 opposite to those at the beginning of the round. In the intermediate plots however, it can be seen that the agent with hidden units starts readapting its associations sooner than the agent without hidden units. In accordance with the above described behavior analysis, the latter reaches a stable and sufficient weight setting after about 16,000 cycles, while the former is much faster to adapt. These findings are in accordance with the performance comparisons, as depicted in Figure 3.6.

The insight that this activation pattern analysis gives us so far, is that agents, whether equipped with a substantial internal environment or not, eventually obtain a direct mapping from input to response actions. However, agents *with* an effective internal environment manage to adapt this mapping much quicker than those without one. Since the structure of the Q-learning network is identical for all compared agents (due to an equal number of IE outputs), the cause for this difference must be sought in the internal environment itself. Therefore, a final subject of inspection is the activation pattern of the internal environment.

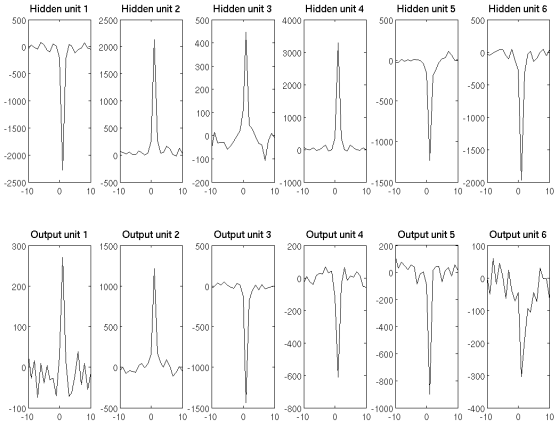
Activations in the IE network Figure 3.11 (page 50) shows cross-correlations of stimulus class and activation levels for the internal environment of an agent with 6 hidden units and 6 outputs. The first six units that are shown are the units of the hidden layer, the remaining units constitute the output layer. Otherwise, it can be read just like the cross-correlation plots of the Q-learning network. The most notable aspect is that most units keep the direction of their association (i.e. their class-activation correlation does not inverse). In other words, the activation pattern with respect to the class of the presented stimulus is relatively stable. This relative stability might indicate a more robust storing of information. The containment of a facility for keeping information over longer periods of time, and,



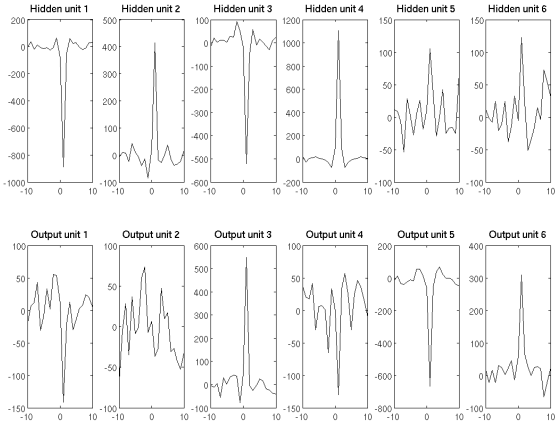
(a) 10,000 – 13,000



(b) 13,000 – 16,000



(c) 16,000 – 19,000



(d) 19,000 – 20,000

Figure 3.11: The cross-correlations of class and activation level of the units of the internal environment network of an agent with 6 hidden units and 6 outputs.

to at least some degree, in a fashion invariant to the current perceptual state is a feature one would expect from a system capable of forming, keeping and using representation. However, from the present analysis it is not possible to conclude that the internal environment in fact provides or contributes to such a facility.

3.7 Discussion

From statistical analysis of the performance of agents with varying configurations, it can be concluded that an effect of the number of hidden and output units of the IE exists. This can be considered evidence for the role of the IE as a mechanism for maintaining internal presentations. What aspects of processing underlie the effect was examined by means of a qualitative analysis of the behavior and internal dynamics of these agents. A general pattern that can be discerned is that agents with more hidden and output units in their internal environment, manage to adapt faster to the reversal of stimulus-reward relations. Quicker adaptations can be seen at the behavior (selected actions, Figure 3.5) and control (the Q-learning and IE networks, Figures 3.7 to 3.11) levels. This adaptation appears to emerge as a result of processing of the IE and interaction between the two modules. Possible implications of these findings for the issue of representation in the context of this model and in an broader sense will be discussed in Chapter 5.

Chapter 4

A situated agent simulation experiment

4.1 Introduction

The results of the reversal learning (RL) experiment imply that an agent supplied with an internal environment can improve its performance on a task by forming presentations. These presentations however, are not necessarily the same things as representations. Recall that the agents in the RL experiment were always perceiving all that there was to perceive. This renders unnecessary the need for internal representations – inner states, *standing in* for extra-neural states from which they are decoupled. The agent's processing, involving internal presentation or not, could in principle be entirely stimulus-driven.

To test the epistemic structuring framework's capabilities beyond presentation-forming, the second experiment simulates agents that are truly embodied and embedded and have to carry out a more natural task that poses *representational* demands. This simulation is based on the epistemic structuring (ES) simulations of C&S (Chandrasekharan & Stewart, 2004, 2007), described in Chapter 2. Whereas C&S separately investigate environment structuring (the first experiment) and internal structuring (the second), the present simulation combines both, by providing the agents with pheromone dropping as well as internal environment training actions. This makes the simulation more natural, as there are no 'magic' sensors

that tell an agent where it has been or non-egocentric actions that simply guide an agent to a location. It also puts to the test the aptness of a single mechanism for internal and external epistemic structuring, which C&S (2007) claim – but do not back up empirically:

> The model provides a unified account of the generation of external as well as internal structures, as the internal structures are stored using the same process as the external structures (...) Given [the] same underlying mechanism, the agent can transform the world or itself, depending on task and resource conditions. (p.343)

The aim of the present experiment is to validate this claim by testing the effect of a task's representational demands on the parameters of the architecture here hypothesized to be involved with representational processing.

4.2 The simulation

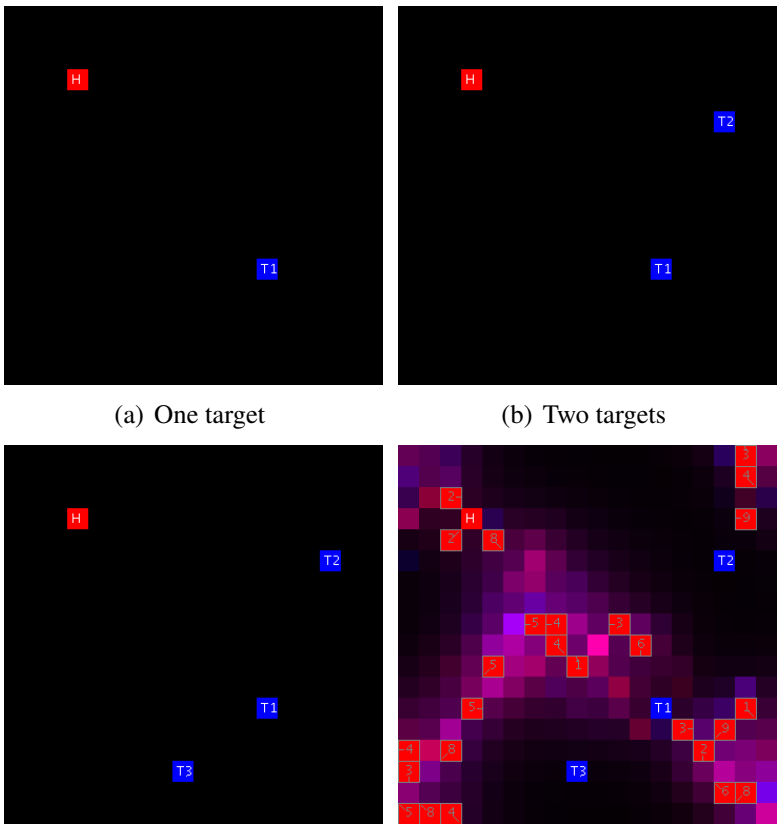
The experiments were carried out in a tailored simulator based on the Q-CON platform (Kapusta, 2008), in which most of the features of the simulators (Stewart, 2006) used by C&S (2004, 2007) were reimplemented. The simulator presents a grid-world environment, and a number of agents in it.

4.2.1 Environment

In contrast to the RL experiment (Chapter 3), the simulation of this experiment has an embodied, embedded multi-agent context. This means that multiple agents simultaneously move about in an environment which they can never sense entirely and act in but only locally (they have a variable location and limited action radius). Figure 4.1 shows visualizations of the environment with different numbers of targets and (Figure 4.1(d)) with agents in it. The environment consists of a 18×18 grid. Although presented as a flat plane, the environment is actually is shaped like a torus: every cell is surrounded by eight neighboring cells. Those cells that appear on the edges of the plane connect, on that side, to the cells on

the opposite edges, such that agents that move over the edge reappear on the other side. Pheromone spreading is not bounded by the edges of the plane either.

Locations of interest are a 'home' location and at least one target location – in contrast to the epistemic structuring simulations of C&S (2007), there can be one, two, or three targets, appearing on fixed positions (see Figure 4.1). All locations occupy a single cell, as do the agents. There can be multiple agents on a cell, and an amount of either kind of pheromone. Pheromones disperse and evaporate according to Equation 2.5 (page 14).



(a) One target (b) Two targets

(c) Three targets (d) With agents and pheromones

Figure 4.1: The environment of the multi agent simulation, with different numbers of targets. The labels H and T_n respectively mark the Home and Target zones. (d) shows the environment with agents in it, which are represented by gray bordered squares. The other squares represent pheromones of different levels, indicated by the intensity of their color. They typically show a mix of H and T pheromones; the brighter the color, the more pheromones.

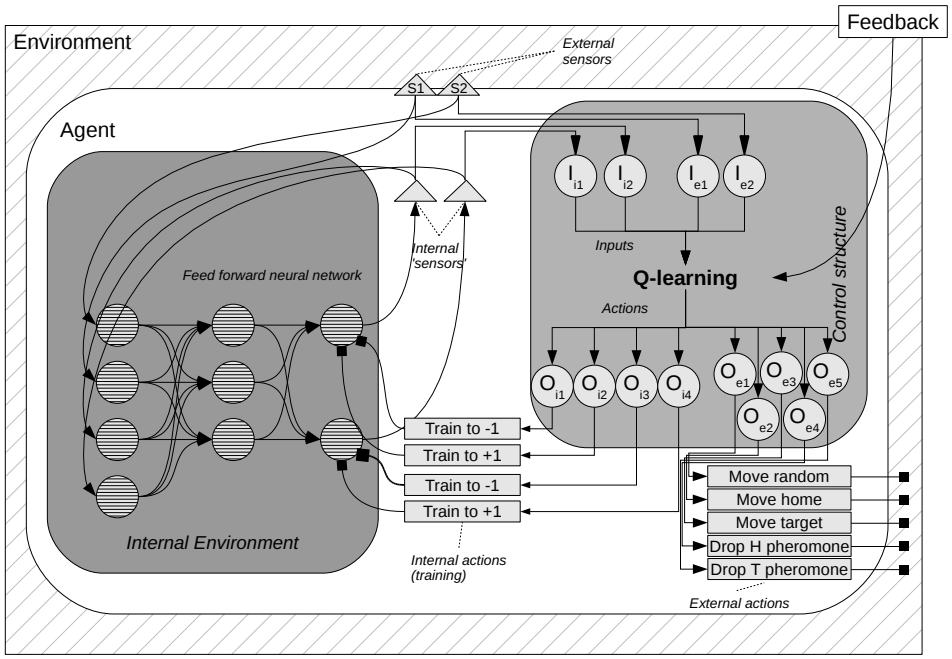


Figure 4.2: A schematic overview of the internal structuring (IS) model as used in the multi agent simulation. It is identical to the model used in the reversal learning experiment (Chapter 3, Figure 3.1), except for the sensors and actions. Lines with arrows are connections, a black square indicates the object of an action. Triangles are sensors, circles are units, either of the IE neural network (horizontal stripes) or of the Q-learning mechanism (solid gray).

4.2.2 Task

The task of the agent in this experiment is based on the foraging task used in the original ES simulations. The agents start their life cycle at a home location and have to visit all (one or more) targets and return to the home location in order to finish their first 'trip'. On every consecutive trip they have to visit all targets and return home again. In some conditions of the experiment an order constraint is introduced that requires the targets to be visited in a *particular order*. A more detailed specification of the task is given below in Section 4.3.2.

4.2.3 The model

This experiment uses the same model as the RL experiment of Chapter 3, except that different inputs and actions are used. This version is depicted schematically in Figure 4.2. To recapitulate the basic workings of the model: A Q-learning based control structure (CS) selects an action on basis of an input vector consisting of values obtained from external sensors and the output activations of the internal environment (IE). The IE consists of a neural network that receives the same input and can be trained by the CS through special training actions. It's outputs serve no function other than as input to the CS.

Sensors and actions

The agents have the same external actions as the agents in the environment structuring experiment (see Section 2.1.3): move randomly, move in a 'home like' direction, move in 'target like' direction, drop a 'home' pheromone, and drop a 'target' pheromone. Like in the RL experiment (Chapter 3), in addition there are two training actions for every output unit of the IE.

Its sensors are as those of the internal structuring experiment (Section 2.1.4). Hence, an agent possesses a 'home' sensor, a 'target' sensor, a 'time since last pheromone drop' sensor and zero or more 'internal sensors' that take the value of the respective IE outputs. The home and target sensors take binary values, the latter of which regardless of *which* target is either present or not in case there are multiple targets. It is therefore a general target sensor, and the agents have no way of *sensing* at which target they are if the sensor is active.

To summarize, an agent can move around randomly and follow trails of pheromones to physically get to the targets and home location, and create pheromone trails by dropping an amount of either kind on its current position. It can sense whether it is on one of the special locations (specifically: home or not, target or not) and how long it has been since it has dropped a pheromone. Finally, an agent has an internal environment which it can sense (and thus potentially use for determining its behavior) and adapt (by setting the desired output activation for the current input) through its control structure.

4.3 Experiment

4.3.1 Agents

The subjects in this experiment are simulated agents, instantiations of the above described model. As in the RL experiment of Chapter 3, the experiment was run with a range of agent configurations, each defined by the number of hidden units (0,1,2,3, or 6) and the number of output units (0,1,2, or 3) in the network of its IE. As explained in the previous chapter, agents with no output units effectively have no IE; agents with no hidden units, but with one or more output units do have an internal environment, although it is irresponsive as there is no coupling between its input layer and its output layer. However, the number of actions and inputs to the control structure is dependent on the number of outputs of the IE. These conditions were included because it cannot be ruled out on forehand that these dimensions have an effect on an agent's performance.

In contrast to the RL experiment, in this experiment the Q-learning mechanism used for the control structure stores its Q-value mappings in a lookup-table rather than a neural network (for a discussion, see Section 2.1.1). For current purposes, lookup-table based Q-learning was considered to be more efficient. As the aim of this experiment is to investigate internal structuring, and no comparison between modes of Q-learning is intended, the use of lookup-based Q-learning rather than connectionst Q-learning is thought to be justified. However, one should be careful when generalizing any results to deviating models (including models with connectionist Q-learning based control structures), as it cannot be ruled out that these are partly due do this specific mechanism. The Q-learning based control structure was equally configured for all types of agents: $\alpha = 0.2$, $\gamma = 0.9$, and $\varepsilon = .1$. The learning parameter of the neural network of the IE was set fixed to $\eta = .2$, and no momentum (see Rumelhart & McClelland, 1986) was used.

4.3.2 Procedure

The aim of the experiments was to investigate the relation between the representational demands of an environment and optimal IE parameters. To do so, the performance of the above described model was measured while varying aspects

of the environment as well as the capacity of the IE.

Factors The representational demands of the environment was defined by two variables: the number of target zones in the environment (1,2,3,4, or 5), and the presence of an *order constraint* (either present or not). The parameters of the internal environment were also varied on two dimensions: the number of hidden units (0,1,2,3, or 6), and the number of output units (0,1,2, or 3).

Measures Performance was assessed by running the simulation for a fixed amount of time under fixed environment and IE conditions. The simulation was run with 30 agents for 80,000 cycles. The average number of trips per agent per cycle was taken as a measure for performance. Per condition, ten runs were carried out, resulting in ten measures for analysis.

To obtain a more easily interpretable and comparable performance measure, 'baseline' runs were carried out such that the performance of a particular run could be normalized with respect to this baseline; that is, expressed as a factor of this baseline performance. Baseline measures were obtained for all environment conditions (see below) by running the experiment with randomly behaving agents. These agents had only external actions (see above), and selected one of these at random at each time step. An average over the performance in ten runs per condition was calculated, resulting in a measure $B_{t,\omega}$ with t indicating the number of targets, and ω a boolean value representing the presence of an order constraint. For each of the experimental runs then, the performance is defined as:

$$P_r = \frac{\overline{trips_r}}{B_{t_r,\omega_r}} \tag{4.1}$$

with r being a run with environment parameters t_r and ω_r (number of targets and order constraint respectively). $\overline{trips_r}$ denotes the average number of trips per agent per cycle in round r.

Expected results The hypothesis that larger representational demands require greater internal structuring capacities in order to perform optimally, predicts that under varying levels of demand (number of targets and with/without order con-

straint), the IE configuration in terms of hidden and output units that leads to the greatest number of trips will vary as well. More specifically: the greater the number of targets, the higher the optimal number of IE units; the order constraint is expected to increase this effect.

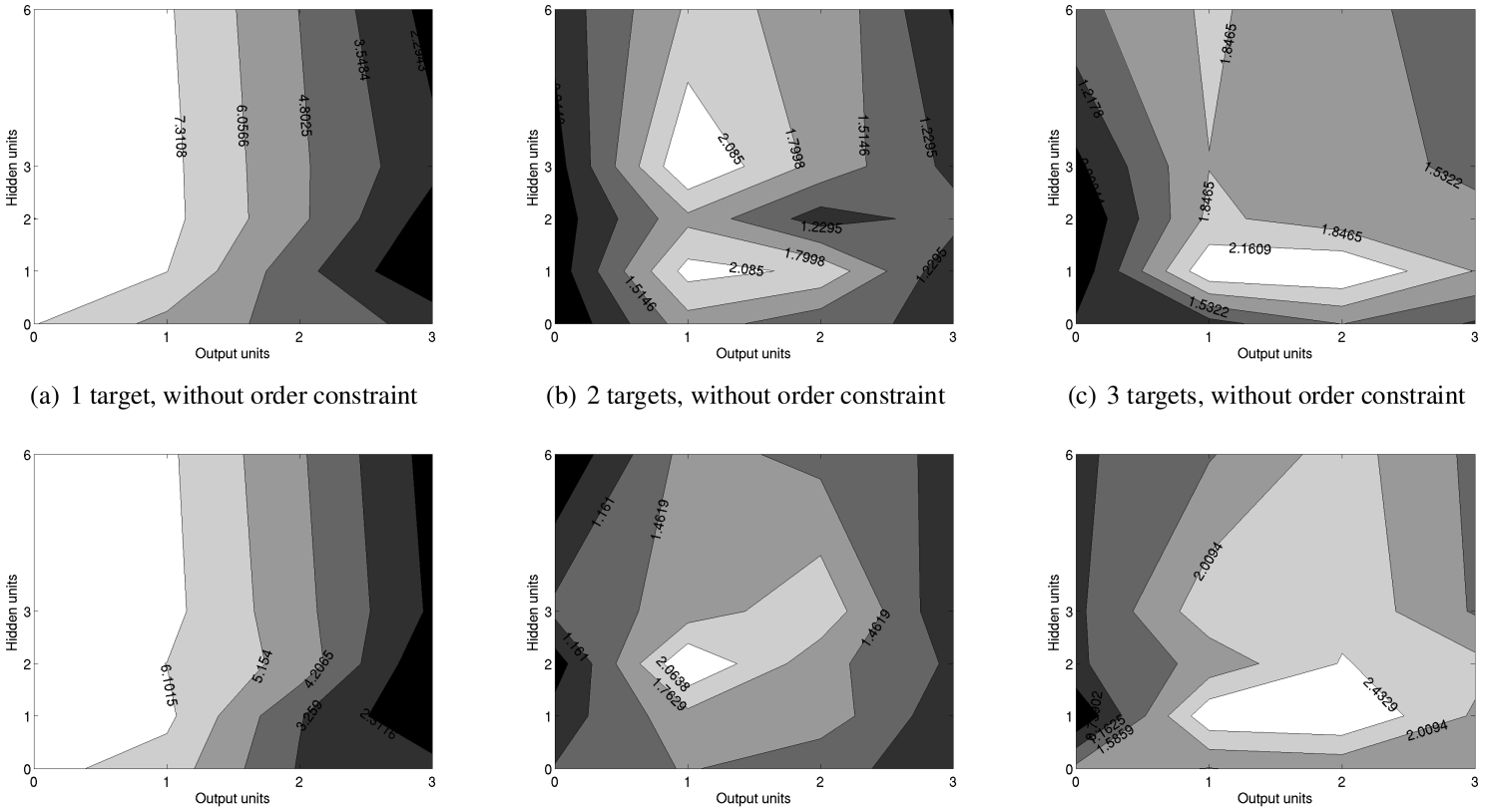
4.4 Results

Figure 4.3 on page 61 presents a first overview of the results of the simulation experiment. The six subfigures show the results of the various environmental conditions: 1,2 and 3 targets, with or without order constraint. Starting with the '1 target' conditions, the 'without order constraint' (Figure 4.3(a)) and 'with order constraint' (Figure 4.3(d)) conditions show highly similar patterns – this should come as no surprise as there is only one 'order' in which to visit a single target. The main trend of the pattern itself is very clear: the more IE output units, the lower the performance; the number of hiddens makes less of a difference. Runs with agents with no effective IE (all conditions with 0 outputs) perform very well in comparison to random behavior: they, on average, manage to complete six to seven as many trips.

In the '2 targets' conditions, the pattern is clearly shifted: agents with no IE output units perform worst, not exceeding chance level. Without or with order constraint (respectively Figures 4.3(b) and 4.3(e)), best performance is achieved by agents with one or two output units: about two times as well as at chance level. However, in the condition without order constraint, the range of the number of *hidden* units that work well is broader, and includes lower numbers. Here, a single hidden unit suffices, whereas 'with order constraint' has two hidden units at its optimum and has its contours skewed to the right a bit more.

The final two conditions, those with 3 targets (Figures 4.3(c) and 4.3(f)) again display a shift albeit less strongly so than between the previous change in number of targets. In general, the 'brighter' areas lie further to the top right areas of the plots, more so for 'with' than 'without order constraint'.

Figure 4.4 shows boxplots for each of the environmental conditions. These boxplots show the distribution for each number of outputs (generalizing over the number of hiddens) within each environmental condition, giving more insight in



(a) 1 target, without order constraint

(b) 2 targets, without order constraint

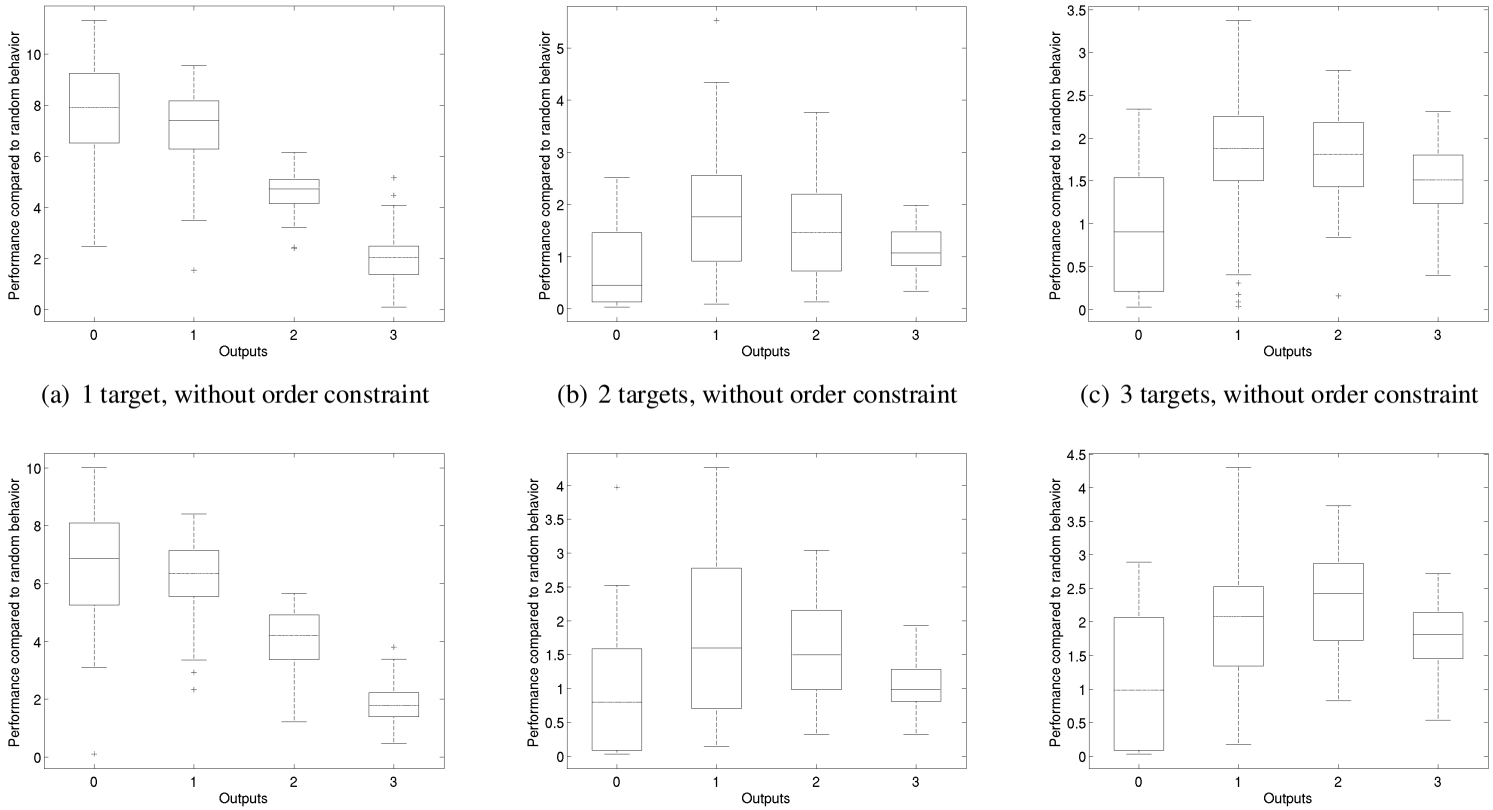
(c) 3 targets, without order constraint

(d) 1 target, with order constraint

(e) 2 targets, with order constraint

(f) 3 targets, with order constraint

Figure 4.3: Contour plots showing the performance of agents of varying configurations on various environmental contexts. The horizontal axes show the number of IE output units, the vertical axes show the number of IE hiddens. The grayscale levels and elevation labels inside the plot represent performance in comparison to baseline performance. Brighter areas indicate higher performance.



(a) 1 target, without order constraint

(b) 2 targets, without order constraint

(c) 3 targets, without order constraint

(d) 1 target, with order constraint

(e) 2 targets, with order constraint

(f) 3 targets, with order constraint

Figure 4.4: Boxplots of the performance of agents of varying configurations on various environmental contexts. The boxplots show the distribution (lowest observation, first quartile, median, third quartile, highest observation and outliers (as crosshairs)) of the agent's performances per number of outputs group for all of the conditions. Plots that include the separate numbers of hidden units are presented in Appendix B. Notice that the vertical axes have different scales.

the distribution of the performance within the varying conditions. Appendix B (page 101) presents more detailed information, splitting out the IE conditions into both number of outputs and number of inputs.

4.4.1 ANOVA results

To confirm the effects suggested by these plots, ANOVA's were carried out for all of the environmental conditions. In all six conditions, significant effects of the number of outputs were found ($p's < .001$). Pairwise comparisons show that in the '1 target' conditions, lower numbers of IE outputs perform significantly better than higher numbers of outputs. In the conditions with 2 targets, no difference is present between 0 and 3 outputs, which have lower scores than 1 and 2 outputs. Of these, the condition with no order constraint has agents with 2 outputs performing slightly worse than agents with 1 output whereas no such difference is present in the condition with order constraint. In the conditions with 3 targets, 0 outputs performs worse than all other number of outputs. Here, in the 'without target' condition, no difference is present between 1 and 2 outputs, both of which outperform 0 and 3 outputs. With targets, the pattern is similar, but has a higher score for 2 outputs all other numbers of outputs, and no difference between 1 and 3 outputs – a skew upwards.

No effect of the number of hidden units is found in most conditions. A convincing main effect of hiddens ($p < .001$) is only found in the condition with 3 targets and no order constraint, which, as can be seen in Figure 4.3(c), has quite a sharp peak at 1 hidden for most numbers of output units.

4.5 Conclusions

The above results are in accordance with the hypothesis as formulated above: greater numbers of targets and, to a lesser extent, the introduction of an order constraint globally increase the optimal number of IE hiddens and outputs.

From the experimental data (as visualized in Figure 4.3 and 4.4), some insights in the underlying causes for these effects can be obtained. To start, in the '1 target' conditions, agents without an effective IE perform best and vastly above

chance level. Even the relatively ill performing configurations perform at least twice as good as randomly behaving individuals. This vantage can be attributed to successful external structuring: making trips back and forth to a single target can efficiently be mastered by collectively maintaining pheromone trails, as was demonstrated by C&S (2004). The configurations with multiple IE output units only get disturbed in the process of learning to structure the environment, as each output adds extra inputs and actions to the control structure that thus burden an agent's resources but add no benefit.

The conditions with 2 or 3 targets provide a qualitatively different case. With multiple targets to visit, especially if in a particular order, mere external structuring becomes increasingly inefficient. In such cases, an agent can benefit from the ability to internalize aspects of the world and its own relation to them, whether or not in a highly task-specific, action oriented matter. Keeping more of such information naturally requires proportionally higher capacities in the resources underlying this function. Hence, the effects that are found in the current experiment correspond roughly to what one would expect in a cognitive system attempting to maintain a balance between representational capacity and computational overhead. Therefore the results of this experiment can be considered evidence for internal environments as a mechanism for storing internal traces of the world that possess a representational character.

4.6 Discussion

In the previous chapter, the behavioral and internal dynamics of the epistemic structuring model were analyzed to gain insight in its workings. While it can be assumed that similar dynamics play a role in more embodied and embedded contexts, such as in the above described experiment, additional questions arise: In what form are representations stored? Are they robust and discrete (like words) or more 'like the core of a bee swarm', as C&S (2007) suggest? In what way do the formed representations influence behavior? How are representations accessed? Can they be actively selected? How do representational processing and reactive behavior relate? And how do internal structuring and external structuring?

These of course are not independent questions. Epistemic structuring, al-

though based on a rather simple mechanism, introduces a potentially very complex dynamics that may not be easy to interpret. A thorough investigation of these dynamics is required before any of these questions can be answered in any satisfactory way. Given that the work presented here provides merely an onset towards gaining sufficient understanding, such an investigation is suggested for further research.

The following chapter rounds up the conclusions that are drawn in this chapter and the previous one, and reviews them in the light of the broader questions stated at the beginning of this thesis.

Chapter 5

Conclusions and discussion

5.1 The forming of internal presentations

In the experiment of Chapter 3, agents equipped with internal environments of varying substantiality were examined while performing a reversal learning task. Performance on such a task should benefit from an agent's ability to form *presentations* (Hurford, 2007; Grush, 1997). From the perspective of representations in the sense of Haugeland's (1991) interpretation, presentations are a weaker kind, as they do not necessarily stand in for things not present to the agent's current perception. However, internal presentation does entail rather sophisticated discrimination and generalization over encountered states. Even if it relies on the stimulus that triggers the presentation to be reliably present, it does attach a kind of *meaning* to it, whether action-oriented or not. Hence, the ability to actively form presentations is an adaptive hookup that gets a long way towards the real thing as far as representations are concerned.

5.1.1 Internal structuring and internal presentations

The RL experiment was intended to investigate whether internal presentations can be employed by agents equipped with internal environments, and, if so, how and to what extent. Performance comparisons showed that agents with more substantial IE's, i.e. networks with more hidden and output units, deal with changes in stimulus-reward relations more swiftly. Investigation of the agents' internal dy-

namics revealed that all agents eventually establish an indexical mapping, which is made up by direct stimulus-response couplings, but that the internal structuring enabled agents manage to readapt relatively early. This readapting coincides with characteristic patterns in both the IE and the control structure. An interpretation of this process is that the reactive, low-level subsystem of the agent is being intervened on by a higher level system consisting of the IE and its interaction with the Q-learning mechanism. In this interpretation, the Q-learning CS is the mechanism that ultimately decides how to act, basing this decision on input from both the external sensors and additional internal information. This information, drawn from the output units of the IE, can be thought of as a *cognitive overlay* from the perspective of the low level CS. This overlay superimposes additional information on top of the original input. The way this can work is immediately clear from the architecture of the model, shown in Figure 4.2: the output values of the IE simply get added to the input array in which the sensory values are kept. Thus, the presentations provided by the IE can be incorporated in the behavior the agent.

An internal dynamics analysis (Section 3.6.2) showed that various units, spread over both networks, show correlations with the class of the current input. These correlations are fairly stable and those in the IE arise as a result of internal structuring actions chosen for execution by the control structure to maximize reward. Hence, they can be considered to serve a *purpose*, which argues (Clark, 1997) for the representational character of the internal structures.

5.2 From presentations to representations

An essential difference between 'mere' presentations as just described, and 'true', or high[1] internal representations is the *standing in* property (Grush, 1997; Clark & Grush, 1999; Chandrasekharan & Stewart, 2007). One way to think of internal representations then, is as presentations – under our current interpretation – that can be accessed, modulated and even instantiated in the perceptual absence of what they refer to (Grush, 1997).

The situated agents of Chapter 4 had to deal with a very limited scope of

[1] referring to their position in the representational continuum, see Section 2.2.

perception. In order to efficiently execute their task, they had to take into account elements in the environment that were not reliably perceivable most of the time. The agents had two ways of dealing with this: first, they could add structure to the environment, thus adapting it to their limited sensory capacities. Second, they could actively structure their IE, thus forming what C&S (2007) call internal traces of the world.

In contrast to the experiments of C&S, the agents in this experiment were capable of internal *and* external structuring, both of which they did, as reported in Chapter 4. Combined internal and external structuring provides a more realistic scenario, where agents interact with their environment and have the potential to keep inner states that represent existing environmental elements, as well as structures that are the result of their own actions. However, it also introduces very complex interactions that make detailed analysis of the agents' individual and collective behavior or their internal dynamics rather complicated. Fortunately, the results of the reversal learning experiments provided a good amount of insight to start with. Considering these results, it can already be assumed that the agents are able to extract presentations from their senses. The remaining question is, whether the agents are also able to apply this ability to elements not present to their current perception. The results of the second simulation experiment did provide evidence for this. In environments with higher representational demands – environments with more targets to visit, in a particular order – higher capacity internal environments appeared to be favorable. The contrast between conditions with relatively low representational demand and those with higher demand is rather stark. In the former, too substantial IE's decrease performance dramatically due to unproportional overhead. As the environment and task get more representationally demanding, the benefits of being able to represent the world start to take over. In a developmental or evolutionary context, once would expect the IE to shrink or disappear (or not appear at all in the first place) in creatures dealing with situations where fast, reactive behavior is the key to survival rather than representation. However, those creatures that end up in environments where going about in a mostly reactive fashion is too risky or inefficient, will adapt by increasing their internal structuring capacities, as putting them to use gets worth it. Hence, the *model itself* conforms to the 'law' of Haselager et al. (2003) mentioned in the

introduction[2]: don't use representations unless it is absolutely necessary.

5.3 Reactive and representational processing

Analysis (Section 3.6) of the epistemic structuring model in action show a striking feature of this model: it integrates reactive and representational processing into a joint system. Without an IE, a Q-learning control structure yields reactive behavior. The addition of an IE does not take away the capacity to have direct mappings between sensory input and behavioral reactions. Rather, an interplay between the two modalities can emerge: direct inputs as well as the cognitive overlay provided by the IE are taken into account by the low-level controller. This way, certain responses can be entirely reactive, while others are mostly the result of IE processing or are reactive under IE supervision.

Reactive processing seems beneficial for an agent when it has to deal with (whether simulated or not) real world situations that impose on the agent unpredictable and potentially harmful conditions at a fast pace. On the other hand, long term planning and even reflection on past events are likely to increase an agent's chance of survival. Therefore, a framework that combines a reactive response mechanism with a representational system constitutes a plausible architecture for embodied, embedded agents.

5.4 External and internal structuring

Thus far, external and internal structuring have been covered as rather distinct modes of epistemic structuring, providing an agent with separate resources, each with their own advantages. However, a complete distinction between the two yields an interpretation that is too simplistic. As C&S (2007) point out, internal structures constitute a "common thread of elements running through contexts and associated actions that lower cognitive load" (p. 344). There is no reason for this action-oriented nature of internal structures to be restricted to task-specific actions only; task-external, or structuring actions, belong to the behavioral repertoire as

[2]Perhaps we should call it *Haselager's razor*?

well. Thus, internal structures, in agents operating in contexts of sufficient complexity and representational demand, can be expected to become connected with an agent's external structuring pattern. This way, a joint internal/external structuring pattern arises in which internal processing can invoke or modulate external structuring and *vice versa*. As a consequence, structures may arise that can be considered to stretch from one end of the entire scope of structuring to the other – crossing the boundary between internal and external environments.

The thought of such a possibility is hardly as esoteric as it might appear, as it accounts for a very common phenomenon. For example, a thesis, at least until completion, can be considered a coherent structure that exists partly in the world as a physical structure, and partly internal to the person writing the thesis, as a collection of more or less explicit ideas, assumptions and intuitions. Until a coherent part of the entire pattern is shifted towards the external mode, the thesis may seem unfinished or even incomprehensible for readers other than the writer. The writer however, will not have such a perception as long as the complementing subpatterns are in his (or her) internal resources. This might explain why it is often hard to detect flaws in one's own writings: to obtain a completely objective perception, one has to disregard all related internal complements. As this example illustrates, compatibility with the *extended mind thesis* (or 'active externalism', Clark & Chalmers, 1998), follows naturally from a tight coupling of internal and external structuring, as already acknowledged by C&S (2007).

5.5 Summary and conclusions

Experimental findings in this thesis suggest that epistemic structuring can be used by agents to form presentations over their sensory input, as well as representations of entities not reliably detectable. The epistemic structuring model allows an agent to engage in both reactive and representational processing. The role of the reactive mechanism as a low-level, reactive controller to which the internal environment adds representational resources, argues for a degree of biological and evolutionary plausibility. Additionally, a joint system of internal and external structuring can account for structural patterns that are distributed over internal and external resources, making the model compatible with the extended mind thesis.

A general observation that can be made, is that the epistemic structuring model provides a framework that fits a broad range of cognitive phenomena, starting with the forming of presentations and representations, hybrid reactive and representational processing and embodied cognition. Admittedly, as yet the framework is broad rather than deep. Both external and internal structures have been considered rather abstractly, or, in the simulations, somewhat ad hoc such as in the case of the pheromones or the choice of a multi layered perceptron as the mechanism of the internal environment. More accurate choices for structuring environments and actions, both internal and external, are likely to exist – an issue that certainly calls for further research.

Chapter 6

Afterword: from labels to language

6.1 Introduction

The previous chapters introduced Chandrasekharan and Stewart's (2004, 2007) Epistemic Structuring framework of internal representations and described two sets of experiments to verify the claim that internal structuring, can provide representational capacities to an agent equipped with such a mechanism.

There are two, related, reasons that make the epistemic structuring theory very compelling. The first, as has been mentioned, is that it describes a, in principle, gapless path from reactive behavior, via simple adaptive behavior and environment enhancement, to cognition guided by internal traces with a representational character. The second is, as will be argued, that this path can be extended gradually into a range of cognitive activities with those collectively called *language* in its extremities. The statement I will speculate on in this chapter, is that the epistemic structuring theory of representations is compatible with established views on the nature, role and origin of language.

6.2 Epistemic structuring and language

The epistemic structuring theory of representation can be said to have had kind of bottom-up development. It originated from a framework of epistemic *environment* structuring and the insight that it could easily be applied to form inter-

nal traces analogue to external pheromone traces. Another unforeseen extension of the framework can be discerned: a number of its functions properties can be aligned with aspects associated with language.

Before further explaining these points, it should be made clear that 'language' will be taken in a pretty broad sense here. It will include the obvious incarnations, such as speech as used in everyday discussions, written texts and sign language. But for the moment, no hard line will be drawn between these meanings and what could be called pseudo-language – in humans – or protolanguage, hypothesized by Bickerton (1990) as a kind of relatively unstructured ancestor of 'true' language. Thus, language here covers a broad range of phenomena, such as tagging items of different categories with different colors (e.g. red boxes go upstairs, the blue ones to the basement), drawing a map, sketching a conceptual schema (or a 'mind map') and the fascinating phenomenon of inner speech.

6.2.1 Similarities between language and epistemic structuring

1. Akin to words, the (internal and external) structures have a *referential* character;

2. Also akin to words, there is no relationship between the form of a trace and its meaning – the form is *arbitrary*.

3. Like language, epistemic structuring can be used to create outer states that lower cognitive load or make the environment more cognitively hospitable;

4. It can be used cognitively in a covert, self-directed matter;

5. It can serve a public goal (as in collective labeling, e.g. pheromone dropping).

Reference and arbitrariness The first two similarities, reference and arbitrariness of form, follow from the representational nature of epistemic structures. In the case of the multi agent simulation, the pheromones (external structures) refer, as far as the agent is concerned, to locations that are out of the scope of perception. The internal traces also have a reference relation, namely with the locations

in the context of a visiting schema. These traces are arbitrary in the sense that their form (e.g. positive or negative output activation) has no a priori relation with the location they refer to.

Reference by patterns of arbitrary form are a key aspect of language. A word composed of the letters R,O,S,E, the Chinese characters that are pronounced *mei gui*, and the spoken utterance /rəʊz/ each form a common noun that can refer to a complete class of flowers or an instance thereof. They also form a proper noun that might refer to one's mother, colleague or any specific other person by that name. Of course there is nothing intrinsically rosy (flowery or girly) about these letters, characters or phonemes – hence, the form of the representations formed by them are arbitrary.

Outer states The third similarity, the possibility of creating outer states that yield cognitive enhancement, needs some more elaboration. Although language may sometimes serve a purely communicative goal, usually, whether intended or not, the symbols of language have the effect of unburdening, enhancing or extending internal processing, and can even be considered to form together a cognitive tool (Clark, 1997, Ch. 10). Performing complex calculation on paper rather than mentally, thinking out loud, or structuring one's confused thoughts into a two dimensional schema are obvious examples of how environmental structures (including those fleeting vibrations of air that are speech) can be used for thought. And, as Bickerton (1990, p. 5) puts it: "A book is a machine to think with." This use of language is not very different from the pheromone trails of the agents of Chandrasekharan and Stewart (2004, 2007) and Chapter 4 of this thesis: some effort is put into creating structures that serve as perceptual input moments – or years for that matter – later.

Self-directedness That the view of language as a mere communicational device is incomplete clearly follows from its *self directed* mode (the fourth similarity). Self directed language can manifests itself overtly as 'private speech', which was recognized by Vygotsky (1934/1986) and Piaget (1959) to play an important role in learning and development. Covert self directed language in the form of 'internal speech' has been studied as a cognitive phenomenon (e.g. Steels, 2003). Whether

through audible utterances or some internal loop, humans obviously use language privately in a form equal or similar to public language. Clark (1997, p. 195) argues that language is employed as a self-directed *tool*, a 'computational transformer that allows pattern-completing brains to tackle otherwise intractable classes of cognitive problems'.

Now how does this relate to epistemic structuring? Physical labeling, a kind of external epistemic structuring, can be though of as imposing categories on objects to decrease future cognitive effort. Internal epistemic structuring does something similar. A context consisting of external input values and internal state values activates the internal environment. By executing an internal structuring (training) action, the agent internally labels that context. In both cases, the agent by means of structuring makes explicit its desired future perception, be it a blue dot or the word 'coffee' on the coffee can, or a positive activation on output unit 3. The latter does not seem to have much to do with language, but the process by which it comes about is remarkably similar to the way private language is used. Take for example studying Ancient Greek architecture. When learning to tell apart the Doric and Ionic orders, a common method is to take a depiction of a column of both orders and alternately inspect these while calling, either out load or with one's inner voice, the names of the respective orders: 'Doric – Ionic – Doric – Ionic', etc. One can also use self-directed language to *provide* a context rather than using what is at hand perceptually: 'Ionic: volutes – Doric: no volutes'. This method of self training seems to serve to reinforce a mapping between a class of external or internal stimuli and some pattern that can be used as a label or trigger for other purposes (e.g. recognizing an architecture style) or as a stimulus in subsequent training.

It can be assumed that in creatures equipped with a system of true language, self-directed syntactical *phrases* can alter internal structures in a similar way. Telling yourself to 'Stop and post that letter when you happen to drive past a post box' may not literally label a concurrent perceptual context, but it does intend to engrave a context-sensitive trigger somewhere outside the loop of short term memory. We will come back to the self-directed mode of language at the end of this chapter, where the relation between cognition and language, and the position of epistemic relative to it will be discussed.

Public role The final similarity concerns the *public* role of language and epistemic structures. As has become clear from the pheromone dropping in the experiments of C&S (2004, 2007) and Chapter 4, external structures can be created and maintained collectively by a population of structuring agents. This allows the agents to profit from each other's structuring efforts and behave more efficiently than they could in absence of a shared, structurable environment.

Another dimension is added if a common form exists for internal traces and external structures. If this is the case, internal traces can be made public by creating isomorphic external structures, which then can be internalized by other agents. Intuitively, this touches language very closely, as internal language can easily be expressed as external language. In fact, the one can be proxy for the other, which is what happens when one internally rehearses an utterance before saying it. Mere shared external structuring though is a far cry from actual language as it does not provide a way of converting back and forth between internal and external structuring. Besides, in many cases such as that of pheromones, internal isomorphisms of external structures are of little added value (the essence of pheromones lies in their *physical* presence and location) or even infeasible (what are internal pheromones?).

6.2.2 Differences between language and epistemic structuring

Besides the above mentioned similarities between language and epistemic structuring, two aspects of language are clearly lacking:

- an intrinsic mechanism of *communication*;[1]
- a notion of *syntax*.

This might appear disastrous for the view of epistemic structuring as a predecessor of language, but such does not have to be the case, as will be argued in the following section.

[1] 'Public role' structuring comes close in that it forms intermediate signals that are sent by one and received by others, but it lacks the directedness (one dog growls at another) and temporary discreteness (a shout illustrates this well) generally associated with communication.

6.2.3 Language: communication or representation?

One of the aspects mentioned to be lacking from epistemic structuring for it to be considered linguistic is that of communication. This might appear to be a major, if not insurmountable objection to it forming the basis for the development of language. For isn't communication the essence of language? And doesn't language find its antecedents in the communication of animals? Bickerton (1990) thinks otherwise:

> [F]or most of us language seems primarily, or even exclusively, to be a means of communication. But it is not even primarily a means of communication. Rather it is a system of representation, a means for sorting and manipulating the plethora of information that deluges us throughout our waking life. (p. 5)

Bickerton's view stands in contrast with a 'classic' view in which language is just a very sophisticated means of communicating beliefs, desires and intentions. His main argument is that not considering language to be primarily representational leads to what he calls the *Continuity Paradox*. It holds that if language is a descendant of animal communication, one would expect to find the difference between the most sophisticated systems of animal communication and human language to be quantitative. Yet, this is not the case: language is open-ended, where all animal communication is restricted to a fixed set of topics that can be communicated; indeed, a qualitative difference (Bickerton, 1990). Therefore, communication cannot be *the* antecedent of language. Bickerton argues that the Continuity Paradox can be solved by showing that language is a system of representation. Then, "we could search for the ancestry of language not in prior systems of animal communication, but in prior representational systems" (p. 23). Given the properties discussed above, ES makes a very suitable candidate for such a 'prior representational system'.

Bickerton (1990) distinguishes two systems underlying representation. He calls them the primary and secondary representational systems (PRS and SRS respectively). Bickerton's concept of representation is a bit off from those common in cognitive science, hence the PRS need not be involved with high representation.

The PRS is a system that incorporates sensory input and is highly species-specific. It is sketched by Bickerton as an input-output system, the complexity of which is determined by the "degree of processing that outputs of sensory cells undergo" (p. 82). In the epistemic structuring model, the Q-learning control structures can be considered equivalent to the PRS.

The SRS then, is the part of the representational system "created by language" (ibid., p. 103). Here, however, Bickerton describes the human case. Later, he states that "Language provides [an SRS], and an SRS is already latent in any creature whose primary system is well developed enough to analyze the world into a sufficiently wide range of categories" (p. 145). Mapping this on the epistemic structuring framework, the IE should be the place where this SRS unfolds from its latent, pre-linguistic nature to the kind of linguistic system Bickerton considers it to be in humans. The ingredient that drives this unfolding is also provided: communication. The presence of rudimentary forms of communication, independent of any latent SRS, may, according to Bickerton, have bootstrapped the latent form up to a linguistic system (p. 146).

Thus, roughly speaking, the PRS can be projected onto the Q-learning mechanism, or reactive control structure, while an IE provides a facility for the development of a non-linguistic proto-SRS, and eventually a true SRS.

Now it is clear how ES fits into a view of language as primarily representational, the path from ES to full-fledged language has to be sketched. The following section presents such a sketch.

6.3 From Epistemic structuring to language

Figure 6.1 shows a scheme that sketches a development from reactive agents to linguistic beings. This sketch should be considered a mere suggestion for such a development. The following elaboration does not provide substantial empirical backing or biological embedding, but rather aims to draw the contours of a model of the origin of language in accordance with Bickerton's theory and the epistemic structuring framework.

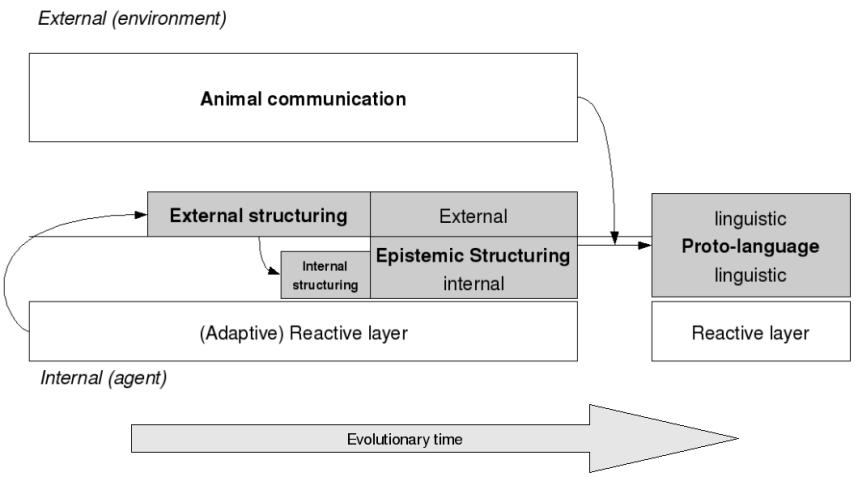


Figure 6.1: Proposed schema of the origin of (proto)language from epistemic structuring. Starting on the left, agents are purely reactive. External structuring develops, and gets internalized, the two turning into a joint system of epistemic structuring. Incorporation of preexisting animal communication leads to protolanguage.

6.3.1 Beyond reactiveness

The external loop that comes to exist when agents add structure to the environment, which later can be perceived and responded to, allows reactive agents to, in a sense, ascend from their reactiveness. Not much is needed for an agent to be able to engage in external structuring, as was shown in the first experiment of C&S (2004): the ability to create appropriate structures and the tendency to minimize loss of energy suffice.

In principle, very complex behavioral patterns may emerge from external structuring. The devices described by Turing (1936), generally known as Turing Machines, clearly demonstrate this: the head and the finite set of rules by which its behavior is governed, essentially form a purely reactive unit with respect to the tape from which it reads (perceives) and writes to (acts upon). Similarly, given the right mapping between perception and structuring actions, which can be learned by the agent on basis of feedback (C&S, 2004), behavior can emerge that goes beyond the a situated agent's reactive capacities. As long as it maintains an ef-

fective mapping, the physical agent can carry out relatively sophisticated behavior without any internal reflection of it. In other words, it does not have to *know* that it knows more as a result of its environmental interactions. Such an agent finds itself (although it won't be aware of it!) very much in a Chinese room (Searle, 1980) situation. For the question whether the agent and its environmental extensions do jointly form an intelligent, or even conscious, whole, one should be referred to the discussions that have been going on since Searle's original paper (for an overview, see Cole, 2004).

6.3.2 The internalization of epistemic structuring

It is now clear how agents can employ and adapt their environment and escape from the trap of their own reactiveness. However, there are clear downsides to a complete dependence on the environment for this leverage. Not only are the environmental structures constrained to specific locations, also is the environment not in complete control by the agent inside it. For example, structures may change or disappear as a result of environmental dynamics or the actions of other agents, causing the agent unable to strongly rely on them.

The advantage of the availability of an *internal* structurable module has been investigated in this thesis and the work of C&S (2007). A proposal of the workings of such a module has been described as well. What remains unclear, however, is how it could come into existence. C&S (2007) point out that one mechanism, Q-learning, can account for both external and internal structuring. Assuming this can be generalized to natural creatures, the remaining question is how the outbound workings of epistemic structuring may have been *internalized* and applied to, presumably, a lump of brain tissue. What is required, in terms of the model, is an action, available to the already present reinforcement learning mechanism, that causes a modification to some remote computational structure – in the model, the IE network. In order to be able to assert that something similar has happened in the brain, a study of its evolution is required. In particular, the evolutionary development of the neocortex seems a good place to start investigating, as it is evolutionary most recent, and is associated with higher cognitive functions. Unfortunately, it lies outside the scope of this thesis to go beyond this suggestion.

For any biological validity of the epistemic structuring model, however, it is vital that traces of analogous developments in the brain can be demonstrated.

6.3.3 From structuring to protolanguage

In Section 6.2.3, Bickerton's (1990) concepts of PRS and SRS have been applied to epistemic structuring. The IE of non-linguistic agents can be said to provide, in Bickerton's terms, at least a latent SRS. A latent SRS bears many of the essentials for language, at least form a primarily representational view of language, but lacks communication.

Non-linguistic communication can be found across a wide range of species in nature (Naguib, 2006) – for a large number of examples, see Hart (2007). Although the linguistic abilities of some species have been debated (see, e.g. Gould & Gould, 1994), most animal communication systems clearly cannot be considered truly linguistic, but rather very *ad hoc* call systems that lack the properties mentioned in Section 6.2.1. Bickerton (1990, 2007) argues that (human) language could not have originated directly from such systems. The proposed role of communication is a different one: it pushes a latent SRS into a full, linguistic one by providing "a set of concrete units that could be handled more easily than raw concepts" (1990, p. 146).

The key to, once again, mapping this development onto the epistemic structuring model, is to consider acts of communication (for whatever purpose) as potential task-external actions, and therefore a means of *external structuring*. The environment that is being targeted, in this case, includes the recipients of the communicative signals. Analogous to the external tracing of the agents of C&S (2004), a systematic, joint epistemic structure of communicative calls – instead of pheromones – can emerge.

Subsequently, there is no reason why an agent capable of producing and interpreting this structure, would not be able to use it for its own purposes, rather than for a public goal. This goes for any other kind of epistemic structuring (like pheromones), so why not for this new mode? A final advantage is gained by short-circuiting the external route and internalizing it by means of a neural pattern isomorphic to the communicational acts. If such sounds far-fetched, think of inter-

nal private speech: is this kind of speech not nearly, or even completely similar to public speech? Introspection strongly suggests this, as do imaging studies of the speech areas of the brain during internal speech (e.g. Hinke et al., 1993; McGuire et al., 1996). In case such a development is accomplished, the Q-learning CS and the IE receive communicational signals as external input, and the CS outputs such signals both externally and internally. Hence, what starts to emerge is a unified form, namely that of the communicative signals, for both external and internal structuring. This obviously provides a great advantage as it allows for straightforward external expression of internal representations.

A solid system, similar of architecture to this, admittedly rather rough sketch, has got what it takes to be considered what Bickerton (1990, 2006) calls *protolanguage*: a rudimentary system of representation and communication that lacks the structure that true language possesses, mostly due to the presence of grammar. For a discussion of the development of full-fledged language out of protolanguage, Bickerton (1990) should be consulted.

6.4 Language, epistemic structuring, and cognition

Clark (2006), building on earlier work (Clark, 1997, 2005), discusses the relations between language, embodiment and cognition. Although his remarks do not get linked to the work of Bickerton (1990), nor do C&S (2007) mention them to be relevant to their work, lining up these various efforts results in a surprisingly broad, coherent and possibly very informative perspective.

Clark (2006) puts his view of language in contrast to what he calls the 'Pure Translation' view of language, which comes down to a view of language items as a mere vessels for thought, used for transmitting such from person to person (p. 370). Alternatively, he argues, language can be considered a cognitive resource in its own right. Specifically, linguistic items can be used to form what Clark calls a thought-enabling *cognitive niche*, a term referring to an "animal-built physical structure (...) that transforms one or more problem spaces in ways that (when successful) aid thinking and reasoning about some target domain or domains" (p. 370). In other words, language can be employed for epistemic structuring. Clark speaks of the "augmentation of biological brain with 'linguaform'

resources" (p. 372), which fits in very well with the concept of internal structuring, although internal structures do not necessarily have to be linguistic, but can also contain, in Bickerton's (1990) terms, raw concepts. The power of linguistic structures, a subset of all possible kinds of epistemic structures, lies in their being transmittable across individuals.

In an effort towards sketching a computational model, Clark (2006) emphasizes the conception of "language as complementary to more basic forms of neural processing", according to which

> language works its magic not (or not solely) by means of translation into appropriate expressions of ['Language of Thought'], but by something more like a coordination dynamics (...) in which words and structured linguistic encodings act to stabilize and discipline (or 'anchor') intrinsically fluid and context-sensitive modes of thought and reason. (p. 372)

The role of language described here, is highly similar to the role of epistemic structuring envisioned by C&S (2007), and as investigated in this thesis. Epistemic structures, whether external or internal, enhance the cognition of an agent by providing very robust traces.

Language – Clark rightly adds the nuance 'and material symbols in general' (p. 370) – in this perspective is the ultimate epistemic structure, as it is extremely robust, fit to store great amounts of information, and can be stored in many forms by many means. The insight that the epistemic structuring model together with Bickerton's view of language add, is that the cognitive role of language discussed by Clark, may actually be that of a larger, developmentally more primitive system of which language is a subclass that fulfills this role the best.

6.5 Conclusions

Although no empirical investigations into the relation between epistemic structuring and language have been carried out, a brief consideration of this topic from the perspective of the model already yields some interesting observations. In par-

ticular, it offers a computational model to two views of language that, as yet, have been motivated theoretically only.

The first view, that of Bickerton (1990), holds that language is primarily representational. The accompanying theory of language evolution holds that a non-linguistic 'primary representative system' predates a, more or less, linguistic 'secondary representative system'. This decomposition can, in a straightforward fashion, be mapped onto the epistemic structuring model – the latter, as has been shown in this chapter, sharing a number of important properties with language.

Viewing language as a kind of epistemic structuring, providing a unified form for its internal and external modes, makes it natural to consider the usage of language as a kind of cognitive niche construction. Precisely this view is being held by Clark (2006). Language can be thought of as the ultimate epistemic structure, boosting all of the advantages of epistemic structuring that opened the door out of reactiveness, into representational processing. Language is the one kind of structure that is very robust, highly transmittable and capable of containing incredible amounts of information. With all the cognitive enhancements it entails, language indeed appears to lie dormant in the internal environment – a representational system, in a cage locked until communication comes by.

References

Arkin, R. A. (1990). Integrating behavioral, perceptual and world knowledge in reactive navigation. In P. Maes (Ed.), *Designing autonomous agents* (pp. 105–122). Cambridge, MA: MIT Press.

Ayala, F. (n.d.). *Natural selection as an opportunistic process.* Retrieved September 2, 2008 from http://www.counterbalance.net/evolution/oppor-frame.html.

Balkenius, C. (1995). *Natural intelligence in artificial creatures*. Unpublished doctoral dissertation, Lund University Cognitive Studies 37. Available from `http://www.lucs.lu.se/Christian.Balkenius/PDF/balkenius.1995.Thesis.pdf`

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

Bickerton, D. (2006). Protolanguage. In K. Brown & A. Anderson (Eds.), *Encyclopedia of language & linguistics* (pp. 235–238). Oxford, UK: Elsevier.

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, *117*, 510–526.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.

Brooks, R. A. (1991). Intelligence without representation. *Artif. Intell.*, *47*(1-3), 139–159.

Brown, K., & Anderson, A. (Eds.). (2006). *Encyclopedia of language & linguistics*. Oxford, UK: Elsevier.

Chandrasekharan, S., & Stewart, T. C. (2004). Reactive agents learn to add epistemic structures to the world. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the cognitive science so-*

ciety, cogsci2004, chicago. Hillsdale, NJ: Lawrence Erlbaum.

Chandrasekharan, S., & Stewart, T. C. (2007). The origin of epistemic structures and proto-represenations. *Adapative Behavior*, *3*(15), 329–353.

Clark, A. (1997). *Being there. putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Clark, A. (2005). Word, niche and super-niche: How language makes minds matter more. *Theoria*, *54*, 255-268.

Clark, A. (2006). Language, embodiment and the cognitive niche. *TRENDS in Cognitive Sciences*, *10*(8), 370–374.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.

Clark, A., & Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*(7), 5–16.

Cole, D. (2004). The chinese room argument. *Stanford Encyclopedia of Philosophy*. Available from `http://plato.stanford.edu/entries/chinese-room/`

Deacon, T. W. (1997). *The symbolic species*. New York, NY: W. W. Norton & Company.

Gould, J., & Gould, C. (1994). *The animal mind*. Scientific American Library.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*(1), 5–23.

Hart, S. (2007). *The animal communications project.* Retrieved october 22, 2008. Available from `http://acp.eugraph.com/`

Haselager, W. F. G., Bongers, R. M., & van Rooij, I. (2003). Cognitive science, representations and dynamical systems theory. In W. Tschacher & J.-P. Dauwalder (Eds.), *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics* (Vol. 10, pp. 229–242).

Haugeland, J. (1991). *Mind design: Philosophy, psychology, artificial intelligence*. Cambridge, MA: MIT Press.

Hinke, R., Hu, X., Stillman, A., Kim, S., Merkle, H., Salmi, R., et al. (1993). Functional magnetic resonance imaging of Broca's area during internal speech. *NeuroReport*, *4*(6), 675.

Hurford, J. R. (2007). *The origin of meaning*. Oxford, UK: Oxford University

Press.

Kapusta, D. (2008). *Qcon java framework.* Retrieved 25 july 2008, from http://elsy.gdan.pl.

Kirsh, D. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science*, *18*, 513–549.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*(3-4), 415–452.

Kuzmin, V. (2002). Connectionist Q-learning in robot control task. In *Scientific proceedings of riga technical university 5.serija. datorzinatne.* (pp. 88–98). Riga, Latvia: Information technology and management science, 10. sejums.

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill.

McGuire, P., Silbersweig, D., Murray, R., David, A., Frackowiak, R., & Frith, C. (1996). Functional Anatomy of Inner Speech and Auditory Verbal Imagery. *Psychological Medicine*, *26*(1), 29.

Naguib, M. (2006). Animal communication: Overview. In K. Brown & A. Anderson (Eds.), *Encyclopedia of language & linguistics* (pp. 276–284). Oxford, UK: Elsevier.

Piaget, J. (1959). *The language and thought of the child*. Routledge and Kegan Paul.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. Cambridge, MA: MIT Press.

Rummery, G. A., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems* (Tech. Rep. No. CUED/F-INFENG/TR 166). Cambridge University Engineering Department.

Rylatt, R. M., & Czarnecki, C. A. (2000). Embedding connectionist autonomous agents in time: The 'road sign problem'. *Neural Processing Letters*, *12*, 145–158(14).

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science*, *7*(7), 308–312.

Stewart, T. (2006). *Epistemic structures.* Retrieved on october 23, 2008. Available from `http://www.carleton.ca/ics/ccmlab/epistemic.html`

Sutton, R. S. (1989). *Implementation details of the TD(λ) procedure for the case of vector predictions and backpropagation* (Tech. Rep. No. TN87-509.1). GTE Laboratories.

Thieme, M., & Ziemke, T. (2002). The road sign problem revisited: handling delayed response tasks with neural robot controllers. In *Icsab: Proceedings of the seventh international conference on simulation of adaptive behavior on from animals to animats* (pp. 228–229). Cambridge, MA: MIT Press.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*(2), 230–265.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: The MIT Press. (Original work published 1934)

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. Unpublished doctoral dissertation, Cambridge University, Cambridge, UK.

Ziemke, T., Bergfeldt, N., Buason, G., Susi, T., & Svensson, H. (2004). Evolving cognitive scaffolding and environment adaptation: a new research direction for evolutionary robotics. *Connection Science*, *16*, 339–350.

Appendix A

Additional figures for the reversal learning analyses

In this appendix, additional figures are provided to accompany the results presented in the behavioral and dynamics analyses of Chapter 3.

A.1 Action selection

The following figures show action plots for agents with varying configurations. The plots show the selection ratio of each of the actions over a run of the experiment. Action selection is discussed in Section 3.6.1.

Notice that all *training actions* are averaged into a single variable, which is plotted as the thickest line. The plots are of averages over ten runs per configurations, and are smoothed with a moving average filter with window size 1000. Reversal takes place every 10,000 trials.

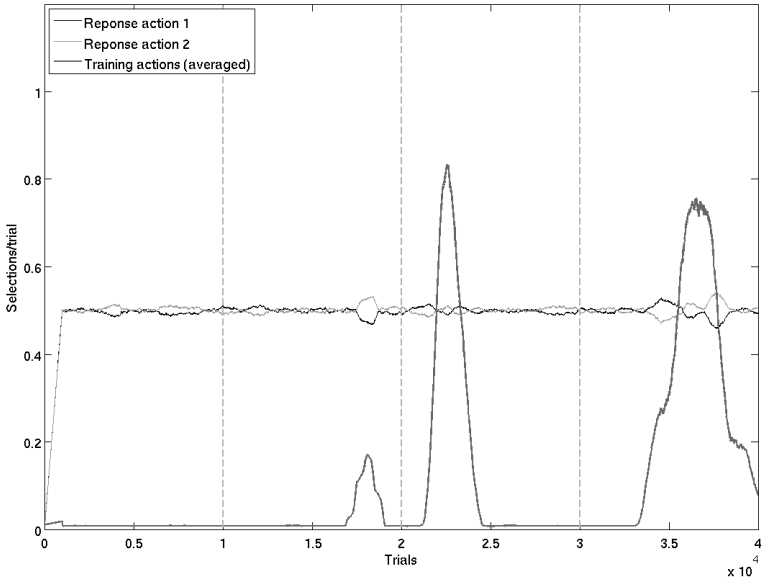


Figure A.1: Action plot of an IE with 0 hidden units and 6 outputs

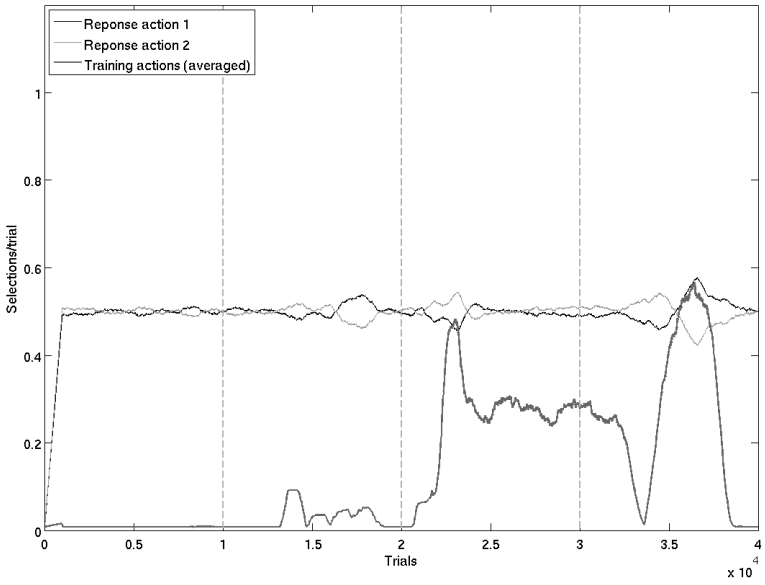


Figure A.2: Action plot of an IE with 2 hidden units and 6 outputs

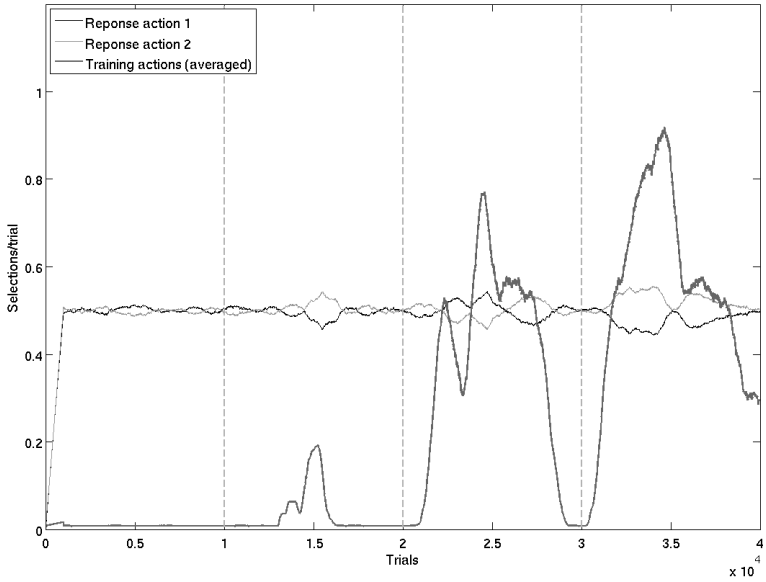


Figure A.3: Action plot of an IE with 6 hidden units and 6 outputs

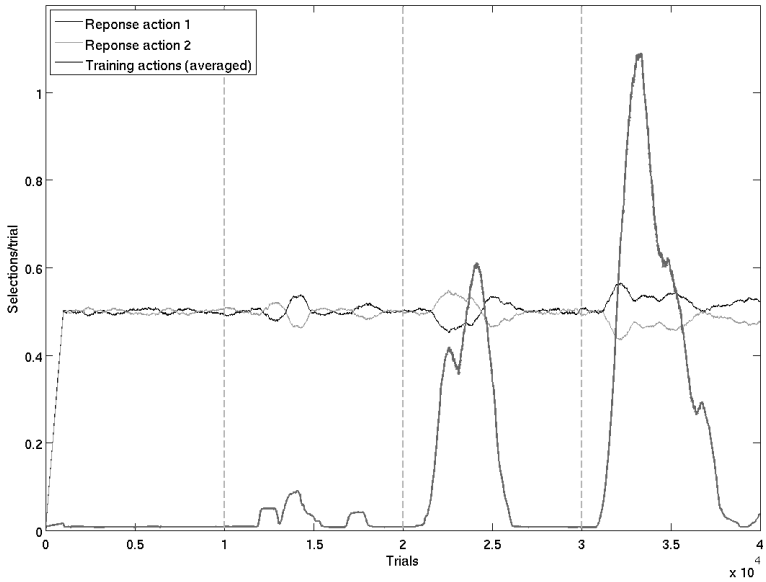


Figure A.4: Action plot of an IE with 12 hidden units and 6 outputs

A.2 Weights of the Q-learning network

The following figures show the developments of the weight settings of the Q-learning networks of agents in the reversal learning experiment, discussed in Section 3.6.2. The plots show the ASW of four subsets of the weight matrix, averaged over ten runs of each of the configurations.

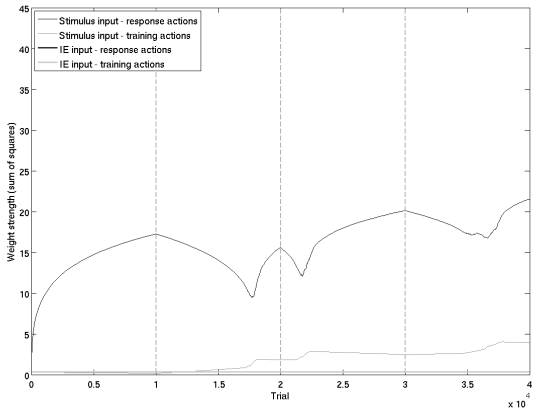


Figure A.5: The development of the weight settings of the Q-learning network of an agent with an IE with 0 hidden units and 6 outputs

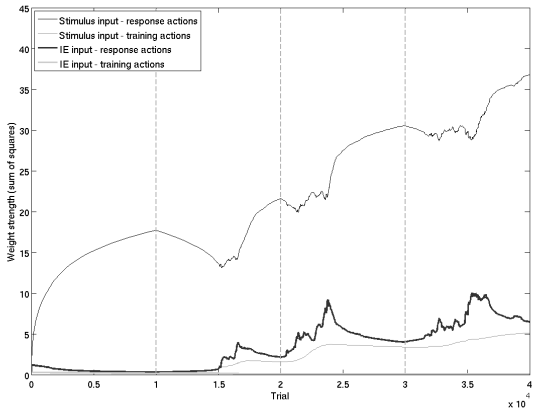


Figure A.6: The development of the weight settings of the Q-learning network of an agent with an IE with 2 hidden units and 6 outputs

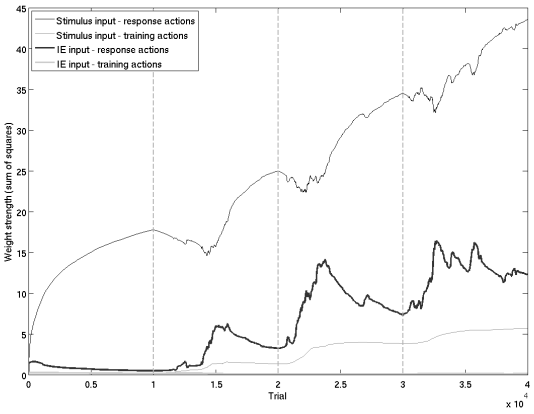


Figure A.7: The development of the weight settings of the Q-learning network of an agent with an IE with 6 hidden units and 6 outputs

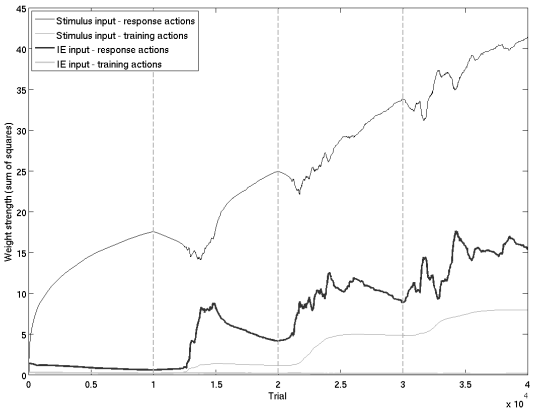


Figure A.8: The development of the weight settings of the Q-learning network of an agent with an IE with 12 hidden units and 6 outputs

A.3 Weights of the internal environment network

The following figures show the developments of the weight settings of the internal environment networks of agents in the reversal learning experiment, discussed in Section 3.6.2. The plots show the ASW of three subsets of the weight matrix, averaged over ten runs of each of the configurations.

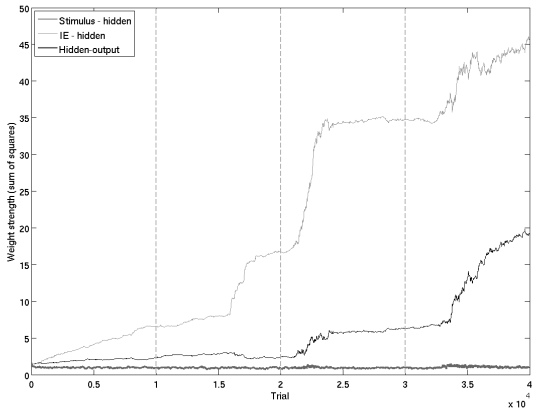


Figure A.9: The development of the weight settings of the internal environment network of an agent with an IE with 1 hidden units and 6 outputs

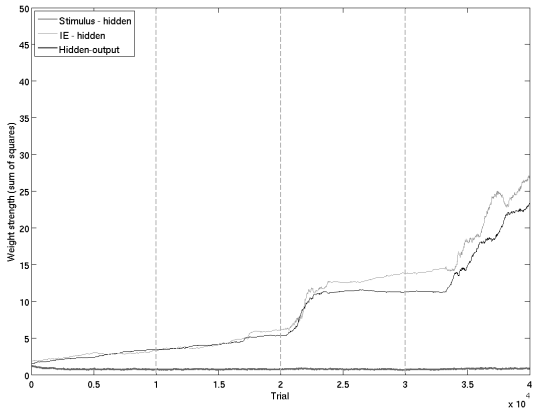


Figure A.10: The development of the weight settings of the internal environment network of an agent with an IE with 2 hidden units and 6 outputs

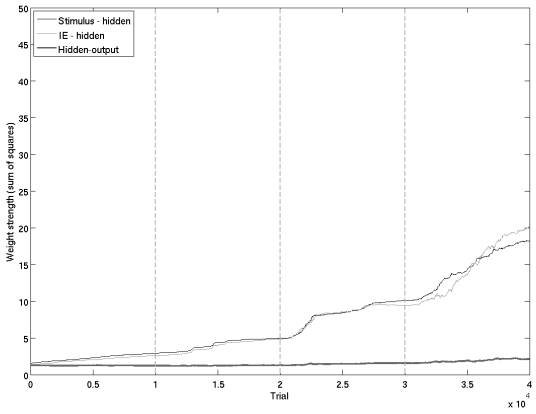


Figure A.11: The development of the weight settings of the internal environment network of an agent with an IE with 6 hidden units and 6 outputs

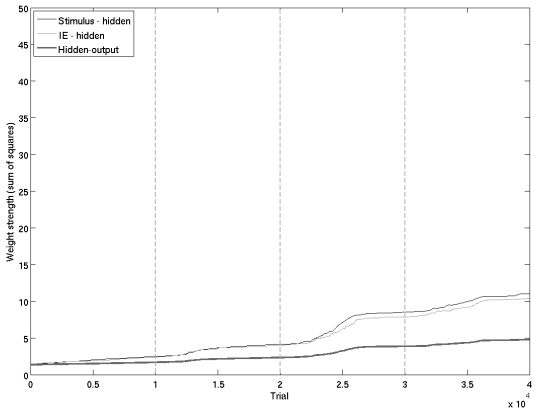
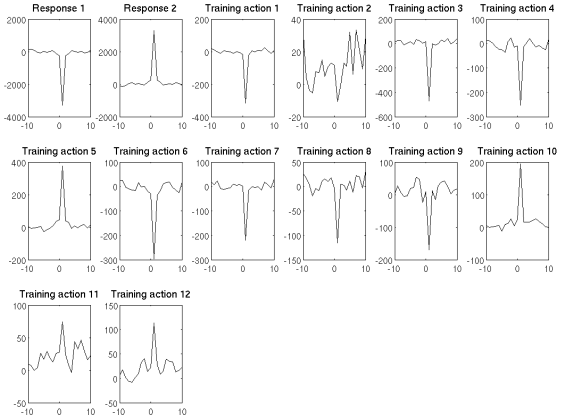


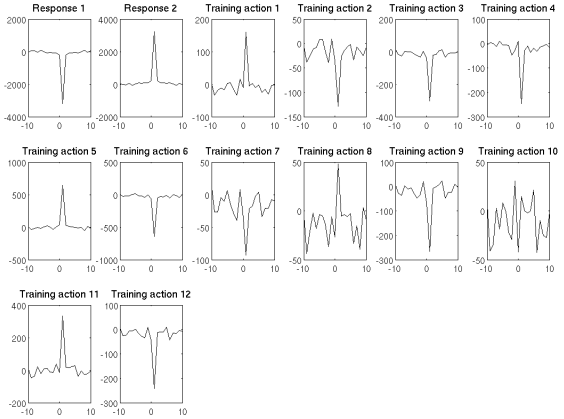
Figure A.12: The development of the weight settings of the internal environment network of an agent with an IE with 12 hidden units and 6 outputs

A.4 Activation patterns of the Q-learning network

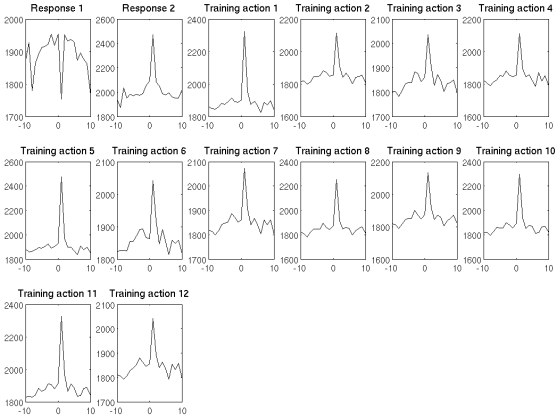
The following figures show cross-correlations of stimulus class and activation level for a single round in two agents, as discussed in Section 3.6.2. The figures below show cross-correlations for all units of the network, hence including those that were left out, for clarity's sake, in the figures included in that section.



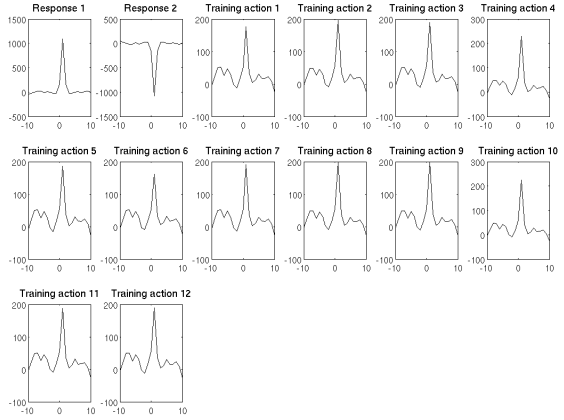
(a) 10,000 – 13,000



(b) 13,000 – 16,000

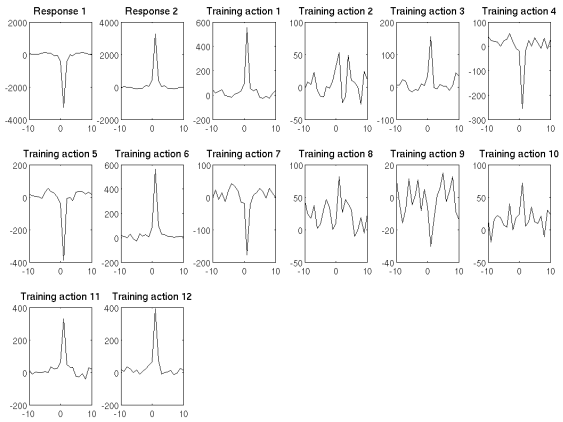


(c) 16,000 – 19,000

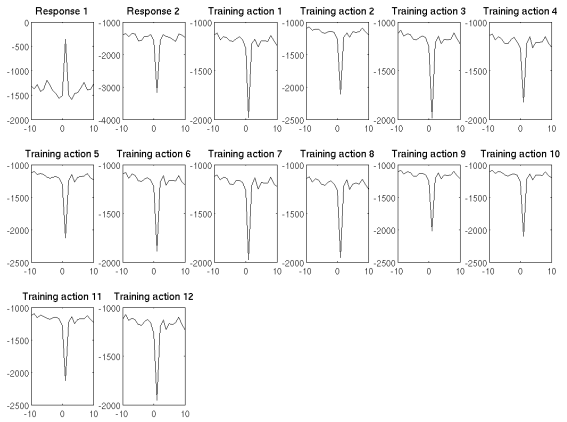


(d) 19,000 – 20,000

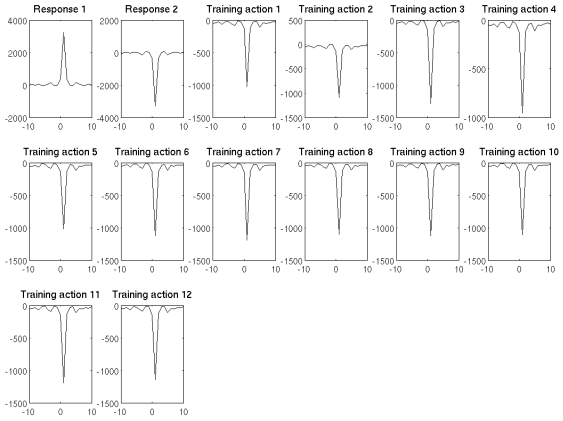
Figure A.13: The cross-correlations of class (-1 or $+1$) and activation level of the units of the Q-learning network of an agent with 0 hidden units and 6 outputs. The subfigure captions indicate the trial numbers over which the cross-correlation was calculated. Lags are on the horizontal axes, correlation values on the vertical axes. The first two units correspond to the response action; the remaining units to the training actions.



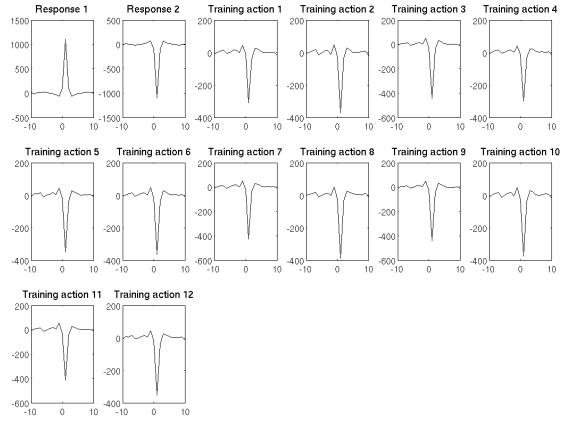
(a) 10,000 – 13,000



(b) 13,000 – 16,000



(c) 16,000 – 19,000



(d) 19,000 – 20,000

Figure A.14: The cross-correlations of class and activation level of the units of the Q-learning network of an agent with 6 hidden units and 6 outputs.

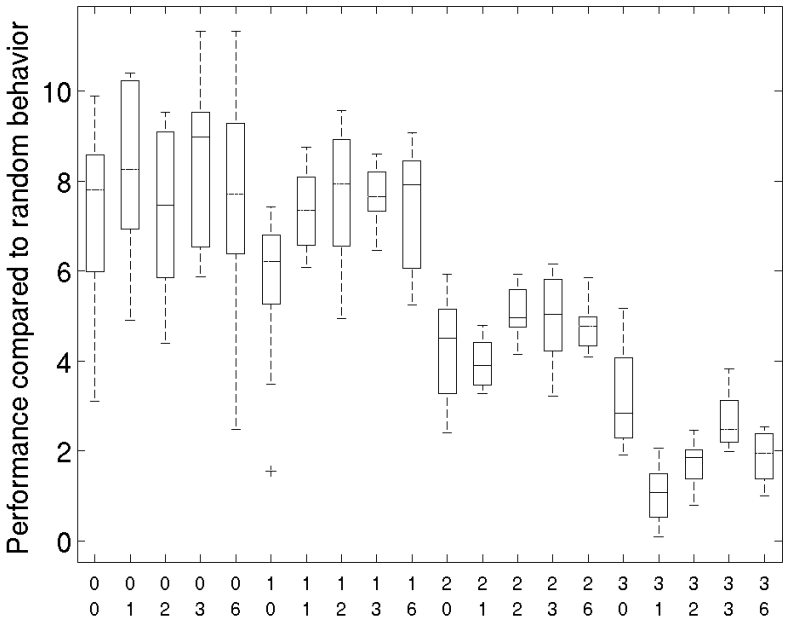
Appendix B

Additional figures for the agent simulation analysis

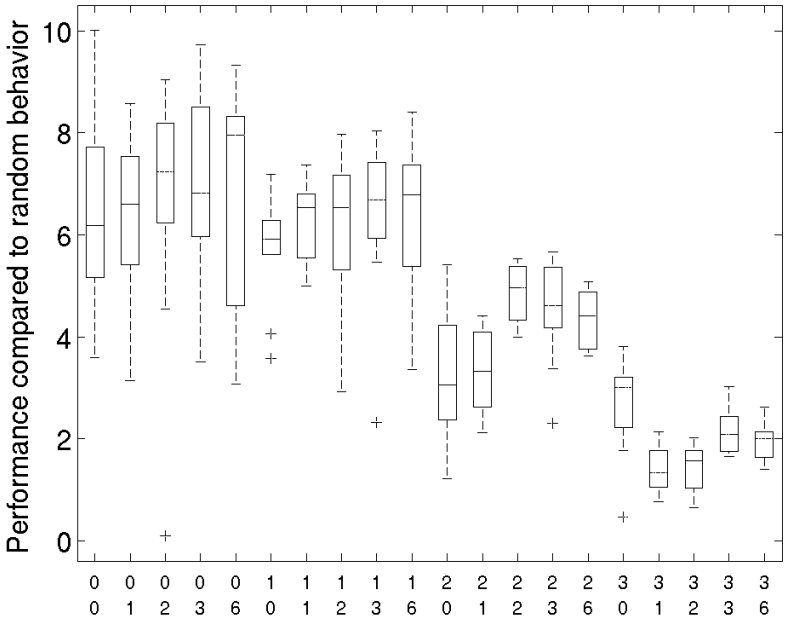
In this appendix, additional figures are provided to accompany the results presented Chapter 4.

B.1 Boxplots

The following figures show boxplots of the performance in the multi-agent simulations of agents of varying configurations in varying environmental conditions. Each of the figures has performance, defined as the baseline normalized number of trips (see Section 4.3.2), on its vertical axis (notice that the plots have different scales). On the horizontal axis are the IE configurations of the agents in the simulation, grouped by number of output units (top row) and number of hidden units (bottom row). The boxplots show, per condition, the distribution of the performance; from bottom to top: lowest observation, first quartile, median, upper quartile and highest observation. Any outliers are indicated by means of crosshairs.

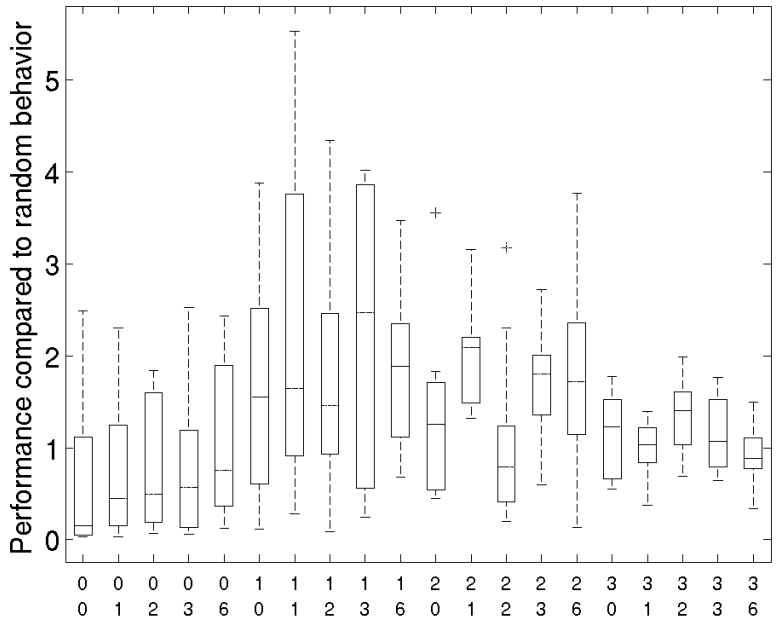


(a) Without order constraint

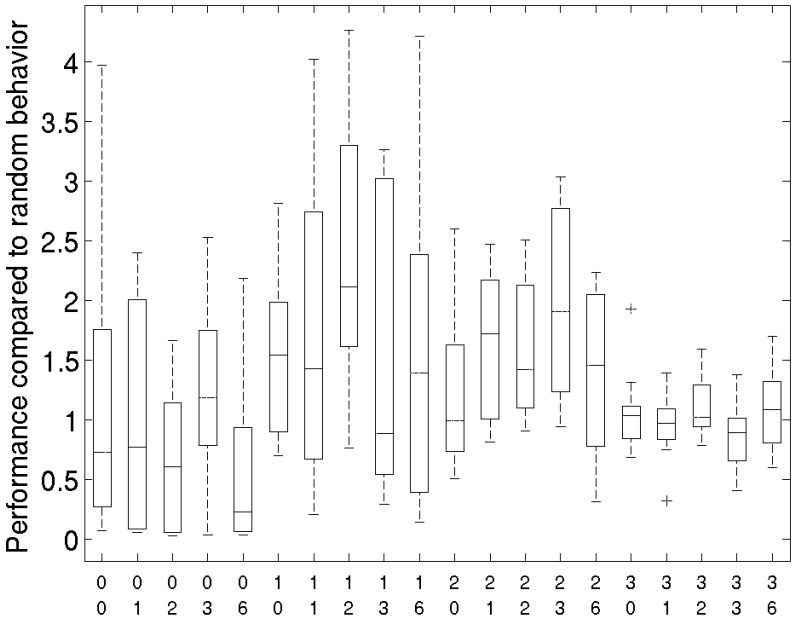


(b) With order constraint

Figure B.1: 1 target

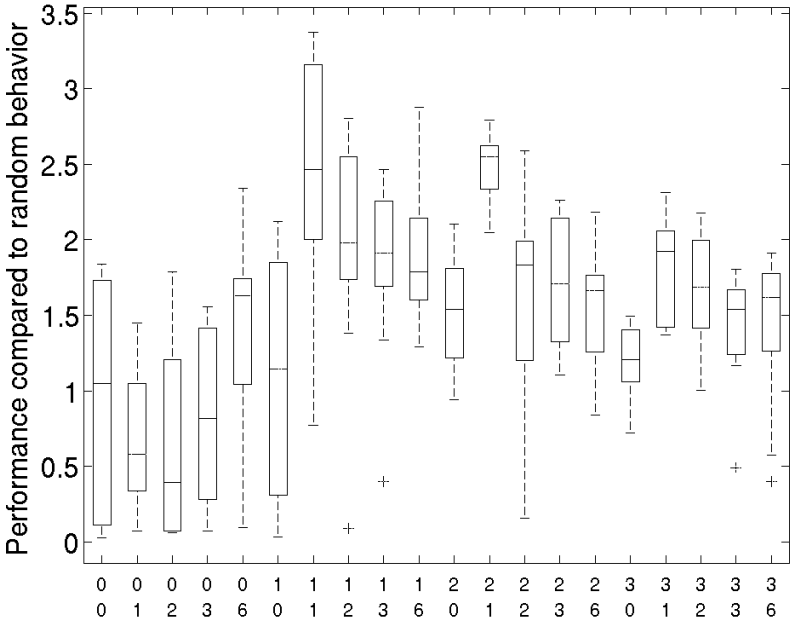


(a) Without order constraint

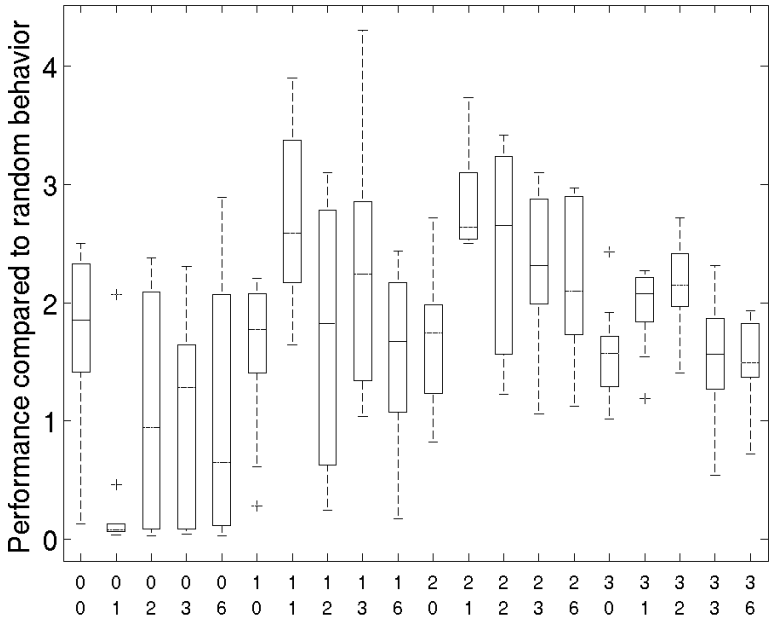


(b) With order constraint

Figure B.2: 2 targets



(a) Without order constraint



(b) With order constraint

Figure B.3: 3 targets