Learning affordance concepts: some seminal ideas

Manuela Viezzer[*†] **C. H. M. Nieuwenhuis**[*]
[*]Thales Research and Technology Nederland, 24 Delftechpark, 2628 XH Delft NL
[†]School of Computer Science, The University of Birmingham, B15 2TT Birmingham UK
{Manuela.Viezzer, Kees.Nieuwenhuis}@decis.nl

Abstract

Inspired by the pioneering work of J. J. Gibson, we provide a workable characterisation of the notion of *affordance* and we explore a possible architecture for an agent that is able to autonomously acquire affordance concepts.[1]

1 Introduction

According to [Gibson, 1979], affordances are perceptual properties of the environment, that become apparent when perception is approached from an ecological perspective:

> The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.

A basic assumption of the ecological perspective is that the animal and the environment have been co-designed through evolution and are therefore 'mutually compatible'. This implies that the animal and the environment can only be adequately described when considered in relation to one another. And that in order to understand an animal's perception one must also understand its environment. What becomes apparent when the system animal-environment is observed in its totality is that animals have been designed to detect properties of the environment that are directly relevant to them, usually because their survival depends on these properties.

In Gibson's theory, the ability to detect affordances is accounted for by learning in an evolutionary sense, which implies not the life-time of one single animal, but rather the history of the whole species. This explains why different organisms will be 'attuned' and have the ability to afford different aspects of the environment. Because of the tight connection with evolution, Gibson's affordances also provide a way of explaining innately preprogrammed reactions to the environment, such as babies' interest and preference for human faces. On the other hand, his account does not provide sufficient insight for behaviour responses that are situationally and culturally determined through an individual history, for example the fact that a red pillar box affords postage.

It is possible, however, to extend Gibson's original notion of affordance to cover the whole individual's attentiveness to the relevant features of the environment. And in a sense he himself provides some cues for the extension, when he recognises that affordances also guide the animal's behaviour as to what are possible or impossible actions within the environment. If acquired and idiosyncratic patterns of action are taken into consideration, alongside instinctive, species-specific ones, the notion of affordance can be enlarged to cover any aspect of the environment that is relevant (in a given situation) because either it affords or prevents the bodily interaction of the animal with the environment.

A full account of conceptualisation in terms of patterns of bodily interactions can be found in [Glenberg, 1997]. According to Glenberg, the representations encoded in memory are patterns of action derived from the projectable properties of the environment (i.e. properties of the environment that can be specified by information available in the optical-flow field) *meshed* with patterns of interaction based on previous experiences. These two sets of patterns are integrated by virtue of their analogical shapes, and both constrained by how our bodies work. Such embodied representations are meaningful *tout-court* because they arise from the world, and are directly grounded by virtue of being lawfully and analogically related to properties of the world and how those properties are transduced by our perceptual-action systems. So, for example:

> The embodied meaning of the cup on my desk is in terms of how far it is from me (what I have to do to reach it), the orientation of the handle and its shape (what I have to do to get my fingers into it), characteristic of its size and material (the force I must exert to lift it), and so forth. Furthermore, the meaning of the cup is fleshed out by memories of my previous interactions with it: pouring in coffee and drinking from it. Those memories make the cup mine.

These considerations bring the notion of affordance closer to AI and robotics themes. Within AI, the problem of model acquisition for an autonomous agent is usually framed in pragmatic terms (see [Davidsson, 1996]), in the sense that what

[1]The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

should be captured by the world-model are those aspects of the world that are somehow relevant or potentially relevant to the agent's goals, rather than some kind of metaphysical world structure. Moreover, part of the model acquisition problem is the acquisition (and mantainance) of the conceptual structures in terms of which the world-model is built (see [de Jong, 2000]).

The aim of this paper is to show how the notion of affordance constitutes a good candidate for the building blocks of an action-based, goal-driven world-model, even though it does not exhaust the variety of conceptual structures that might be needed by an intelligent autonomous agent. We will first provide a more workable characterisation of what an affordance is, and then explore the possible architecture of an embodied agent that is able to learn affordance concepts. The overall idea is that objects fall into the same basic category because they can be used to accomplish the same interactive goal, and since the same object may be useful for accomplishing a variety of goals, categorisation can be flexible and context dependent.

2 Affordances: towards a characterisation

Inspired by the works of Gibson and Glenberg, we will say that an affordance concept is a concept denoting a category of objects that are relevant for an agent because they can be used to achieve the agent's goals. For example, if the agent has the goal of increasing its internal energy and eats an object O to achieve such goal, then, if the action succeeds, O can be categorised as something that affords eating. Let's call such a category 'Food', and let's assume that Food items are red, sweet and soft (like strawberries). How can we tell whether an agent has an affordance concept that corresponds to the category Food? In an agent architecture, a condition/action rule expressing something like *If hungry then, if visual field not empty then, if an item in visual field is red, small and round then approach it and eat it, else move on* will prompt an hungry agent with a behaviour that an external observer might describe as 'looking for food': that is, the agent is not just trying to eat whatever happens to be near it, but continues moving around until it finds something that has a certain appearance, something that 'looks like food to it'. Shall we say that such an agent has an affordance concept FOOD that corresponds to the category Food, as described above? Our answer is no, or at least not if this is the whole story about our agent. Probably, something like the rule we sketched above should be present in the agent's architecture, but what really matters in order to establish whether the concept FOOD is present, is *how* such a rule happened to be part of the architecture, in particular whether it is innate or acquired, and whether it has special relations with other rules. Our intuition is that we should say that the concept FOOD is present, if the rule has been acquired and the acquisition process has been using the relation between the action of eating and the expectation of an increase of the internal energy as a result of eating. This is not a definition, however, because it is not the starting point of our analysis; rather, it is its conclusion. Let's now try to explain the reasons that led us to this conclusion.

Along the same lines of other authors [Smith and Medin, 1981; Langley, 1996], we started by taking a category to be a class of entities in the world that are united by some principle(s), and a concept to be an agent's internal representation of a category. Inspired by Gibson's analysis, we then understood affordance concepts as internal representations of categories of things that are relevant to an agent, because either they enable or they prevent the agent from doing something (for example, if we consider the class of things that enable sitting, then the agent's internal representation of such a class, provided the agent has one, is an affordance concept). At this point, our question has been: How can an agent acquire an affordance concept? To be able to answer, however, we needed to be more specific on what affordance concepts really are, on what is the structure of these internal representations. So, we refined our initial question and asked:

1. How does an internal representation that is an affordance concept look like, and
2. How can such a representation be acquired in case it is not yet there when the agent is 'turned on'.

We will first provide an answer to question 1. This will guide us in the process of answering question 2. And finally we will use this last answer to refine and make more precise our initial intuitions about representations.

3 Intuitive answer to question (1): the internal representation of affordance concepts

In AI, 'concept formation' is a problem addressed within machine learning. Two successive phases are involved in concept formation [Langley, 1996]: first of all, certain items have to be clustered together, and secondly a characterising description for the cluster can be searched for. In a later stage, this description can be used to decide whether or not a new item is also to be considered a member of the cluster. One could argue that something like this description, that you search for later on, should already be there in the first place, as the principle according to which the clustering operation is performed. The two however can be distinct: for example, the clustering could be done using a fast pattern-matching procedure relying on some superficial similarity (like when someone initially puts together the pieces of a puzzle because they have roughly the same surface colour patterns) and then the characterising description can be searched for at a more structured similarity level (like checking whether the colour patterns of the various pieces can fit together to form a connected spatial region, such as the region representing part of the tulip field that the puzzle depicts).

On the basis of this distinction between clustering and characterisation we claim that an internal representation standing for an affordance concept should be a two-tiered structure: one part should capture the functional aspect of the affordance, the fact that an affordance enables grouping together things according to their potential use; and another part should capture the discriminational aspect of the affordance, the fact that things serving the same use can sometime be recognised and grouped together also because they are similar under some other respect (for example, graspable things can be similar in shape). It is important to notice that such a similarity relation could be accessible *before acting*, so that its acknowledgement could be used to guide action.

4 Intuitive answer to question (2): how to acquire affordance concepts

Once we got an idea about how affordances should be represented, the next step has been working out a strategy an agent can use to acquire affordances.

Assume the agent has no initial concepts according to which external objects can be categorised, but has a set of condition/action rules relating the performance of an action to definite expectations — for example in terms of values of internal sensors — on the result of that action. The idea is that any such rule could be exploited to perform the initial clustering of external sensors data, so that later on a characterisation algorithm could be run on those clusters to produce a set of discrimination rules. Here is an example:

In cycle c_n the agent, who has no concepts to categorise the objects it encounters, is at target object O, performs the action `eat` and expects the value of its internal sensor for energy to increase. Now, it happens that O is actually food, and therefore the interaction actually causes the agent's energy to increase. In cycle c_{n+1} the agent can therefore evaluate its expectations against the actual modifications of the values of its internal sensors: in this case, the judgement is that expectations has been fulfilled. Assume now that in cycle c_n, besides acting, the agent has also been storing the values of its external sensors as activated by the presence of the object O, for example that it is red, sweet and soft; then in cycle c_{n+1} the judgement about the fulfillment of the expectations on eating can be used to store some further information about O, namely that it is a positive example of the concept THINGS THAT AFFORD EATING. Had the expectations failed, the agent could have stored the information that O is a negative example of that same concept.

As time goes by, an agent like the one considered in the example could develop two related internal databases, one storing the condition/action rules (where the condition part is 'performing an action' and the action part is 'setting an expectation') and the other storing external sensor data labelled as positive or negative. The keys relating the two databases can be seen as the internal designators of the agent's concepts (we can think of them as symbols in the agent's "mentalese" to denote affordance concepts, i.e. *affordance-labels*). In the service of cognitive economy, the database containing the external sensor data can then be processed using a supervised learning algorithm, so that, for every class of positive and negative items associated to the same key, only a characteristic description is actually stored and kept in memory (rather than the whole list of items). The formation of these two related databases however does not yet cover the whole concept formation process, because simply the presence of the two databases will not make any difference in the agent's course of action. So the next bit that we envisaged has been the use of the database information to change, whenever possible, the agent's behaviour. More precisely, we decided to exploit the presence of affordance-labels.

As we said, affordance-labels are associated both to the condition/action rules and to the characterising descriptions of affordances, so the idea has been to provide the agent's architecture with the following control rule: whenever the intention to perform a certain action is triggered, check whether an affordance-label is associated with it; if so, check whether a non-empty characterising description is also associated; if so, use such a description in the procedure to select the target object upon which to perform the intended action (e.g. approach objects that match the description; or avoid objects that match the description), otherwise rely on your default target-selecting procedure.

For the agent architecture just described we can definitely speak of 'ability to acquire affordance concepts' and, after some time, also of 'presence of affordance concepts' and 'ability to use affordance concepts'. On the one hand, by making reference to the agent's architecture these abilities have been made precise [Sloman and Scheutz, 2002]. On the other hand, to get the analysis started and to reach a provisional end-point, certain simplifications and initial assumptions have been made: we limited ourselves to the consideration of primitive, unanalysable actions, and we assumed the presence right from the start of the associations between initial internal state and action, and between action and expected final internal state. This prevents the acquisition of completely new affordances, so that we should say that affordances are learnt in the sense of made explicit: they are already present in a sort of hard-wired way, but a mechanism is present that allow them to be re-presented [Karmiloff-Smith, 1992] in a new representational form, that allows some new form of processing. We also assumed the innateness of the notion of external object: to start with, the agent has no concepts to categorise the objects it meets, but it knows when it is in presence of an object.

5 Deepening the analysis

5.1 Towards a minimal architecture

The architecture described above takes as a starting point an agent which is autonomous in the sense that it has a motivational structure that prompts it with some behaviour, so that, when turned on, it starts interacting with the environment. To start with we assumed that the initial information coded into the architecture, as if it were innate, was a set of condition/action rules associating (i) particular internal states with the intention to perform certain specific actions (for example, the energy level being below a certain threshold and the intention of eating), and (ii) the performance of the action with a content-specific expectation on the results (for example, eating and expecting the energy level to increase). We will now discuss how to weaken this assumption.

First of all, let's assume that the agent, when 'turned on' for the first time, is only able to discriminate between feeling OK and not feeling OK, even though its physiology is such that many different internal states may actually correspond to these feelings: for example, the agent feels OK when it is happy, when it is replenished, when its body assumes a certain posture, etc., and the agent does not feel OK when it feels pain somewhere, when it is hungry, when it is depressed, etc. Let's also assume that i) the agent has a planning mechanism and a reasoning mechanism but ii) at the beginning, the agent has only one goal, namely the goal of feeling OK, so that whenever this is not the case the agent will start acting in order to re-establish the good internal feeling. Since at the beginning the agent knows very little (for example, it does not know which of its action transform which state into which other state, or which state will turn into which other state because of the world laws), it cannot plan or reason simply because there is nothing to plan or reason about. So our question is: what are the knowledge contents that can (and should) be acquired and how is this acquisition going to happen? Let's assume that the agent actually acts only when feeling not OK, and let's consider the following control sequence governing the agent:

- check internal state I; if internal state is OK then do nothing; else act with action A;
- check internal state again; if new internal state I' still not OK continue acting, otherwise store in the memory that a transition is possible from negative internal state I to positive internal state I' via action A.

The idea is to store both rules such as *If internal state I and want to achieve internal state I' then perform action A* and rules such as *If internal state I and intend to perform action A then expect internal state I'*. The first rules are to be used for planning while the second rules are to be used to learn affordances along the lines sketched in the previous section. After a while, an agent with such learning abilities will have at its disposal a set of associational structures relating an internal negative state X with an internal positive state Y via an action A. The problem however is that only motivational internal states that are emotionally connotated (feeling OK vs. not feeling OK) have been taken into consideration. What about representational internal states, that is internal states that have some semantic content? For example, what about a structure associating a certain visual experience with a certain tactile experience via an action of reaching out and touching? Or structures capturing the fact that when the head moves or an object is moved (and hence certain kinesthetic sensations are experienced), the visual scene also changes?

In other words, how can we make our mechanism look for discriminating data among the information that is provided by the external sensors? In [Drescher, 1991] the Schema Mechanism is proposed as a solution to the problem of the permanence of an object. We are more concerned with the problem of the *constitution* of an object. The next example should give some insights.

5.2 A worked-out example

Consider the following initial situation: the agent is hungry and object O is outside hand-reaching distance (for example, the agent must walk 3 steps to reach O with its hand). Also, the world laws (here what matters is probably the agent's physiology) are such that, in order to remove hunger, the agent should eat. Let's assume that O is indeed food, and that the agent knows nothing of its environment, and very little of itself (event though, as we will see, it has some innate capabilities). How could our agent possibly end up knowing that O affords eating?

Since hungry, the agent feels bad and an impulse for acting is triggered: this is an innate bit, in the sense that when the agent feels bad, it automatically starts acting with the goal of removing the bad feeling and with a generic expectation about not feeling bad anymore. At the beginning, however, the agent does not know exactly what to do: it does not know which action, or sequence of actions will remove the bad feeling in this particular case. This means that, in the initial situation, the agent has innate knowledge of feeling bad and feeling good, but does not know that feeling bad (and feeling good) can actually take many different forms (feeling hungry, having head-ache, feeling replenished, feeling that the head is OK, etc.) which depends on different configurations of the internal sensor values (the agent has no prior knowledge of all these possible forms, they are discovered by living, in the same sense in which you discover what is feeling the pain of a broken leg only the first time you actually break one leg). So, there is something like an innate capacity of judging the degree of badness/goodness of an internal state, but not yet a taxonomical knowledge of types of bad/good states. Also, the agent does not know that some bad internal states can be changed into less bad or even good internal states by acting. The only thing that is so to say pre-programmed is the impulse to act somehow when feeling bad, with a generic expectation of feeling better as a result of the action. So, let's assume that, in order to remove hunger, the agent decides to move its hand, reaching for something to grasp.

While the agent moves the hand trying to grasp object O, it will experience a sequence of internal sensations of movement coupled with a sequence of external sensors contents, such as a sequence of visual scenes (since the visual scene keeps changing, for example in terms of relations of the visible parts of the body with the background). Let's assume that this stream of sensations is recorded (not only the two sequences, but the two sequences associated in such a way that the association is to be understood as the fact that the changes in the first sequence — the internal sensations of movement — are *causes* for the changes in the second sequence — the external sensors contents). What can be learnt from such records? Statistical data analysis techniques such as those discussed in [Cohen et al., 2002] could extract from such records (i) characteristic patterns of hand movements HMs, (ii) characteristic patterns of visual scenes VSs, and (iii) characteristic patterns of associations between HMs and VSs. These last association patterns could then be translated into rules such as 'if visual scene VS and hand movement HM then visual scene VS' (with probability p)' where the 'if ... then ...' is understood temporally, that is something like 'visual scene VS can be transformed into visual scene VS' by performing hand movement HM (with probability p)'. Another association pattern that could be learnt and then transformed into a rule similar to the one just discussed would relate feelings of posture (internal sensations about the body configuration) and hand movements (we are here assuming a distinction between kinesthetic sensations of movements and other kinesthetic sensations). All of these rules are extremely important because they can later be used by the agent to plan in a mean-ends fashion. So, let's assume that a data analytic mechanism capable of producing such results is part of our agent.

Up to now we have been assuming that the agent is moving the hand trying to grasp object O. We said however that O is out of reach, so this means that nothing happens in terms of removing the bad feeling of being hungry. So after a while the agent stops moving the hand and tries something else. Here we are making an assumption about another innate bit within the agent's architecture: a control structure that checks the internal state after an action has been performed, to see whether the generic expectation on feeling better has been fulfilled; if not, the impulse of acting should be directed towards another action. In order to perform the statistical data analyses mentioned in the previous paragraph, however, more than a single stream of data should be present: for this, we assume that 'trying an action' is not just trying one precise movement (x degrees East, y degrees North and z centimeters forward); rather, it means exploring for a while the workspace of one effector. So, the control structure that we have in mind would first of all recognise the presence of a 'bad' internal state; it would then set a generic expectation about a better internal state; then it would select an effector, reserve part of the memory to store the data that are about to come, and then yield part of the control to an action procedure that explores the workspace of the selected effector. The internal state will be monitored continously, and after a while, if nothing happens in terms of removing the badness, the higher control structure will resume control, give the stored data to the data anlysis procedure to extract patterns, and then select another effector to make a new attempt of removing the bad internal state. This means that the agent's temporal flow of sensations will automatically receive some 'delimitations': one delimitation is given by the transitions from bad internal states to good internal state, and within these big chunks there will be delimitations corresponding to the exploration of the workspace of different effectors.

Now, let's assume that the agent starts exploring a new action, namely the workspace of the successive activation of three effectors, something like approach-grasp-eat — AGE for short — and that at a certain point the agent ends up in a situation where it is not hungry anymore. Like in the previous case, we assume that all the sensor data are stored in order to be analysed. This time the agent could learn something like: 'if bad feeling of type hungry and action AGE then good feeling of type replenished'. It seems therefore that there are three types of 'if . . . then . . . ' structures that could be learned:

1. if bad feeling of type X and action A then good feeling of type Y
2. if internal feeling of posture P and action A then internal feeling of posture P'
3. if visual scene V and action A then visual scene V'

With the induction of rules such as 1. and 2. the agent learns about causal chains that pertains to its inner flow of sensations; let's call them I-causal-chains. While with the induction of rules such as 3. the agent learns about causal chains that pertains to its outer flow of sensations (data coming from external sensors); let's call them E-causal-chains.

In order to learn that an object affords eating, the agent should learn that certain I-causal-chains and certain other E-causal-chains are related. More precisely, the agent should discover that sometimes an I-causal-chain is the case *also because* an E-causal-chain is the case, that is, because the action is actually an action upon the environment. In order to achieve this type of learning, we propose the following mechanism. At the beginning the agent only learns that certain actions enable the transition from a certain internal state to a certain other internal state and *it assumes* that it is so no matter what the external state is. This means that, whenever an internal state transition happens, the agent so to speak focuses only on the internal side of its flow of sensations, and learns about actions enabling the transition independently of the external state. When the rule expressing such an internal transition is learnt and the agent can therefore have expectations on the results of its actions, there is room for the agent to learn about the external environment. An example of such a progression from internal experience to external experience is described in [Cohen et al., 1996]. In particular, whenever the expected internal transition does not happen the agent could reason as follows: Given the rule I learnt, my internal state X was such that, by doing action A I would then feel Y, but this was not the case now, so the reason for this has to be searched in the external state.

6 Conclusions and future research

To start with, we took the notion of affordance to mean a concept denoting a category of objects that are relevant for an agent because they can be used to achieve the agent's goals. We then suggested that an affordance should be internally represented by means of a two-tiered structure: one part capturing the functional aspect of the affordance, and one part capturing its discriminational aspect. The third step has been the proposal of a strategy that can be used to acquire affordance concepts. First of all we showed how an agent with innate associations between initial internal state, action and expected final internal state could learn affordances and successively use them to direct its behaviour. Secondly, through a worked-out example, we discussed how these associations could be acquired in the first place, and how the learning of transformations of internal states seems to be prior to the learning of action preconditions that relates to the external state. These last considerations involved some preliminary distictions among different types of internal states. We will now suggest how these different types of internal states can be related to different kinds of concepts, as concrete ideas for future work to be carried out. Along these lines we will also hint at research issues that need further attention.

Consider sensations like feeling hungry or experiencing a certain kinesthetic sensation on the one hand, and experiencing a certain visual scene on the other. Feeling hungry, like feeling a certain movement, depends on a configuration of the body (changes in the body cause changes in the feeling), while experiencing a certain visual scene, or having a certain tactile experience depends on a configuration of the body *and* the environment (changes in the body and/or changes in the environment cause changes in the experience).

We propose therefore to make a distinction between sensations like feeling hungry, feeling pain or kinesthetic sen-

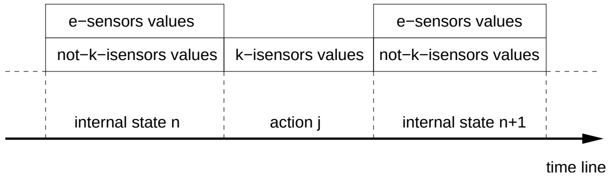

Figure 1: A schematic representation of the stream of experience of an agent.

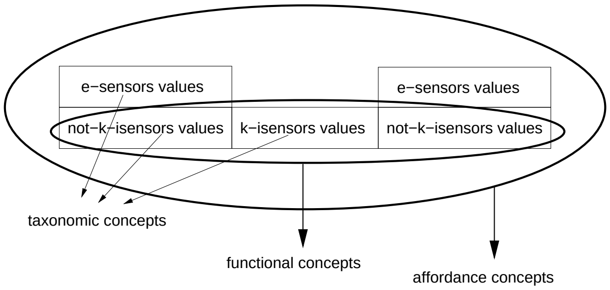

Figure 2: Different types of concepts and their relations with different sensor field values.

sations, and sensations like being-appeared-to-red. The first sensations correspond to values of internal sensors (i-sensors), the second to values of external sensors (e-sensors). Then, among inner sensors, we propose to distinguish between those that provide sensations of movement (k-isensors) and all the others (not-k-isensors). The k-isensors provide sensations that mediate the transformation of one internal state into another, that is: sensations that corresponds to actions in the sense of body movements. An internal state is a combination of values of not-k-isensors and e-sensors, while a combination of values of k-isensors is an action. For sure it is not easy to isolate, by introspection, something like 'only kinesthetic sensations' and we always have the feeling of a continous, ininterrupted stream of experiences, but if the distinctions are accepted, then it is possible to represent the phenomenon of perceiving the external world as a succession of phases such as in Figure 1.

So, the stream of experience is a sequence internal state, action, internal state, action, internal state, etc. As a matter of fact, through time we assist to a continuous and correlated transformation of the fields of e-sensors and not-k-isensors on one hand, and k-isensors on the other. The idea is that clustering upon series of sensor fields of the same type gives rise to taxonomic concepts (similar external states, similar internal states and similar actions), clustering upon series of transformations of not-k-isensors fields by means of k-isensors fields gives rise to functional concepts, while clustering upon series of e-sensors fields associated to the same functional concept gives rise to affordance concepts (see Figure 2.)

According to the notion of conceptual space discussed in [Gärdenfors, 2000], to each sensor corresponds a sensor field which has certain quality dimensions endowed with topological and/or ordering structures. These dimensions corresponds to the different ways stimuli can be judged to be similar or different within that field. Whenever the sensor is active, its sensor field is 'shaped' in a certain way, that is, quality dimensions for that field happen to have certain specific values. An important issue that need to be addressed is therefore determining the field characteristics of each sensor field in Figure 1. A further research issue that needs attention concerns the notion of action and the relations between the sensory-motor system and higher cogntive skills such as categorisation. Our analysis was based on the assumption that the sensory-motor system has the right kind of structure to characterise both sensory-motor and more abstract concepts. For primitive actions such as grasping an initial account is given in [Gallese and Lakoff, 2005], but there is a long way to go to establish what happens when more complex actions are involved.

References

[Cohen et al., 1996] Cohen, P. R., Oates, T., Atkin, M. S., and Beal, C. R. (1996). Building a baby. In *Proceedings of the Cognitive Science Conference*.

[Cohen et al., 2002] Cohen, P. R., Oates, T., Beal, C. R., and Adams, N. (2002). Contentful mental states for robot baby. In *Proceedings of the 18th National Conference on Artificial Intelligence*, Edmonton, Alberta, Canada.

[Davidsson, 1996] Davidsson, P. (1996). *Autonomous agents and the concept of concepts*. PhD thesis, Department of Computer Science, Lund University.

[de Jong, 2000] de Jong, E. D. (2000). *Autonomous formation of concepts and communication*. PhD thesis, Artificial Intelligence Laboratory, Vrije Universiteit Brussel.

[Drescher, 1991] Drescher, G. (1991). *Made Up Minds: A Constructivist Approach to Artificial Intelligence*. MIT Press, Cambridge, Mass.

[Gallese and Lakoff, 2005] Gallese, V. and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 21.

[Gärdenfors, 2000] Gärdenfors, P. (2000). *Conceptual Spaces*. MIT Press, Cambridge, Mass.

[Gibson, 1979] Gibson, J. J. (1979). *The ecological approach to visual perception*. Laurence Erlbaum, Hillsdale, NJ.

[Glenberg, 1997] Glenberg, A. M. (1997). What memory is for. *Behavioral and brain sciences*, 20:1–55.

[Karmiloff-Smith, 1992] Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press, Cambridge, Mass.

[Langley, 1996] Langley, P. (1996). *Elements of machine learning*. Morgan Kaufmann, San Francisco, CA.

[Sloman and Scheutz, 2002] Sloman, A. and Scheutz, M. (2002). A framework for comparing agent architectures. In *UKCI'02: Proceedings of the UK Workshop on Computational Intelligence*, Birmingham, UK.

[Smith and Medin, 1981] Smith, E. and Medin, D. (1981). *Categories and concepts*. Harvard University Press, Cambridge, Mass.