ATTENTION AS A MINIMAL CRITERION OF INTENTIONALITY IN ROBOTS

Lars Kopp and Peter Gärdenfors

Lund University Cognitive Science

Kungshuset

S-222 22 Lund, Sweden

Lars.Kopp@lucs.lu.se, Peter.Gardenfors@lucs.lu.se

Abstract: It is proposed that the capacity of attention is a minimal criterion of intentionality in robots. An attentive system must be able to identify relevant objects in the scene; select one of the identified objects; direct its sensors towards the selected object; and maintain its focus on the selected object. We describe the robot R1 which exhibits behavior that seems intentional. However, when the robot is confronted with a situation where more than one object is present on the scene, the fact that the behavior of R1 is determined merely by S-R rules becomes apparent. We also defend the position that a robot with attention would have a minimal level of intentionality, since the attentional capacity involves a first level of goal representations. This criterion is also useful when discussing intentionality in animal behavior.

1. INTRODUCTION

One of the first situated robots, called Herbert and built by Connell (1989), could stroll around in a MIT department searching for soda cans. The most interesting novel feature of Herbert was its capability to act in an environment that was not specially prepared for robots. The platform of this robot was based on the so called subsumption architecture, which is based on a hierarchical decomposition of behaviors. Brooks (1986) has argued that this kind of *reactive* system is sufficient to model the appropriate behavior for robots in unstructured environments.

The subsumption architecture of a reactive system is able to produce complex behaviors. For example, Herbert was not pre-programmed to perform long complex schemes, but it was instead activating appropriate behaviors according to the state of the environment or the signals that the robot received.

To an outside observer, Herbert looked as if it was a robot with *intentions* – as if it had plans and a purpose with its behavior. But how can we know whether Herbert is intentional or not? What could be a *criterion* for determining this property in relation to the behavior of a robot? Perhaps it is not meaningful to talk about intentionality in robots at all, but this property can only be ascribed to biological systems?

As a background for a discussion of these questions in this paper, we will start by presenting some experiments with a grasping robot, called R1, that has been constructed by the first author. The behavior of R1 is determined by a set of stimulus-response rules (S-R rules). This kind of architecture falls under the subsumption paradigm.[1]

We will use the performance of R1 as a basis for an analysis of when a robot can be said to be intentional. In particular, we will argue that R1 is not intentioal because it has no capacity to *attend* to one object at a time. On a more general level, we will argue that attention is indeed a first level of intentionality in robots (as well as in animals).

2. THE ROBOT R1

2.1 Architecture

The robot R1 (see figure 1) consists of an arm that can move along the x-axis (from left to right and back again). On the arm there is a gripper that can be lowered and raised and rotated ±90 degrees. A

1 Brooks is a bit unclear on exactly what is allowed in subsumption architecture. To be on the safe side, we start by discussin systems based on S-R rules.

Lund University Cognitive Studies - LUCS 89 ISSN 1101 - 8453

peripheral camera is placed on the arm above the gripper. Another camera is placed near the gripper and it rotates with the gripper. In addition to this, there is a conveyor belt that transports various objects (for example, Lego bricks of different sizes and colors) into the visual field of the robot.

The robot is constructed to locate, categorize and grasp a moving object. The location and categorization processes are designed to function independently of the direction the object was placed in on the conveyor belt.

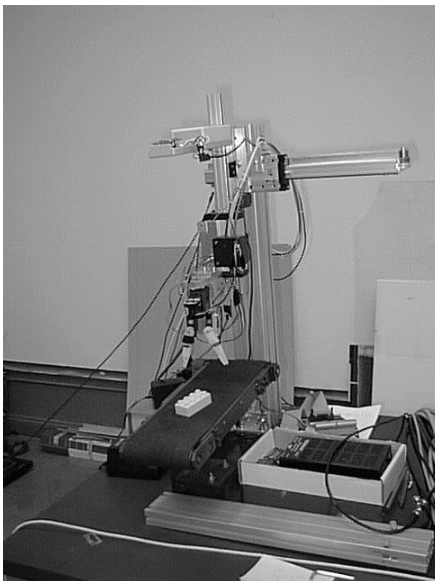

Figure 1. *The reactive grasping robot R1.*

The performance of the robot is based on two independent vision systems: One that handles peripheral vision (input from camera 1) and one that takes care of focus vision (input from camera 2). The electronic hardware architecture of R1 is presented in the appendix.

2.2 Peripheral vision

The purpose of the peripheral vision system (figure 2a) is to locate a moving object on the conveyor belt somewhere in the view of the camera. The system is designed to categorize an object from different angles (see below). It is organized in such way that different views are associated with different responses in a reactive (behavioristic) fashion. When the peripheral vision system finds an appropriate object, it directly moves its arm in the direction of the object. By moving its arm, the robot is also moving the second camera closer to the object.

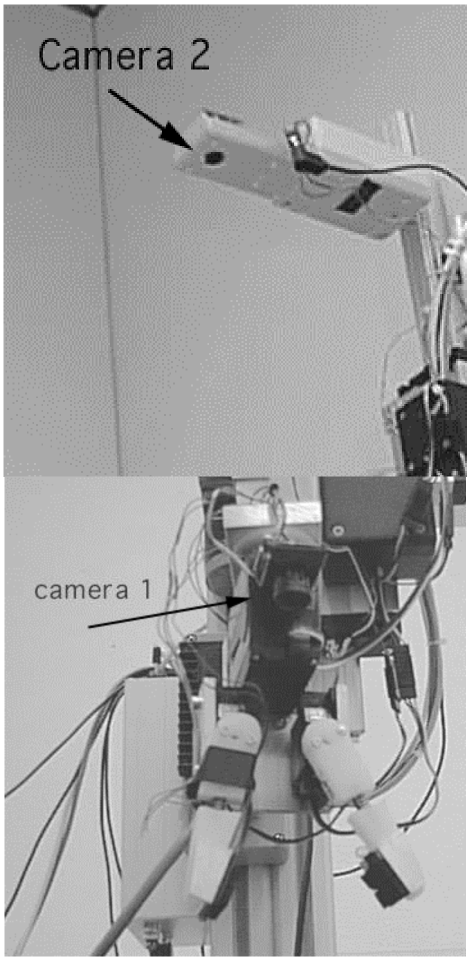

Figure 2a. (Top) *The peripheral camera which is placed on the arm of the robot above the conveyor belt and above the gripper.* ***2b (Bottom)*** *The focus camera is placed near the gripper and is rotated along with the hand.*

By repeating that loop, the peripheral vision system soon reaches its goal, which is that the object will be moved into the center of the image of camera 2. It should be mentioned already here that this stimulus-approach behavior was successful when only one object was present on the scene. The robot was also tested when several objects were present on the scene. In these cases, however, a large number of grasping failures occurred.

2.3 Focus vision

The focus system (figure 2b) is concerned with how an object should be grasped correctly. That system is responding to the orientation of an object. The camera can only see a small region and it only responds when the object is within that region. That also means that it is necessary to place the object close to the hand if the camera of the focus system is to be activated. Basically, this is the task of the peripheral system.

To categorize a certain object from a variety of views, it is necessary that the focus system has a set of stored representations of views of the relevant objects in its internal memory. This knowledge has been implanted by the programmer into the system and is thus not learned by the robot itself. In addition to this, the vision system has stored a set of appropriate stimulus-response behaviors that control the behavior (orientation of the hand and grasping and letting go of an object) of the robot.

2.4 The control of the orientation of the hand-wrist

The appropriate control of the hand-wrist of the robot was more complicated to achieve than controlling the x-position of the arm. The orientation of the moving object is not easily determined. In computer vision applications, constructors often use a rotation matrix to estimate a large numbers of transformations, which have to be matched against the input representation of the object in focus.[2] To be able to perform such a matching, it is necessary to first estimate the gravity-center of the object. It is important to notice that such operations are computationally costly since all such generated representations must be matched with the input representation of the object.

Instead, we solve the problem of determining the orientation of the object by utilising the stored views of an object that were mentioned above (Poggio and Edelman 1990). These views are associated with a certain response, which in this case means a certain angular rotation of the hand (figure 3). When the picture of the object matches one of the stored views, the corresponding response is performed. This kind of procedure is typical of a reactive response system. It is computationally quicker than the procedure presented above – the burden is carried by the vector of stimulus-response pairs. However, it puts higher demands on the memory of the system.

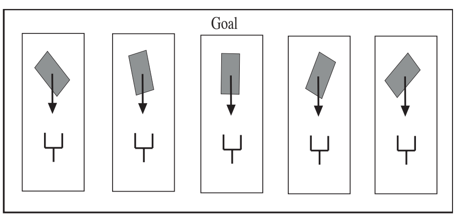

***Figure 3**. Memorized views of a brick. The focus vision has been taught to associate a certain view of a brick with a certain response. In this example five different views have been memorized and are associated with a response to rotate the wrist of the gripper an appropriate angle.*

2.5 Multiple scales

In many categorization tasks, it is not necessary to match the whole image in all its details, but basic shape information is, in general, sufficient for categorization. However, if two or several of the objects are quite similar in shape and form, it is useful to integrate more details in these models. The idea is to stop the categorization process when a sufficient matching is achieved between the object categories.

To achieve a fast responding system, all representations are not stored at the same spatial resolution (figure 4). Such a resolution pyramid of multiple scales has many interesting properties. For example, it is useful when generalizing a class of objects, it results in faster estimations of a reduced picture, and fewer stored views are needed (Balkenius and Kopp 1997a, 1997b).

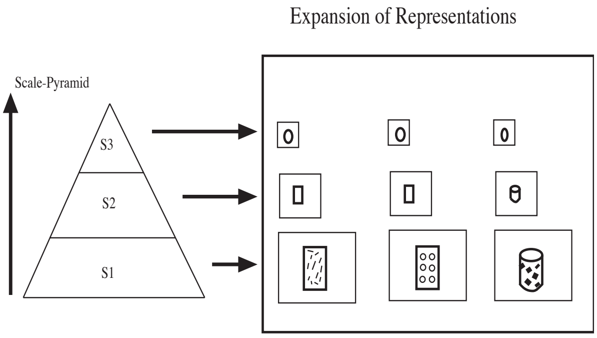

***Figure 4**. An example of expansions of representations involving multiple scales.*

For example, in the experiments with R1, the representation of a brick was stored at different spatial resolutions to achieve a fast responding vision system (this boils down to a reduction of the size of matrix calculations). A brick that was placed at an angle with the direction it was moving in (with a difference in orientation greater than 10 degrees) was stored as a separate schema. The coarseness of the scale makes it possible to rapidly estimate the position of the object. It is not necessary to determine the exact orientation of an object when it is far from the goal orientation. Rather more details (concerning resolution, position and orientation of the object) should be calculated when the object is close enough to the gripper. This approach makes the robot much more adaptive. We also believe that this approach corresponds roughly to the way the brain identifies objects.

2 A rotation matrix represents information about the appearance of an object in different rotational positions.

3. THE PERFORMANCE OF R1

3.1 Actions on a single object

As described in the previous section, R1 was programmed to categorize a number of objects and taught how to orient its hand independently of the position and orientation of the object. Furthermore, pairs of stimuli and responses (small pieces of reactive behaviors) were stored in the memory of R1. None of these behaviors were linked or associated with longer sequences in the stored model of the grasping behavior of the robot.

The scenario for most of the experiments consisted of one object, which was moved rather quickly by the conveyor belt towards the robot arm and the camera that is mounted on the gripper (see figure 5). The velocity, orientation and position of the object was not determined in advance, so the robot had to compensate or adapt to these parameters to perform a successful action sequence of recognizing the object, moving the arm, rotating the gripper, grasping and removing the object from the conveyor belt.

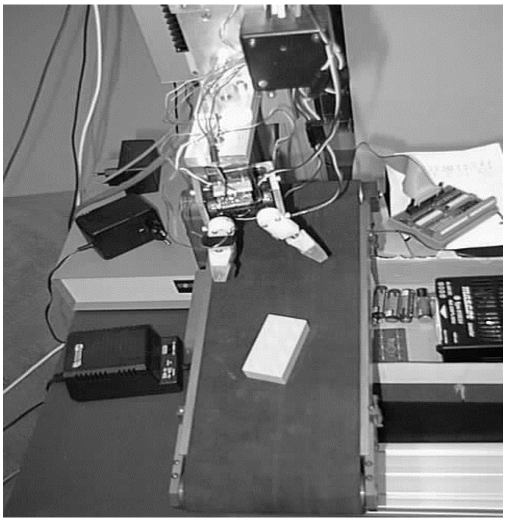

***Figure 5**. Robot R1 in action. The goal of R1 is to grasp bricks independently of position, orientation and speed of the object. A conveyor belt transports the object towards the arm and hand of R1.*

In the experimental interaction, the robot was able to co-ordinate the start and finish states of separate actions and in this way generate a complete sequence which ended in a successful grasping and removal of the object. For example, when the object was close to the gripper and the camera with the gripper had the correct orientation, one of the stored stimulus-response pairs was activated which resulted in a grasping behavior. The grasping response or behavior was an end state of a sequence of previous reactive behaviors – a sequence that, to a large extent, was unpredictable. In a satisfactory number of situations, the reactive system of R1 managed to grasp and remove the moving object. The interaction between the different parts of the systems takes place by ”communication through the world” as Brooks (1991) calls it. That is, one part of the system is activated when another part has changed the environment in a particular way. For example, the robot Herbert, menitioned in the introduction, only moved towards the trash bin when it had succeeded to grasp a soda can.

However, the success of the reactive system was to a large extent depending on the fact that there was only one object present on the scene. When two objects were present, however, the robot did often not succeed in grasping any of the objects. The problem for a system that builds on reactive behavior is to focus on only one object, if several are present on the scene. We will return to the crucial role of attention for solving this kind of problem in section 4.

3.2 The interaction between perception and action

In traditional AI, robots are pre-programmed to perform certain behaviors under certain conditions. In such an approach all situations that can be problematic for the robot must be foreseen by the programmer. However, in complex unstructured environments, such programs will often fail because it is impossible to predict all problematic situations. As a background to a solution of this kind problem, the following points from Gibson’s (1979) theory of cognition are relevant:

(1) Perception and action form an integrated system. Perception guides action; and action provides opportunities for perception. Human activity can be fruitfully described in terms of perception-action systems.

(2) We perceive a meaningful world. This meaningfulness resides neither in the agent (person) nor in the environment, but derives from the interaction of the two. Gibson described meaning in terms of *affordances*. For example, some surfaces offer support for walking, whereas others do not. What the environment affords depends on what kind of animal you are, which involves not only your physical receptors and effectors but also your level of cognitive and behavioral development.

These two principles are supported by the subsumption architecture proposed by Brooks (1986) (also see Connell (1990)). Gibson’s thesis (1) has also been an inspiration in the construction of the robot R1. The robot was not only built to show that affordances can be useful in robotics, but also to bring forward problems that frequently occur in connection with behavior based robots. In (1), Gibson focuses on the importance of interactions between actions and perceptions. The point is that such interactions can facilitate the performance of the robot and govern the selection of the next S-R pair. In particular, such interactions are important when a robot tries to grasp a certain object. The interactions simply facilitate the opportunities for the next action to happen.

The interaction between action and perception can be exemplified with the behavior of R1 (see figure 5). When the robot tries to grasp a brick at a certain location and orientation, the object is first located with the aid of the peripheral vision system. That system then estimates a distance-error signal that is used to decide how much the arm should be moved in the x-direction (only) in order to approach the object. The object is constantly moving towards the arm and hand (the object is transported in the y-direction with aid of the conveyor belt). The success of the first action system creates the opportunity for the focus vision system to perform a long sequence of appropriate actions. The implicit purpose of these actions is to finally reach and grasp the object.

In summary, the two systems that are electronically completely separated were able to produce the complex behavior of grasping an object in motion. This shows that a form of "emergent" behavior can appear as a consequence of external interactions. For example it is impossible for the focus vision system to stick to the approaching object and control the rotation of the hand, if the control of arm location in x-direction is not properly managed by the peripheral vision system.

What has been shown by the construction of R1 is that a simple set of S-R rules can result in a kind of "quasi-intentional" behavior. The question to be addressed now is when a robot really becomes intentional. In particular, what should be added to R1, if anything, to make it an intentional system?

4. Attention and Intentionality in Artificial Systems?

4.1 The problem with lack of attention in a reactive system

When we are observing the action of a typical grasping situation, the robot R1 appears to be intentional. This impression is particularly strong when you interact with the object by moving it, since then the robot arm and the gripping mechanism are adjusting to the movement. This seems like an adaptive and intentional behavior. But since we know that the system is merely reactive and does not have any representation of objects, plans or goals, there is nothing intentional built into the robot.

The robot R1 is interacting the world by using merely a set of S-R pairs. It can thus be viewed as a behavioristic system, where "mental" notions like intentionality are not needed to describe its function. R1 is nevertheless able to perform functional actions like locating and grasping a moving object. The architecture of the robot, however, admits no internal representations of goals that can be used for reasoning and planning in such a system.

So, why would a grasping robot need representations that enable, for example, planning of actions? R1 seems to function well anyway. The limitations of the S-R architecture of R1 showed up, as was mentioned above, when there was more than one object present on the scene. In such a situation, the robot would seemingly shift it's "attention" randomly between the different objects, which would lead to inconsequent actions that resulted in no object being grasped. A more accurate description is that R1 had no capacity at all to *attend* to an object. We will argue that such a capacity makes a crucial difference for the performance of a system.

4.2 When is a system intentional?

Before we turn to what should be required of an attentional system, we first comment on what properties a system must have to be intentional. We do not believe that there is a unique answer to the question, since a subject can exhibit different levels of intentionality.[3] A first condition an intentional subject must satisfy is that it should be capable of having certain kinds of representation.

Some kinds of animal behavior, like phototaxis, is determined directly by psychophysical mechanisms that transduce information about the environment. In such cases, representations are not involved at all. The actions that follow on the transduction are mere reflexes that connect the signals received by the animal with its behavior. Such biological mechanisms corresponds to reactive systems in the robot domain.

In other cases, animals and robots use the incoming information as cues to "perceptual inferences," which add information to what is obtained by the psychophysical receptors. Whenever information is added in this way to sensory input representations are obtained. Representations can be seen as intermediate variables that connect stimuli and responses and thereby reduce the number of links required. The intermediate variables also make the behaviour of the system more flexible.

For example, von Uexküll (1985, pp. 233-234) argues that as soon as an animal can map the spatial

3 For a discussion of some of the different levels, see Brinck and Gärdenfors (1999).

structure of its environment by a corresponding spatial organization of its nervous system, the animal constructs

> a new world of excitation originating in the central nervous system that is erected between the environment and the motor nervous system. […] The animal no longer flees from the stimuli that the enemy sends to him, but rather from the mirrored image of the enemy that originates in a mirrored world.

We submit that the capacity to represent the world is a necessary condition for intentionality. Von Uexküll (1985, p. 231) expresses the difference between animals capable of representation from those not capable of it in the following drastic way: "When a dog runs, the animal moves its legs. When a sea urchin runs, the legs move the animal." In brief, intentionality is necessary for agenthood.

Representations are necessary for planning, reasoning and rational behavior in general.[4] In particular, representations of the *goals* of the system are central for achieving intentionality. This is what gives a system a "directedness" (Brentano 1973). It should be noted that we do not require that representations are expressed by some form of *symbolic notation* in the system. There are many other ways of making systems, natural as well as artificial, representational (see e.g. Kirsh (1991)).

The advantage of intentionality in a system is that it is able to *adopt its actions* to the situation at hand (Tomasello and Call 1997). To be reached, the same goal may require different actions in different contexts. And the same actions may afford different goals in different contexts. Due to its use of representations as mediating terms, intentional behaviour becomess flexible. It depends on the ability of the system to adjust to the distinctive character of each context it encounters. It also depends on the ability to learn about new contexts and how to represent them.

In our discussion of the robot R1, we have been focusing on S-R systems. In such systems, there is no way to *represent* a goal. The perceptions and actions that form the elements of the S-R pairs does not allow a "goal" to creep in on either side of such a pair.

However, other kinds of non-symbolic architectures have been proposed for robotic systems. One of the most well-known is the *subsumption architecture* proposed by Brooks (1986) (see also Connell 1990) Systems based on the subsumption architecture are more advanced than S-R systems since, firstly, there may be internal links between several links of the S-R type (in other words, internal "responses" may function as "stimuli" for other links); and, secondly, such a system may contain internal control loop, which improves the goal-directedness of the system. Nevertheless, such a system does not contain any internal representations of the goals of the system. As a matter of fact, Brooks (1991) argued in an early paper that robotic systems do not need representations.[5] Hence systems based on the subsumption architecture do not satisfy the criterion of intentionality as presented above.[6]

4.3 What is required of an attentional system

What is then required of a visual robot if it is to be able to attend to something? First of all it should be noted that there are different levels of attention. Brinck (2001) distinguishes between scanning, attention attraction and attention focusing. *Scanning* is the continuous surveying of the environment which is directed at discovering possibilities to act. *Attention attraction* occurs when something happens that is at odds with the expectations of the system. This kind of attention is triggered by events in the environment. *Attention focusing*, finally, is the intentional form of attention where the agent itself choses what to attend to. In this case the agent may even have its attention directed to something that does not exist in the current environment. For example, if you are looking for chanterelles in the forest, your attention is focused on yellow mushrooms, even if there is not a single chanterelle in the environment. In the current paper, we are only concerned with attention focusing.

We submit that a visual robot that is capable of attention focusing must be able to

(1) *identify* relevant objects in the scene;

(2) *select* one of the identified objects;

(3) *direct* its sensors towards the selected object; and

(4) *maintain* its focus on the selected object.

A special case of (4) is that the robot should be able to *track* an object, that is, focus on the object even if it moves across the scene. Since we see attention focusing as a minimal form of intention, it should be noted that the four criteria proposed here fit well with Cohen and Levesque's (1990) proposal that intention is choice with committment.[7] The capacity to follow an object over time and over varying perceptual circumstances is what, following Piaget, is called

4 For a general discussion of representations in animals, see Roitblat (1982), Gopnik (1982), Lachman and Lachman (1982), Gulz (1991), and Gärdenfors (1996a, 1996b).

5 This position was strongly criticized by Kirsh (1991).

6 At least not "classical" systems based on the subsumption architecture. Brooks has included further aspects in later systems, e.g. in the COG robot, that may make them pass the criterion.

7 We agree with the basic idea of the paper, although we believe that the logical formalism chosen by the Cohen and Levesque is not appropriate for the problem.

object permanence in psychology. Thus there is a close connection between a system exhibiting object permanence and the system's attentional capacities.

Capacities (1) and (2) demand *representations* of objects which is a radical departure from behavioristic principles. The point is that no set of S-R couplings is sufficient for identifying and selecting an object in a changing environment. For this the robot needs a way of *internally marking* a set of features as being characteristic of a particular object. It is such a set that constitutes the representation of the object.[8]

This kind of representation is an unevitable control variable in the attention mechanism of the robot. The representation "binds together" the S-R pairs of a reactive system and makes them cooperate in new ways. In other words, the perception-action interaction is enhanced by the perception-representation-action combinations that are made possible by adding, for example, object representations. The representation acts as a "hidden variable" in the control mechanisms of the system.

Another aspect of (2) is how it is decided *what* the system should attend to. Here the goals of the system are of course the fundamental driving forces. But often a system has several, often conflicting goals. Thus a mechanism for prioritizing the goals is needed. Such a system, which we will not discuss here (but see Balkenius 1995 ch. 6, and Rolls 1999) provides the systems with its *motivation*. The motivation of the system then determines what should be attended to – in a sense, it determines the current value of different objects. For example, if hunger is your strongest motivation when you are in an unknown town, you will attend to restaurants and other places where food is provided, but if you are tired you will attend to hotels and houses providing bed and breakfast. As a matter of fact, the very act of choosing one object may increase the motivation of the system to attend to the object.[9]

An example of a reactive high-level control system is the robot Vision Car developed by Newton Research Labs. The robot is able to capture balls that move randomly over the floor and deliver the balls at a goal position. The balls have different colors which represent different values for the robot. The robot is programmed to find and approach static and moving balls and to attend to the best target.

The control system of the Vision Car has four basic states: (1) Find and approach a ball; (2) lift the ball; (3) find and approach the goal; and (4) drop the ball. The system that visually classifies and evaluates the static and moving balls consists of a preprogrammed strategy. The motivation of the robot derives from this internal routine.

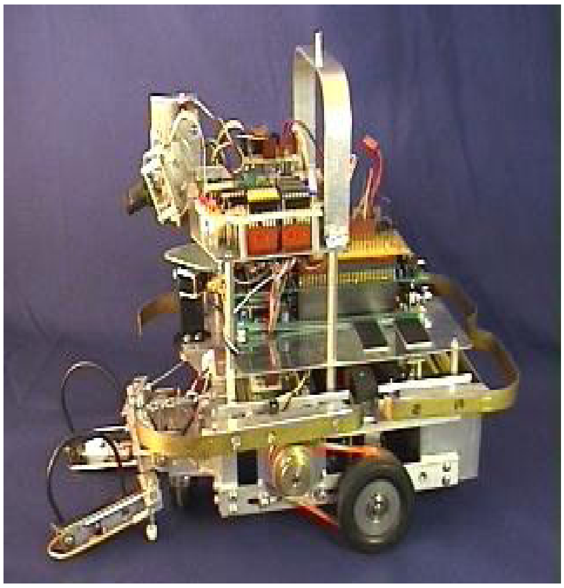

Figure 6. The Vision Car developed by Newton Research Labs

The robot showed opportunistic behavior in the sense that it had the competence to direct its attention to other targets while it was holding a ball (on its way to the goal position). The robot could drop the ball it was carrying and approach another ball of a higher value. This kind of behavior requires a value system that is used by the motivation system to attend and lock on a certain object. In this example, attention is used to lock on a target (a ball) and then behave appropriately. However, a meta-system seems to be involved here, since the robot is able to drop the ball it already has in its grip. The machine is thus able evaluate a new target, while it is carrying another towards the target. This kind of opportunistic behavior is a sign that the Vision Car approaches the intentional level in its behavior. The exact judgment of what level of intentionality it achieves depends on the flexibility of its attentive system in relation to the criteria (1) – (4) presented above.

4.4 Attention as a criterion of intentionality for robots

Our analysis of the functions of attentive mechanism suggests the that a basic architecture for an attentional robot consists of the following components: a reactive

8The set of features can take various forms: it could, for example, be just a name of the object; or it could be an equilibrium point in the activities of an artificial neuron network.

9 See McFarland and Bösser (1993), section 6.4 for a modelling in terms of utilities of this seemingly paradoxical phenomenon.

system, a value component, a selection system, and an attentive system.

As argued above, an attentive system satisfying criteria (1) – (4) presumes that the system has “object permanence“ which in turn presumes that the system has representations of objects that are, to some extent, independent of the perceptual input and the actions performed on the object.

We now propose that these capacities are sufficient to give the system a minimal form of intentionality. When a system attends to an object, the motivation of the system is the force that makes the system keep track of the object and to gather further information about what happens to it. The striving for these goals is what makes the system intentional.

Then, of course, the basic goal of the attentive process can be embedded in higher level system goals that would be connected with more advanced forms of intentionality. But the point we want to make is that merely attending to an object already results in an intentional system. Furthermore, the higher level goals of a more advanced intentional system that, for example, acts on a certain object depends for its success on the performance of an attentive system that can keep track of the relevant object.

The fundamental point is that it is the representation of the object and its potential motivational value that determines the goal of the system. This internally represented goal controls the behavior of the system rather than just the perceptual input as in a S-R system. The perceptions of the system may change the representation of the object, and thereby the behavior of the system with respect to the object, but this is a secondary effect. For example, because the representation “glues together“ the various perceptions of an object, a system with representations will be much less sensitive to disturbances in the perceptions of an object than a S-R system is. In consequence, the erratic behavior of the robot R1 when it was confronted with more than one object on the scene will not be present in a goal-directed attentive robot.

Another way of describing the fundamental difference between an attentive system based on representations and a reactive S-R system is that an attentive system can have *expectations* about the world, but this cannot be achieved by a pure S-R system. For example, the representations of an attentive robot makes it “expect“ that an object will have a comparatively stable shape, that it will move in a continuous manner in space, and that the object will continue to exist even if it is temporarily occluded from the visual field of the robot. A system that only contains S-R pairs, consisting of perception-action couplings, will not be able to handle the “it“ that comes with the representation of an object, let alone make any predictions that depends on the spatio-temporal continuity of an object.

4.5 Object representation and attention in animals

In this context, it is interesting to compare with the representational capacities of different species of animals. Mammals (and birds) exhibit object permanence, but reptiles don’t. In order to illustrate the behavioral differences this leads to, we present an example borrowed from Sjölander (1993, pp. 3-4) comparing how snakes and cats hunt. It seems that a snake does not have a central representation of a mouse but relies solely on perception-action (S-R) couplings. The snake exploits three different sensory systems in relation to prey, like a mouse. To strike the mouse, the snake uses its *visual* system (or thermal sensors). When struck, the mouse normally does not die immediately, but runs away for some distance. To locate the mouse, once the prey has been struck, the snake uses its sense of *smell*. The search behaviour is exclusively wired to this modality. Even if the mouse happens to die right in front of the eyes of the snake, it will still follow the smell trace of the mouse in order to find it. Finally, after the mouse has been located, the snake must find its head in order to swallow it. This could obviously be done with the aid of smell or sight, but in snakes this process uses only *tactile* information. Thus the snake uses three separate modalities to catch and eat a mouse. Since there is no communication between the three sensory systems (except that one takes over when the other finishes), it has no central *representation* of a mouse.

In comparison, the cat is able to represent objects, which among other things leads to object permanence. When the cat hunts, it relies on a combination of information from several sensors: eyes, ears, nose, paws, and whiskers. It can predict that the mouse will appear at the other side of a curtain when it disappears on one side. It can “infer“ information about the mouse even if there is no immediate sensory information, like when it is waiting outside a mouse-hole. In this sense it has a central representation of a mouse that is, at least to some extent, independent of the perceptual information.

Of course, the ability to represent will also affect the attentional capacities of animals. As every cat owner knows, the cat has no problem in intensively attending to a mouse during the hunt, even when there are several disturbing factors. In contrast, we conjecture that the snake would have severe problems if there, for example, were more than one mouse present on the scene (although we have no empirical evidence for this). If the snake, for instance, had struck one mouse, but happened to follow the smell track of another unhurt mouse, it would fail miserably in its hunting.

5. CONCLUSION

In this paper we have suggested that the capacity of attention is a minimal criterion of intentionality in robots. We submit that an attentive system must be able to identify relevant objects in the scene; select one of the identified objects; direct its sensors towards the selected object; and maintain its focus on the selected object.

We have described the robot R1 which exhibits behavior that at a first glance may seem intentional. However, when the robot is confronted with a situation where more then object is present on the scene, the fact that the behavior of R1 is determined merely by S-R rules becomes apparent. In brief, the robot has problems attending to a specific object.

We have also defended that position that a robot with attention would have a minimal level of intentionality, since the attentional capacity involves a first level of goal representations. This criterion also seems to be useful when discussing intentionality in animal behavior.

ACKNOWLEDGEMENTS

We wish to thank Christian Balkenius, Ingar Brinck, Antonio Chella and Shimon Edelman for helpful discussions.

REFERENCES

Balkenius, C., (1995): *Natural Intelligence in Artificial Creatures*, Lund University Cognitive Studies 37, Lund.

Balkenius, C., Kopp, L. (1997a): "Elastic template matching as a basis for visual landmark recognition and spatial navigation". In Nehmzow, U., Sharkey, N. (Eds.) *Proceedings of AISB workshop on "Spatial reasoning in mobile robots and animals",* Technical Report Series, Department of Computer Science, Report number UMCS-97-4-1. Manchester University, Manchester.

Balkenius, C., Kopp, L. (1997b): "Robust self-localization using elastic template matching", in Lindeberg, T. (Ed.) *Proceedings SSAB '97: Swedish Symposium on Image Analysis 1997*. Computational Vision and Active Perception Laboratory, KTH, Stockholm.

Brentano, F. (1973): *Psychology from an Empirical Standpoint*, Routledge and Kegan Paul, London.

Brinck, I. (2001): "Attention and the evolution of intentional communication", *Lund University Cognitive Studies 82*, Lund.

Brinck, I. and Gärdenfors, P. (1999): "Representation and self-awareness in intentional agents", *Synthese 118*, pp. 89-104.

Brooks, R. A. (1986): "A robust layered control system for a mobile robot", *IEEE Journal of Robotics and Automation, 2(1):*14-22, April 1986.

Brooks, R. A. (1991). "Intelligence without representation", *Artificial Intelligence 47*, pp. 139-159.

Cohen and Levesque, H. (1990): "Intention is choice with committment", *Artificial Intelligence 42*, pp. 213-261.

Connell, J. (1989): "Behaviour-based arm controller", *IEEE Journal of Robotics and Automation, 5(6),* December 1989, 784-791

Connell, J. (1990): *Minimalist Mobile Robots: A Colony Architecture for an Artificial Creature*, Academic Press, Boston, MA.

Gärdenfors, P. (1996a): "Cued and detached representations in animal cognition", *Behavioural Processes 36*, pp. 263-273.

Gärdenfors, P. 1996b): "Language and the evolution of cognition", in *Penser l'esprit: Des sciences de la cognition à une philosophie cognitive*. V. Rialle and D. Fisette, editors. Presses Universitaires de Grenoble, Grenoble, 151-172.

Gibson, J. J. (1979): *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA.

Gopnik, A. (1982): "Some distinctions among representations", *Behavioral and Brain Sciences 5*, 378-379.

Gulz, A. (1991): *The Planning of Action as a Cognitive and Biological Phenomenon*, Lund University Cognitive Studies 2, Lund.

Kirsh, D. (1991): "Today the earwig, tomorrow man?", *Artificial Intelligence 47*, pp. 161-184.

Lachman, R. and Lachman, J. L. (1982): "Memory representations in animals: Some metatheoretical issues," *Behavioral and Brain Sciences 5*, pp. 380-381.

McFarland, D. and Bösser, T (1993): *Intelligent Behavior in Animals and Robots*, MIT Press, Cambridge, MA.

Poggio, T. and Edelman, S. (1990): "A network that learns to recognize threedimensional objects", *Nature 343*, pp. 263-266.

Roitblat, H. L. (1982): "The meaning of representation in animal memory," *Behavioral and Brain Sciences 5*, pp. 353-372.

Rolls, E. T. (1999): *The Brain and Emotion*, Oxford University Press, Oxford.

Sjölander, S. (1993): "Some cognitive breakthroughs in the evolution of cognition and consciousness, and their impact on the biology of language," *Evolution and Cognition 3*, pp. 1-10.

Tomasello, M. and Call, J. (1997): *Primate Cognition*. Oxford University Press: New York, NY.

Von Uexküll (1985): "Environment and inner world of animals", in *Foundations of Comparative Ethology*, G. M. Burghardt, editor. Van Nostrand Reinhold Company, New York, 222-245.

APPENDIX

Hardware description of R1

R1 was constructed using industrial mechanical components. The computers are standard PCs. The frame grabber is an inexpensive construction by the first author. The cameras are cheap black-and-white and the motion control consists of standard radio controlled servos. A one stepper motor is used to move the arm in the x-direction. Figure 7 shows how the different computers and electronics are connected.

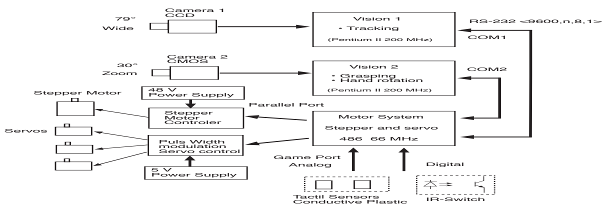

Figure 7. *The hardware of R1*.

Design of the robot hand

The robot-hand has the following sensors: an IR-light switch and two tactile sensors (made of resistive plastic and with variable resistance proportional to the force on the sensor area). Figure 8 illustrates the hand sensors.

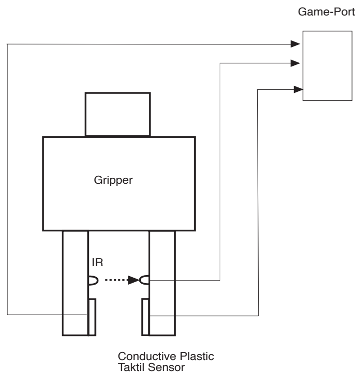

Figure 8. *Gripper with tactile sensor and an IR-switch*.

When the robot prepares to grasp an object it focuses on the light bar in order to know when its time to stop the lowering of the arm. When the light beam is broken the hand is assumed to be at the correct height to grasp the object. The “fingers“ of the gripper can never touch the conveyor band because of physical constraints in the construction. The grasping behavior depends on several kinds of information: the orientation of the hand, the distance (height) to the object and the appropriate force for grasping. The tactile sensors measure the force of the grasping. A servo-loop is continually measuring and controlling the grasping force applied to the object.

A color picture of the robot can be seen at http://www.lucs.lu.se/Projects/Robot.Projects/Snatte.html.