

REPRESENTATION AND SELF-AWARENESS IN INTENTIONAL
AGENTS

ABSTRACT. Several conditions for being an intrinsically intentional agent are put forward. On a first level of intentionality the agent has representations. Two kinds are described: cued and detached. An agent with both kinds is able to represent both what is prompted by the context and what is absent from it. An intermediate level of intentionality is achieved by having an inner world, that is, a coherent system of detached representations that model the world. The inner world is used, e.g., for conditional and counterfactual thinking. Contextual or indexical representations are necessary in order that the inner world relates to the actual external world and thus can be used as a basis for action. To have full-blown intentionality, the agent should also have a detached self-awareness, that is, be able to entertain self-representations that are independent of the context.

1. LEVELS OF REPRESENTATIONS

The question in focus in this paper is: What properties must a subject (an organism or possibly a computer) have in order to be intentional? Before we can answer this question, we need a working definition of intentionality. In everyday parlance, we call those subjects intentional whose behavior can be predicted and explained with the help of a folk-psychological vocabulary, i.e., by ascribing the subject states like belief, desire, etc., and taking these states as either reasons for, or causes of, that behavior.¹ Let us initially take this as a criterion for being an intentional subject.

Now, the question is what properties a subject must have to be intentional in the manner described by the criterion. We do not believe that there is a unique answer to this question, since a subject can exhibit different levels of intentionality. Below we will put forward several conditions of intentionality. The level of intentionality of a subject will depend on which of these conditions it fulfils.

A first condition an intentional subject must satisfy is that it should be capable of having certain kinds of *representation*. Representations are necessary for planning, reasoning, and rational behavior in general. In this section, we want to present a classification of the different kinds of representations that one finds in biological systems.²



Some kinds of animal behavior, like phototaxis, are determined directly by psychophysical mechanisms that transduce information about the environment. In such cases, representations are not involved at all. The actions that follow transduction are mere *reflexes* that connect the signals received by the animal to its behavior.

In other cases, animals use the incoming information as cues to “perceptual inferences”, which add information to what is obtained via the psychophysical receptors. Whenever information is added in this way to sensory input *representations* are obtained.³ For example, von Uexküll (1985, 233–234) argues that as soon as an animal can map the spatial 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 *sine qua non* for intentionality. Von Uexküll (1985, 231) expresses the difference between animals capable of representation and 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.”

We view *categorization* as a special case of representation. When, for example, a bird not only sees a particular object, but sees it *as* food, the bird’s brain is adding information about the perceived object that, for instance, leads to the bird’s swallowing the object. Since information is added, *mistakes* become possible. A mistake is made when the behavioral conclusions drawn from the categorization turn out to be disadvantageous to the animal.

For our analysis of the different levels of intentionality, we need to distinguish between two kinds of representation, namely, *cued* and *detached*. A *cued* representation stands for something that is present (in time and space) in the current external situation of the representing organism. Say that a chicken sees a silhouette of a particular shape in the sky and perceives it as a hovering hawk. The chicken has then used the perceptual stimuli as a cue for its hawk representation. Most cases of categorization are instances of cued representations.

An advanced form of cued representation is what Piaget calls “object permanence”. A cat can, for example, predict that a 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 (see Sjölander 1993).

The representation is nevertheless prompted by the presence of the mouse in the actual context.

In contrast, *detached* representations stand for objects or events that are not necessarily present in the current situation. In other words, such representations are context-independent. A representation of a phenomenon that happens to be present is also detached if the representation could be active even if the phenomenon *had not been* present. This means that sensory input is not required to evoke a detached representation, instead the subject generates the information by itself.⁴

For an example of a detached representation, consider the searching behavior of rats. This behavior is best explained if it is assumed that the rats have some form of “spatial maps” in their heads. The maps involve detached representations because the rat can, for instance, represent the location of the goal even when it is distant from the rat’s present location. Evidence for this, based on the rat’s abilities to find optimal paths in mazes, was collected by Tolman as early as the 1930s (see Tolman 1948). However, his results were swept under the carpet for many years, since they were clear anomalies for the behaviorist paradigm.⁵

It is useful to make a further division within the class of detached representations. Sometimes the representation is *dependent* on an external referent, although the referent does not have to be present in the subject’s immediate surroundings. This is the case with the spatial maps as in the example above. The other sub-class of detached representations are those that are completely *independent* of an external referent in the environment (see Gulz 1991; Gärdenfors 1996b). Say that a chimpanzee walks away from a termite hill to break a twig. It does so in order to peel the leaves off to make a stick that can be used to catch termites. In this case, the animal has a referent-independent representation of a stick and its use. The representation of the stick has not been triggered by the presence of a stick in the environment – the chimpanzee may not even be able to find a twig to make one. In the case of humans, a fantasy about an object that does not exist or a situation that has never occurred are even clearer examples of referent-independent representations.

The distinction between referent-dependent and referent-independent detached representations thus concerns the origin of the representations: whether they could occur without the subject that entertains them ever having met with the phenomena that the representations are about. In this sense, cued representations are all referent-dependent. In the following, we hope to show that the distinctions between the major kinds of representation are instrumental in that they direct our attention to key features of the representational forms and thence to different types of intentionality.

2. THE INNER WORLD

As mentioned above, von Uexküll (1985, 233–234) argues that animals capable of representation have “a new world of excitation [...] that is erected between the environment and the motor nervous system”. He calls this new world the “counterworld” of the animal.

The environment as reflected in the counterworld of the animal is always a part of the animal itself, constructed by *its* organization, and processed into an indissoluble whole with the animal itself (1985, 234).

Von Uexküll refers to a mirrored world and not to the external world when he talks about representation. This should not lead one to think that the animal does not interact with or perceive the external world itself. Rather, as we understand it, the counterworld mediates between perceptions and actions. Perception is necessary for the emergence of the counterworld. The representations of the counterworld are “tools of the brain determined by its plan of organization. These tools always stand ready to become active in response to appropriate stimuli from the external world” (1985, 234).

Von Uexküll’s counterworld contains both cued and detached representations. However, in putting forward a second condition of intentionality we want to focus on the role of detached representations. The role of such representations in the mental life of an organism can be explained by relating it to an idea introduced by Craik (1943, 61):

If the organism carries a “small-scale model” of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which are the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react on a much fuller, safer and more competent manner to the emergencies which face it.

We define the *inner world* of an organism as the collection of all the detached representations of the organism.⁶ Loosely speaking, the inner world consists of all the things the organism can actively “think” about in addition to what is given by the cued representations. The inner world constitutes an intermediary between perception and action, where the detached representations are systematically interrelated and provide the subject with a coherent model of the external world. It is as such a model that the inner world is instrumental to intentionality.

The inner world thus consists in representations of objects (like food and predators), places (where food or shelter can be found), actions and their consequences, etc., even when these things are not present in the environment. Accordingly, Jeannerod (1994, 2) says that “actions are driven by an internally represented goal rather than directly by the external

world". The external world does not impose a behavior on the subject, but instead the subject, by behaving in a goal-directed way, imposes a structure or an order on the external world. This structure is reflected in the inner world.

By exploiting its inner world, the animal can *simulate* a number of different actions in order to "see" their consequences and evaluate them. After these simulations are done, it can choose to perform the most appropriate action in the external world. An animal with cued representations can only rely on *trial-and-error* behavior when trying to solve a problem. One of the main evolutionary advantages of an inner world is therefore that it frees an animal who is seeking a solution to a problem from such dangerous behavior.

Of course, the *success* of the simulations depends on how well the inner world is matched to the perceptions of the external one – a monkey who imagines a branch where there is none is soon a dead monkey. Evolutionary selection pressures have led to a strong correspondence between the perceived world and the simulated inner world of organisms. However, this does not guarantee that an organism will never make any mistakes.

The inner world of a subject must form a *unity* or the subject would not be an *agent*. Different subjects have different inner worlds and they act on the representations that constitute their own world. But what is it that unites one world and distinguishes it from another? Is it the existence of some sort of center that controls the representations and the way they are used, or is it a property of the representations themselves?

We think that it is a mistake to assume the existence of a central control unit of the inner world. First, we want to avoid positing a unifying factor, like a self, if there is a possibility of making do without one.⁷ More importantly, we believe that unity is a property of the inner world as such and not something that is imposed on the world by an external element. Our model does not require a control element, as it were, a ghost in the machine, that surveys the operations of the inner world.

It would be preferable if unity could be explained by reference to the representations themselves. One way to do so might be to describe the inner world, or consciousness, as a self-regulating control system. The idea would be that unity arises as an emergent property of mutually interacting representations. However, this suggestion involves the peculiar idea that the representations themselves interact, while it seems more natural to say that representations do not operate on their own, but are put to use by an organism. Representations are not only *about* something, they are also *for* somebody, in the same way as a tool is made for or used by somebody.

The inner world is a tool that helps the organism to find its way through the world.

Unity can instead be explained by saying that the representations, or rather, the inner world composed of them, owe their existence to a complex of different elements, each contributing to the overall functioning of the whole organism, and that these elements together guarantee the unity of consciousness. If any of them malfunctions, unity is threatened. The different elements that we have in mind are the functional units of the brain together with the perceptual apparatus that feeds information to these units. The inner world will then emerge from these elements. Thus, the unity of consciousness does not supervene exclusively on the brain, but on the functional unity of the organism as situated in the environment.⁸

The functional unity of the organism arises from the parts of the brain taken together with those other parts of the organism that are necessary for perception and action. For instance, perception presupposes that the organism has the means to interact with the environment. Perception is *active* in the sense that the agent does not take the input as something provided by an independent unit. In contrast, the agent actively seeks, by various mechanisms of attention, the perceptions that are most relevant to the problem at hand.

This view of how perception takes place can be compared with Merleau-Ponty's (1962) conception of motility. He writes that

bodily space and external space form a practical system, the first being the background against which the object as the goal of our action may stand out

and that

movement is not limited to submitting passively to space and time, it actively assumes them, it takes them up in their basic significance which is obscured in the commonplaceness of established situations (1962, 102).

Merleau-Ponty thus considers perception as an ongoing activity in which the environment becomes significant to the subject.

As an example of this kind of functional model, Luria (1973) distinguishes between three functional units within the brain: one which provides a basic state of arousal, one which analyses and synthesizes information, and, finally, one which organizes and controls action and reasoning. The three units work concertedly. They are not localised at different areas of the brain but should instead be understood as composing different levels of activity of the whole brain. Normal functioning of the brain thus demands co-operation of all centers of the brain and cannot be localized at separate areas.

Luria (1973, 39) conceives of mental activity as a kind of self-regulating functional system. He writes that

each area of the brain concerned in this functional system [of mental activity] introduces its own particular factor essential to its performance, and removal of this factor makes the normal performance of this functional system impossible.

Lesions of the brain thus threaten the unity of consciousness, a fact which is evident from all kinds of brain damage. Luria's view of the workings of the brain supports the thesis that unity of the inner world emerges from the complex interaction of different units of the brain. It cannot be found within a specific area. On the contrary, it depends on the activity of many different areas.

3. INDEXICAL SELF-AWARENESS

A system consisting exclusively of detached representations cannot be used for reasoning about actual events or for planning actions. The reason is that detached representations, as used in reasoning are not related to specific contexts in the external world.⁹ We take it as a third condition for intentionality that the subject is capable of entertaining *indexical* representations in addition to detached ones. If the subject cannot do so, it will not be possible for us to ascribe an intentional behavior to it. The reason why the subject will not exhibit intentionality is that intentionality discloses itself in action. If the subject cannot entertain indexical representations, it will not have the capacity to act. Thus the subject will not fit our initial criterion of intentionality (see Section 1).

Indexical representations rely on an indexical relation to what they represent. Such a relation is characterized by a contiguity in time and space and/or a causal relation between the representation and its object. In contrast to cued representations, indexical ones do not have to be descriptive at all, but can function only as indicators or "pointers".

Indexicality is necessary not only for executing actions, but also for the preparation of action. The subject has to reason or plan for herself in a specific setting if the plan is to be feasible. For instance, to keep your appointment with the dentist, it is not enough that you know that Liz Taylor is due there at noon on the 1st of April. You must also know that you are Liz Taylor and that the 1st of April is today. The subject must have an indexical self-awareness or she will not realize that a certain plan concerns herself.

The same goes for the mental map of the rat: its self-representation must be from a certain point of view or the information in the map will not connect to the actual context. The map is used when the animal is planning a route through, for example, a maze (Tolman 1948). For such a plan to

function it is necessary that the rat can represent *the present location of itself* on the map. Otherwise it would not know where to start planning its route. However, this does not entail that the rat can *imagine* itself being in a place other than it actually is. Nor does it entail that the rat can have different *attitudes* (for instance, desires) concerning its being in different locations. Presumably, it cannot “think” things like “I wish I were at that T-junction, because then I would be very close to the food bowl”. Even more remote would be to assume that it can represent its future desires, for example, that it will be hungry in two hours, so it had better start moving now (since it is such a long way to the goal).

The indexical representations necessary for action emerge in the interaction between the subject and its surroundings. They depend on the subject’s ability to orient itself in the perceptual field. The subject perceives the world from its own perspective: its point of view is anchored to its body. The subject moves around in different directions for different purposes and its movements gradually impose a structure on the perceptual field. It is placed in the center of the perceptual field with the surrounding items organized around it. The subject can adjust its position in the field, on its own initiative or as a response to the acts and movements of other individuals and to the character of the environment, and thereby update the information and keep the structure coherent.

Merleau-Ponty (1962) has emphasized that intentionality is first and foremost a *bodily* capacity and not a mental one.¹⁰ It follows from this that our representation of the surrounding world is anchored in the indexical perspective that derives from the motility of the body. For instance, he writes (1962, 279):

In so far as the body provides the perception of movement with the ground or basis which it needs in order to become established, it is a power of perception, rooted in a certain domain and geared to a world.

The content of indexical representations is, accordingly, determined by the interaction between perceptual input and behavioral output. It is not perspectiveless or neutral, rather every specification of the content of a certain representation involves the subject’s relation to the specified item. Such content depicts what things are *like* to a specific subject, not what they are in an objective or generalized sense.

Hence a minimal condition on an agent is that it has an egocentric representation – a point of view. For the rat in the maze, for example, the egocentric representation of the location provides a point of departure.

Indexical representations are connected with an *indexical self-awareness*.¹¹ Indexical self-awareness emerges when the subject gradually creates an egocentric space for herself. Having a point of view, or locating

beliefs (i.e., that represent the present location of the agent), demands of course the ability to distinguish between oneself and the rest of the world. That a subject has an indexical self-awareness means that it experiences itself as being placed in space and time. It also means that the subject can conceive of objects as being related to itself.

To act purposefully, an agent with an indexical self-awareness must be able to discriminate and locate the objects of its actions, that is, those objects that occur in its perceptual field.¹² The agent must grasp the notion of an object, which means that the agent has to learn how to categorize perceptual information in a way that is appropriate for action. One way to do that without first explicitly thinking about objects conceptually, *as* objects, or without behaving intentionally towards them, is by interacting nonintentionally with them. Both in succumbing to action and in resisting it, they reveal themselves *as* objects to the agent.

Merleau-Ponty has brought attention to the central role of the body in categorizing the perceived world. He writes (1962, 326):

A thing is, therefore, not actually *given* in perception, it is internally taken up by us, reconstituted and experienced by us in so far as it is bound up with a world, the basic structures of which we carry with us, and of which it is merely one of many possible concrete forms.

Subject and external world are entwined in the inner world. Their coexistence in the inner world is conditioned by the body.

Perceptual content could not as such give rise to indexical self-awareness. Interaction with the environment is necessary, or the agent will not grasp the relation between the objects and itself, but only the relations between objects. The agent becomes aware of itself through other objects, by simultaneously using its body and its different senses in interacting with them.

John Campbell (1995, 32) maintains that having the idea of something being an object involves grasping that it has a two-dimensional causal structure: it is at once internally causally connected over time and a common cause of many phenomena. That an object is internally connected means that its state at any moment depends upon its preceding states (Campbell 1994, 27). Campbell's principle of the common cause, on the other hand, concerns the external relations between objects and the ways in which they interact. It presumes that an object forms a unit in space. An agent could not understand how objects act upon each other if it did not grasp this causal structure.

It seems to us that a successful indexical representation of an object would have to involve the conditions both of unity and extension over time. Agency is, moreover, impossible without a minimal grasp both of oneself as a causal power and one's position in relation to other objects in the

context of action. To act, one must experience the world as distinct from oneself and the objects in it as items (and not fluctuating collections of properties or features) that extend over time.

Indexical representations do not only provide the agent with a spatial map of the environment as seen from its point of view. In experiencing and interacting with the environment, the agent takes a location in relation to other objects. Henceforth, it is itself located on the map. This is, however, not sufficient to support the conception of oneself as an object among others. To look upon oneself, so to speak, from the outside, or from a third-person perspective, demands well-developed conceptual capacities (although not necessarily linguistic ones). Indexical self-awareness can obviously be coupled with detached representations of oneself and the world. But the indexical representations necessary for action are independent of such representations. One can have one without the other.

For instance, small children have indexical self-awareness, but they do not have a detached representation of their point of view. There is a wealth of evidence for this, the classic example being the “three-mountain problem” (Piaget and Inhelder 1956). In this experiment three “mountains”, one bigger than the other two, are placed in a triangle on a table. The child to be tested sits in front of the small mountains, while a doll is placed on a chair facing the large mountain. The child is asked to draw what the doll “sees” from where it is sitting. A child in the “preoperational stage” (Piaget’s term) draws how the scene looks from its own perspective, independently of where the doll is seated. However, a child in the “concrete operational stage” can take the doll’s point of view and draw the “correct” perspective. This suggests that once children have this capacity, they can view the world from many different points of view independently of the perceptual input they are currently receiving.¹³

We have suggested that agency requires an indexical self-awareness that emerges from the interaction of the subject with the environment. It may seem that the account is circular, since we claim that agency requires indexical self-awareness, but the development of such self-awareness in turn appears to depend on agency. Nevertheless, the circularity is avoided, since the initial interaction between subject and environment that brings about indexical self-awareness does not have to involve representations. Agency, which requires the use of representations on the part of the agent, is thus not presupposed by indexical self-awareness.

Indexical self-awareness consists in contextual information gained from the interplay of perception and behavior, both of which depend on the body. Thus, it seems, egocentricity would be impossible without interaction with the environment, and interaction would, in turn, be impossible

without embodiment. But this is not altogether true. Interaction is necessary for egocentricity, but this could, however, take place without bodies. Imagine a severely handicapped person who can only communicate via readings of her brain activities. This would be sufficient for her to interact with the surroundings, even though she could not use her body to communicate. The body would then not be necessary for establishing a way of communicating with others. This means that embodiment is not necessary for indexical self-awareness, but situatedness and locating beliefs are.

Let us sum up the discussion concerning levels of intentionality. A subject is intentional in a minimal sense if it has representations. To reach an intermediate level of intentionality, the subject must be capable of having detached representations that form an inner world. The subject should also have an indexical self-awareness.

4. DETACHED SELF-AWARENESS AND SELF-CONSCIOUSNESS

A higher level of intentionality is reached when indexical self-awareness is combined with a detached one. An agent with a *detached self-awareness* has at least some self-representations that are cut loose from the actual context. She can think of herself generally, as a subject that may instantiate different properties in different domains. This kind of generality also paves the way for self-representations that attribute properties to the subject that she actually does not have and thus for counterfactual thoughts about oneself. Such thoughts are useful in planning for circumstances other than the actual one, for instance, when the agent considers possible solutions to a problem. An example would be a subject believing that she will become unemployed and who ponders different strategies to cope with that situation.

The general ability to envision various actions and their consequences is a necessary requirement for an animal to be capable of *planning*. Following Gulz (1991, 46), we will use the following criterion: An animal is planning its actions if it has a representation of a goal and a start situation and it is capable of generating a representation of a partially ordered set of actions for itself for getting from start to goal. The representations of the goal and the actions must be detached, otherwise the animal will only be capable of trial-and-error behavior. In brief, planning presupposes an inner world with detached representations.

Representation of future or possible events does not demand object-centered self-representations, that is, representations of oneself as an object among others. Such non-indexical or object-centered self-representations are rather unusual. On the other hand, indexical

self-representations are necessary for planning and agency. A self-representation totally void of indexical content would not move the subject to action.

Some cases actually demand that the agent can take a view of itself as an object among other objects. This happens if the agent needs to plan for a team, and its own role is confined to be one of the members of the team, all of which are on an equal level. There are many examples of team-work of this kind, from a group's joint defense of its camp against an anticipated attack from enemies, to the strategy of a football team in anticipation of an important match.¹⁴ In these cases, the agent is not primarily planning for itself, but for the whole group.

Using the distinction between referent-dependent and referent-independent detached representations, one can go further and distinguish between the corresponding kinds of *goals* of an agent. Thus an animal that only has the referent-dependent type of representations, could not have a non-existent object as a goal. However, with the more advanced referent-independent form one can, for example, truly seek a unicorn.

There are several clear cases of planning among primates and less clear cases in other species. However, all evidence for planning in non-human animals concerns planning for *present* needs.¹⁵ Apes and other animals plan because they are hungry or thirsty, tired or frightened. Humans seem to be the only animal that can plan for *future* needs. Gulz (1991, 55) names planning for present needs *immediate planning* while planning for the future is called *anticipatory planning*. Humans can predict that they will be hungry tomorrow and save some food, and they realize, for instance, that the winter will be cold and are therefore able to start building a shelter in the summer.

The crucial distinction is that for an organism to be capable of anticipatory planning it must have a *detached representation of its future needs*. In contrast, immediate planning only requires a cued representation of the *current* need. There is nothing in the available evidence concerning animal planning, notwithstanding all its methodological problems, that suggests that any species other than *Homo sapiens* has detached representations of their desires and goals. Anticipatory planning requires that the agent can suppress the feelings and desires of the current situation and evoke memories, context-independent desires or fantasies, during the planning.¹⁶

Fullblown *self-consciousness* requires both indexical and detached self-awareness. A subject must be capable of making inferences that involve both kinds of self-representation, to go, for instance, from the thoughts "I am sad" and "That tall creature is sad" (for example, when looking at a mirror image of itself) to "I am that tall and sad creature". This means that

the subject connects first- and third-person beliefs about itself.¹⁷ An agent with only third-person beliefs about itself would not be self-conscious, since one cannot connect general beliefs to oneself without an indexical self-awareness. General beliefs must make contact with the actual world to concern a particular subject. They must be tied to the context or the agent will not be conscious of *itself*, but of, for instance, Liz Taylor (whoever that is).

Self-consciousness is essentially from the first-person perspective; it does not depend on reidentifying oneself from context to context. The self-awareness that arises from the subject's relation to and constant interaction with her environment suffices to guarantee self-identity, at least in one sense of the word. Subjects do not fundamentally conceive of themselves from a third-person perspective and thus do not primarily think about themselves as objects. As long as the subject takes an active part in life in this way, she does not run the risk of losing track of herself in the common, objective world.

As Merleau-Ponty and others have emphasized, perception depends on the subject's ability to engage in interaction with the environment over time. A completely passive subject would not be able to impose a structure on the external world. This means that it could not perceive the external world as constituted by different objects where different events take place, all falling into separate categories. The subject would then not have an inner world.

This further implies that the mode of existence of agents guarantees self-identity in a fundamental sense, as of being a mobile point of view. The reason is that self-identity is a consequence of the subject's interaction with the world. Agency, self-awareness and a basic kind of self-identity go hand in hand. Of course, this does not exclude that a person doubts whether she is exactly the same person, physically, psychologically, or socially, as, say, ten years earlier, or that she has, for instance, a split personality. A subject can also go through a gradual change without any grave disturbance or interruption to her perception of the external world, as long as there is a *continuity* over time of herself (physically, psychologically, and perhaps also socially) – continuity being necessary for having an inner world.

5. CONCLUSION

We have in this article formulated several conditions for intentional subjects. These conditions hold for intentionality viewed as an intrinsic property of subjects, in contrast to Dennett's intentional stance. Basically, we have identified three levels of intentionality. To qualify for the lowest

level the subject must be capable of having representations. We distinguish between two kinds of representations: cued and detached. Cued representations are prompted by the context. In contrast, the referent of a detached representation does not necessarily have to be present.

The intermediate level of intentionality is achieved by having an inner world, that is, a system of detached representations that form a coherent model of the external world. The inner world can, for instance, be used for generating possible consequences of actions. In order to use the inner world for planning and prediction, indexical representations are needed. The reason is that agency requires indexical self-awareness.

To reach the highest level of intentionality, the subject must have a detached self-awareness, that is, self-representations that are cut loose from the current situation of the subject. This makes it possible for the subject to think of herself from a third-person perspective. Fullblown self-consciousness requires both indexical and detached self-awareness. A special case of self-representation is when the subject has detached representations of her future desires. Such representations are necessary for anticipatory planning.

NOTES

¹ Daniel Dennett (1978; 1981; 1983) uses a similar characterization to define intentional systems. He says that an intentional system is “a system whose behaviour is reliably and voluminously predictable via the intentional strategy” (Dennett 1981, 55). For Dennett, intentionality seems to lie in the eye of the beholder and not primarily in the system itself. In contrast, we think that intentionality is an intrinsic property.

² For a general discussion of representations in animals, see Roitblat (1982), Gopnik (1982), Lachman and Lachman (1982), Fodor (1986), Gulz (1991), and Gärdenfors (1996a, 1996b).

³ Representations, as we conceive them, can carry information in non-conceptual, conceptual, as well as linguistic form. Having linguistic capacities is not necessary for being an intentional subject: Representations are necessary, but not linguistic ones. Our theory of representation is, moreover, compatible with various naturalistic theories of content such as covariation, causal, and teleological theories.

⁴ In order to use detached representations effectively, the organism must be able to suppress interfering cued representations (compare Deacon 1996, 130–131).

⁵ Vauclair (1987) provides a more recent analysis of the notion of a “cognitive mapping”.

⁶ This notion of an inner world is much more restricted than the “Innenwelt” in von Uexküll’s writings, since his notion includes all kinds of “effects evoked in the nervous system by the factors of the environment” (1985, 223).

⁷ This possibility is explored by Pallbo (1997).

⁸ This idea reminds somewhat of Dennett’s (1991) “multiple drafts model”. One difference is that Dennett does not emphasize the interaction between subject and environment, what is sometimes called the situatedness of the subject (see Clark 1997).

⁹ The representations may have originated in a specific setting, but their content is independent of the context of use.

¹⁰ Reuter (1999, this volume) writes: “Merleau-Ponty’s basic intentionality is the body-subject’s concrete, spatial and pre-reflective directedness towards the lived world”. Her paper extensively discusses Merleau-Ponty’s notion of pre-reflective intentionality and its bodily basis.

¹¹ For a discussion of different kinds of self-awareness, see Brinck (1997).

¹² We are not presupposing any particular metaphysics of objects.

¹³ Piaget held that the child could do this from about the age of seven. However, recent research suggests that this capacity is acquired at a much earlier age.

¹⁴ Here we assume either that it is not the coach that plans the strategy but one of the players, or that the coach is playing in the team. Team and coach both have the same goal.

¹⁵ Squirrels and other animals who collect food for the winter have no representation of the goal and hence they are not planning. Their behavior is just instinctive, as can be shown by different kinds of experiments.

¹⁶ For the role of memory in suppressing current information see Glenberg (1997) and Gärdenfors (1997).

¹⁷ This issue is discussed in chapters 5 and 6 of Brinck (1997).

REFERENCES

- Brinck, I.: 1997, *The Indexical “I”*, Kluwer Academic Publishers, Dordrecht.
- Campbell, J.: 1994, *Past, Space, and Self*, MIT Press, Cambridge, MA.
- Campbell, J.: 1995, ‘The Body Image and Self-Consciousness’, in J. L. Bermúdez, A. Marcel, N. Eilan (eds.), *The Body and the Self*, MIT Press, Cambridge, MA.
- Clark, A.: 1997, *Being There: Putting Brain, Body and World Together Again*, MIT Press, Cambridge, MA.
- Craik, K.: 1943, *The Nature of Explanation*, Cambridge University Press, Cambridge.
- Deacon, T.: 1996, ‘Prefrontal Cortex and Symbol Learning: Why a Brain Capable of Language Evolved Only Once’, in B. M. Velichkovsky and D. M. Rumbaugh (eds.), *Communicating Meaning: The Evolution and Development of Language*, Lawrence Erlbaum, Mahwah, NJ, pp. 103–138.
- Dennett, D.: 1978, *Brainstorms: Philosophical Essays on Mind and Psychology*, MIT Press, Cambridge, MA.
- Dennett, D.: 1981, ‘True Believers: The Intentional Strategy and Why It Works’, in A. F. Heath (ed.), *Scientific Explanation*, Clarendon Press, Oxford.
- Dennett, D.: 1983, ‘Intentional Systems in Cognitive Ethology: The “Panglossian Paradigm” Defended’, *Behavioral and Brain Sciences* **6**, 343–390.
- Dennett, D.: 1991, *Consciousness Explained*, Little, Brown and Company, Boston, MA.
- Fodor, J. A.: 1986, ‘Why Paramecia Don’t have Mental Representations’, *Midwest Studies in Philosophy* **10**, 3–23.
- Gärdenfors, P.: 1996a, ‘Cued and Detached Representations in Animal Cognition’, *Behavioural Processes* **36**, 263–273.
- Gärdenfors, P.: 1996b, ‘Language and the Evolution of Cognition’, in V. Rialle and D. Fisette (eds.), *Penser l’esprit: Des Sciences de la Cognition à une Philosophie Cognitive*, Presses Universitaires de Grenoble, Grenoble, pp. 151–172.

- Gärdenfors, P.: 1997, 'The Role of Memory in Planning and Pretense', *Behavioral and Brain Sciences* **20**, 24–25.
- Glenberg A. M.: 1997, 'What Memory Is For', *Behavioral and Brain Sciences* **20**, 1–19.
- Gopnik, M.: 1982, 'Some Distinctions Among Representations', *Behavioral and Brain Sciences* **5**, 378–379.
- Gopnik, A.: 1993, 'How we Know our Minds: The Illusion of First-Person Knowledge of Intentionality', *Behavioral and Brain Sciences* **16**, 1–14.
- Gulz, A.: 1991, *The Planning of Action as a Cognitive and Biological Phenomenon*, Lund University Cognitive Studies 2, Lund.
- Jeannerod, M.: 1994, 'The Representing Brain, Neural Correlates of Motor Intention and Imagery', *Behavioral and Brain Sciences* **17**, 187–202.
- Lachman, R. and J. L. Lachman: 1982, 'Memory Representations in Animals: Some Metatheoretical Issues', *Behavioral and Brain Sciences* **5**, 380–381.
- Luria, A. R.: 1973, *The Working Brain*, Basic Books, New York.
- Merleau-Ponty, M.: 1962, *Phenomenology of Perception*, Routledge and Kegan Paul, London.
- Pallbo, R.: 1997, *Mind in Motion: The Utilization of Noise in the Cognitive Process*, Lund University Cognitive Studies 57, Lund.
- Piaget, J. and B. Inhelder: 1956, *The Child's Conception of Space*, Routledge and Kegan Paul, London.
- Reuter, M.: 1999, 'Merleau-Ponty's Notion of Pre-Reflective Intentionality', *Synthese* **118**, 69–88 (this issue).
- Roitblat, H. L.: 1982, 'The Meaning of Representation in Animal Memory', *Behavioral and Brain Sciences* **5**, 353–372.
- 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**, 1–10.
- Tolman, E. C.: 1948, 'Cognitive Maps in Rats and Men', *Psychological Review* **55**, 189–208.
- Von Uexküll, J.: 1985, 'Environment and Inner World of Animals', in G. M. Burghardt (ed.), *Foundations of Comparative Ethology*, Van Nostrand Reinhold Company, New York, pp. 222–245.
- Vauclair, J.: 1987, 'A Comparative Approach to Cognitive Mapping', in P. Ellen and C. Thinus-Blanc (eds.), *Cognitive Processes and Spatial Orientation in Animal and Man: Volume I Experimental Animal Psychology and Ethology*, Martinus Nijhoff Publishers, Dordrecht, pp. 89–96.

Department of Philosophy
Lund University
Kunghuset
S-222 22 Lund
Sweden