

## Chapter 14

Ruth Millikan

► **To cite this version:**

Ruth Millikan. Chapter 14. The Jean-Nicod Lectures 2002 (expanded version), 2003.  
<ijn\_00000388>

**HAL Id: ijn\_00000388**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_00000388](https://jeannicod.ccsd.cnrs.fr/ijn_00000388)**

Submitted on 12 Sep 2003

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## CHAPTER FOURTEEN

### DETACHING REPRESENTATIONS OF OBJECTS

Beginning with minimally articulate P-P representations, the evolution of inner representations seems likely to have paralleled evolution writ large. First, representations have become more articulate, so that more and more of what they represent is represented explicitly. More complex functions are then built up out of more specialized functions of the articulated parts. Then ways to perfect these more specialized functions somewhat independently have developed, sometimes by the development of new generate and test procedures. These articulated specialized functions are then recombined and reintegrated in new ways. The general strategy involved --disassemble, tune the parts separately and recombine-- is typical of evolutionary developments more generally. We encountered it before in chapter two when discussing the evolution of evolvability. The following chapters outline some of the results of this strategy during the evolution of inner representations, keeping a special eye out for developments that begin to separate the pushmi from the pullu sides of representations, that is, the descriptive from the directive.

Primitive behavior releasers are often activated by quite crude patterns of proximal stimulation. Anything that is red and gaping of about the right size on the retina will serve as a stimulus to drop in a worm for a parent bird at the edge of its nest or for a bird engaged in courtship feeding. Jackdaws that had imprinted on Konrad Lorenz as if a member of their own species tried to drop worms in his ears (Lorenz 1952). Imprinting itself tends to occur on any object at all that is moving about near the animal during the earliest moments of its life. Anything decorated with two blobs above and a sort of horizontal line below will attract the gaze of a human infant in the first weeks of life, thus focusing its attention on human faces which are important things for it to study. The greylag goose will try to pull into its nest just about any proximate object with rounded contours that is not bigger than the goose itself. If other objects similar to the targeted object are unlikely to appear in the contexts in which perception of a certain affordance is potentiated, these crude perceptions of affordances may serve quite well. But for other tasks, more discriminating methods are needed.

The central problem for an animal that needs to discriminate more carefully among various affording affairs involving distal objects is that the same kind of distal affair is likely to have myriad alternative effects on the sensory surfaces of the animal, depending both on its spatial relation to the animal and on mediating or intervening conditions such as lighting conditions, atmospheric conditions, sound absorption and reflectance properties of surrounding objects, obscuring conditions such as intervening objects, masking sounds and odors, and so forth. This doesn't always matter. If all the animal needs to know is that winter will be along within a month or so, then there may be quite a number of telltale proximal stimulations that it is likely to encounter sometime within the necessary time frame, such as frosty nights, certain smells, the look of the sun appearing lower in the sky and so forth. But most distal affairs that matter to an animal need to be registered more immediately. This means that the animal needs to be able to recognize the affording affair from a wide variety of perspectives and under a wide variety of conditions. To be as useful as possible, the apparatus that recognizes the affording distal object or situation must recognize it over as wide a range of relations to the animal as possible, near, far, overhead, underfoot, left, right, partly occluded, and under a variety of mediating conditions such as lighting conditions, sound transfer conditions, when the animal is moving or still, when the object is moving or still, despite "static" such as fog, wind and other extraneous noises, dappled shadows,

other entwined smells, and so forth.

Similarly, affording distal affairs typically need to be registered not merely as somewhere within the animal's vicinity but also as currently bearing some quite definite relation to the animal, for the appropriate reaction on the part of the animal will vary as a function of this relation. For example, the animal needs to know not just that a certain predator or prey is present but in what direction and at what distance. Thus the animal may also need to command a variety of different ways to recognize the same relation to the affording object or affair, depending on conditions. It may need to hear and feel direction as well as to see direction, for example, and it may need to employ a variety of alternative means of depth and distance perception, using several sensory modalities as well as several methods within in a single modality.

This kind of need is graphically illustrated by a story that has been circulating fairly widely, and though it is probably apocryphal, makes its point very well.<sup>1</sup> The story is that certain venomous snakes perceive mice for purposes of striking by sight, trace the path of the dying mouse by smell, and find its head so as to swallow that part first by feel, and that none of these jobs can be accomplished using any other sensory modality. A snake that was wired up this way would merely perceive first a "strike me", then a "chase me" and finally a "swallow me," having no grasp at all that what it struck, followed and swallowed was the same thing. The story is probably apocryphal, but it has a certain plausibility. Both objects themselves and their perceived relations to oneself show up differently through different sensory modalities. Consolidating one's inner representational system so that the same object and the same relation to that object are always recognized as being the same is not a trivial task. But surely it would be more efficient for the snake to be able to use each of its various ways of recognizing a mouse in some relation to itself for whatever purposes it might have for a mouse at the moment. Moreover, it would be good if its ways of perceiving its relation to the mouse when striking, following, and swallowing could be detached for use in perception of similar relations to other kinds of affording objects.

In a similar vein, recall the connectionist face recognizer, VisNet, from Chapter Nine that recognizes seven individual faces from each of nine angles but would have to start all over again to learn to recognize an eighth face. Clearly this sort of holistic approach to face recognition is not very efficient. If you have need to recognize many different faces, especially if you need to learn to recognize new faces quickly, it is more efficient if you first put in place the more abstract but multipurpose capacity to recognize any shape, as such, from any arbitrary angle and at any arbitrary distance. It is better if you have already in place the capacity that psychologists call "shape constancy." Generalizing this, suppose that a great many different objects need to be recognized by an organism, each under numerous different conditions. It will be efficient to divide or disassemble the function of the of the organism's object-recognizers into parts or stages, each of which recognizes members of a certain class of properties such as shape, color, size, texture, quality and direction of movement, identity of sound at origin (e.g., voice constancy), solidity or malleability, weight or angular inertia, and so forth, under many different conditions. Then these various recognition capacities can be combined as necessary, using each over again in as many new configurations as possible. That is, you build in a set of prior apparatuses that register simple objective physical properties and relations as constant through changing perspectives and intervening media. In different combinations, these properties are likely to be locally recurring signs of a wide variety of affording objects, kinds, stuffs, events and other world affairs, each of which it will now be easy to learn to recognize.

Also, just as the same property such as shape, color or texture may affect the sensory

surfaces in different ways on different occasions, which of an object's various properties are currently observable to an animal also varies from occasion to occasion. It is best then if the animal can recognize affording objects by way of many different alternative sets of diagnostic properties. This is a theme that I have developed at length elsewhere,<sup>2</sup> however, and I will not say much more about it here.

Representations of properties used for the purpose of detecting affording objects are not dedicated to particular practical purposes settled in advance. They are not pushmi-pullyu representations. They tell of the disposition of properties among various objects in the environment without yet saying what is to be done as a consequence. They are steps on the way to perception of affordances, but they are not themselves perceptions of affordances. The same is true, of course, for any representations that are prior to representations of properties of objects, such as representations of lines or edges with a particular orientation, right angles, ocular disparity, directional movement, color edges, and so forth. These represent detached facts. So we have discovered one small way in which pushmi-pullyu representations may begin to come apart, descriptive representations detaching from directive ones. It would be a bad mistake, however, to assume that use of these descriptive signs required the animal to have concepts of properties. The capacity to discriminate a property is not, as such, a concept of a property. (Many one-celled organisms discriminate dark from light; it does not follow that they employ any concepts.) The concept of a property, presumably, is the kind of thing that can play a role in propositional judgment and mediate inference, and there is no reason to suppose that simple property detectors are employed towards any such ends.

The identification of affording objects and situations is, however, only half of what is needed for guidance of effective action. Recognition of an affording object tells the animal what to retreat from or what to approach, what to pick up or to eat or to climb up on and so forth. But the animal will not perceive how to perform any of these maneuvers unless it also perceives its own relation to the affording object or situation. Perceiving that there is an apple which affords eating is one thing; perceiving exactly how I would need to reach from here to obtain it is another. These two perceptual aspects need to be combined in a single articulate pushmi-pullyu representation to guide the procurement and eating of a suitable object. Correlatively, two general-purpose skills must be combined here to make up a single-purpose skill. First is the general skill that allows perception of an apple, hence perception of an eating affordance, through any of a variety of media from any of many perspectives. Second is the set of skills that allows perception of the current relation of the perceiving animal to the affording object so as to guide activities of approaching, picking up, and conveying to the mouth. Skills of this latter sort can be practiced, of course, in many contexts other than that of apple eating. We have here another example of disassemble, tune the parts and recombine.

The distinction between skills involved in identification of affording situations and objects versus skills involved in identification of relations of objects to the animal for purposes of interaction or manipulation seems to have relatively clear neurological correlates. Vision has been the most closely studied of our perceptual capacities and current neurological data on vision indicate these different aspects of perception quite clearly.<sup>3</sup> As mentioned in Chapter Nine, the origin of vision involves, first, the translation of gradients of luminance across larger or smaller areas of the retina to detect various rudiments of visual form (lines, edges, orientations, angles, movement, direction etc.) which are later processed to yield information about properties of

objects. But even at the level of the ganglion cells in the retina, a division already occurs between what are to become two relatively separate neural pathways of visual perceptual analysis, usually referred to as the *dorsal* and the *ventral* pathways or channels. Roughly speaking, the dorsal pathway is concerned with guidance of the organism's movements in relation to perceived objects, while the ventral pathway is concerned with identification of objects.

Ganglion cells that feed information into the dorsal channel process information from all over the retina including all peripheral areas, feed into channels that process information at high temporal frequencies, help to produce special sensitivity to the larger patterns on the retina and special sensitivity to ocular disparity and to motion. As the object or the organism moves, ocular disparity and motion displacement are fundamental, for example, to the ability of the dorsal channel to detect direction, distance, angle, location and size relative to the organism. In sum, the dorsal channel processes information relevant to performing motor movements in relation to perceived objects and events, walking or running towards or away from things, between or through them, climbing, pointing, reaching, grasping and so forth: "[H]ow large should the gap between the thumb and forefinger be in order to pick up that block?" (Norman 2002, 3.4.6). It processes information about affordances at the level of immediate movement and contact with objects-in-general. Thus Jeannerod calls the representations processed by the dorsal channel "pragmatic representations" and says of them that they "refer to rapid transformation of sensory input into motor commands" (1997, p. 77). A crucial fact about these representations may be that they are always used only for this purpose. For example, they cannot be used as the basis of expressed perceptual judgments or discrimination tasks of other kinds. They cannot be "brought to consciousness."<sup>4</sup>

Those retinal ganglion cells that feed mainly into the ventral channel lead to slower processing of information but at higher spatial frequencies. The ventral channel processes information from more central areas of the retina and it yields more detailed form, pattern and color analysis. It helps to produce special sensitivity to features necessary for object identification, such as more exact shape and size. It detects relations between objects better than relations to the animal. In general, it detects *what* objects the animal confronts so that it will know whether to go towards or away from them or what to do with them, separating this information from the animal's current accidental relations to these objects. Jeannerod calls representations processed by the ventral channel "semantic representations" and says of them that they "refer to the use of cognitive cues for generating actions." A dramatic illustration of the separation of these two systems is the experience of an experimental subject who is required, over a period of days, to adjust to wearing glasses that reverse right and left so that the world appears as a mirror image of itself. After a time he may be perfectly capable, for example, of riding a bicycle through traffic, everything seeming to him perfectly normal again, except that the license plate numbers on the cars are still backwards!

The dorsal and ventral visual channels are, of course, but two faces of one system. One's peculiar momentary relation to an object of interest needs to be canceled out in order to recognize the object, but of course it must be figured back in again if one is to manipulate the object or alter one's relation to it. "Thus, when one picks up a hammer, the control and monitoring of the actual movements is by the dorsal system but there also occurs intervention of the ventral system that recognizes the hammer as such and directs the movement towards picking up the hammer by the handle and not by the head" (Norman 2000, 3.4.8). A differentiation between these two channels is

apparently very old, being found in the visual systems of mammals from hamsters through monkeys to humans. A similar division between two information channels is found within the auditory system as well.

The ventral channel is often called the "what" channel while the dorsal channel is called the "where" channel or the "how" channel. These names are not very helpful, I believe. The ventral channel may typically represent "what it's for" rather than merely "what," showing not detached facts, primarily, but affordances. Thus Gibson told us that we perceive apples as affording eating and post boxes as affording letter-mailing (1979, p. 139). This may well be so most of the time for people, and perhaps all of the time for most animals.<sup>5</sup> In the first instance, at least, perception of objects immediately serves practical, not theoretical purposes. The dorsal system, on the other hand, is not alone in representing either where objects are or other attributes relevant to performing motor movements in relation to objects. "A large set of attributes are in fact relevant to both the semantic and the pragmatic processing:....shape, size, volume, compliance, texture, etc....An essential aspect of object-oriented behavior is therefore that the same object has to be simultaneously represented in multiple ways,..." (Jeannerod 1997 p. 78-79). Nor does the dorsal system merely represent "how." In order to direct how to move appropriately it has to represent what the relevant relations are between the animal and what it would act with respect to. The difficulty with all three of these designations, "what," "where," and "how," is that each fails to recognize the double aspect of the representations generated by these systems. Each fails to recognize either the pushmi or the pullu aspect. For dorsal and ventral systems each produce P-P representations, though not complete ones. These representations need to be joined to each other to make, as it were, a complete P-P sentence. On the other hand, it is not very good to refer to these two kinds of representations just as "dorsal" and "ventral" either, for the important distinction to be drawn is one of function, the degree to which the anatomical division between dorsal and ventral channels in humans or other animals accurately reflects the distinction being, of course, a matter for empirical investigation. Indeed, it is known that the ventral and the dorsal systems are able to exchange some information when necessary, although less accurately and at a delay. I will suggest what may be a more perspicuous terminology below.

It is commonly claimed that dorsal-system signs are "viewer centered" or "body centered" or "ego-centric" whereas ventral-system signs must be "object centered" or "allocentric." Often this difference is interpreted as a difference in the coordinate systems used for the two types of representation. In this connection it is worth noting that it is not a logical requirement for spatial representations that they employ any coordinate system at all. A model ship is a detailed spatial representation of a ship, but no coordinate system is employed in this representation. The sentences "New York is south of Boston," "New York is larger than Boston" and "New York is Twenty miles from here" are all spatial representations that employ no coordinates. The important difference between the two kinds of representations, I suggest, has nothing to do with coordinate systems but is as follows.

The dorsal part of the full P-P sign or the part that represents the relation of the affording object to the animal needs, of course, to represent the perceiving animal itself, but this representation needs only to be implicit. These signs are used in the context of showing relations of situations and objects to the animal for immediate guidance of the animal's motions. We can call these relations "enabling relations." Obviously the animal can only act for itself, only move its own limbs and so forth. It is perfectly clear then that there will be no intentionally significant

transformations of these signs showing other animals bearing these same enabling relations to these objects. These signs will contain no variable parts or aspects representing the animal itself. They will represent the animal implicitly in exactly the same way that the bee dance represents the hive, the sun and the nectar implicitly, there being no transformations of it that talk about the big oak tree, the moon or peanut butter (Chapter Seven). This sort of sign is "ego-centered" in that the ego is so central to it that it doesn't even need to be mentioned. Only the enabling relation needs to be mentioned. I will say that these signs represent "enabling relations" and that they are "ego implicit" signs.

On the other hand, the part of the full P-P sign that represents the affording object, configuration, or state of affairs but without representing its relation to the animal as needed for action may or may not represent the animal itself, but if it does represent the animal, it will represent the animal explicitly. It will permit significant transformations yielding intentional representations of things other than the animal in place of the animal's self. For example, if this kind of sign can represent an apple being about so far (a yard, say) from me, it can represent an apple being about so far from you in exactly the same way.<sup>6</sup> This seems a good reason to call this second kind of sign "objective," for if such a sign represents the self it represents it as one object among other objects. But, of course, an animal need not be capable of explicitly representing itself at all. It may well be that most animals do not have the capacity to represent themselves as objects, so that they harbor only ego-implicit representations and egoless representations, never ego-explicit representations. None of their objective representations include themselves as objects among other objects.

The ego-implicit versus objective distinction is an important one for the animal that needs to learn many new behaviors, for it partitions these tasks into two aspects, each of which can be learned and practiced separately, then are combined. On the one hand, there is the capacity to recognize the same individual or kind of object or kind of objective situation again. This rests on recognizing the same objective properties again, a skill that is practiced whenever objects sharing properties in the same range need to be identified. On the other hand, the animal can develop general skills for manipulating arbitrary objects and its relations to these objects. It can learn how to move among objects, climb up on them, jump from one to another, move them at will, grasp and pick them up, turn them over, throw them, and so forth. Young mammals often seem to be practicing such general skills while playing, with no more distant goals in focus. The playing cat doesn't care whether it is a mouse or a leaf or its tail that it chases. It is practicing chasing just things.

The task for the animal designed to learn a great deal by operant conditioning, which requires appropriate generalization and discrimination following successes, would be staggering if the animal were working with no prior knowledge of what variety of proximal stimulations might be signs of the same sort of distal object or affair, or of what variety of proximal responses might produce the same distal effects. Further, the attainment of objective representation may allow the animal to analyze its various activities into distinct achievement stages, each stage being recognized as what-the-objective-situation-is-now, apart from the animal's momentary position within that situation. Activities can then be understood as series of transitions from one objective situation into another objective situation, as stages in an objective process. There is evidence, for example, that it is by recognizing the objective completion stages of a process that animals sometimes learn from one another by imitation. Despite folklore about monkeys and apes who

"ape" others, the evidence is that what is "aped" in nature is not bodily motions but a completed series of project stages (R. Byrne, 1999, 2002). Taking a homely example, many cats will make a try at opening a door by reaching for the doorknob, even taking it between their two paws in an attempt to turn it. The human model whose results in action they are attempting to replicate does not use two paws, however, but one hand. A few animals can learn to imitate observed bodily motions, for example, surprisingly, dolphins can (Merman forthcoming). But the natural focus of animals' capacities to represent objective processes in the world is not on their bodies but on the objects they manipulate. Their bodies are represented only implicitly by the systems that must perceive relations of their bodies to other objects for guiding manipulation of these objects.

I have emphasized that perceptual representations that represent objects objectively, representing them apart from their momentary enabling relations to the acting animal, are not representations of facts but of partial affordances. Prey are perceived as for chasing, predators as for escaping from, and so forth. On the other hand, the same object may be perceived by an animal as having different affordances on different occasions, depending on the animal's current projects and needs. The ability to recognize water, for example, probably has a considerable variety of uses for most land animals. For the snake, the mouse affords striking, then following, then swallowing. If the snake were able to recognize the mouse for each of these purposes through any of its three sensory modalities, would this general ability to recognize a mouse be an ability to recognize the detached fact of the presence of a mouse? The kitten sees the mother cat as a source of food, a source of warmth, as protection, as a friend to play with, and so forth. Can it represent the detached fact of the presence of its mother? If learning how better to recognize an object for one purpose is carried over and applied to recognizing the same object for many other purposes, doesn't this amount to the ability to perceive a pure fact, detached from any practical use it might have?

Representations of pure fact are representations that are not dedicated to any particular purposes. They stand ready to be combined in the production of actions with purposes that have not been determined in advance, and perhaps with other factual knowledge that has not been determined in advance. The animal that represents a variety of objects and objective situations, recognizing different affordances of these at different times, certainly might be said to represent facts. But we should distinguish such an animal from one that can represent facts that it is not interested in exploiting at the moment, or that it doesn't yet know any uses for at all. In Chapters Eighteen and Nineteen, I will argue that this is quite an important distinction, perhaps one that helps to demarcate the peculiar intellectual capacities of humans from those of other species.



## FOOTNOTES

- 
1. The original source seems to be the Swedish zoologist Sverre Solander, who gives no references, however, and despite several requests from myself and others, has offered no data yet to my knowledge.
  2. in Millikan (2000).
  3. Most of the following information on vision can be found in (Jeannerod 1997) or in (Norman 2000). The latter is a paper with a thesis, but it contains a balanced review of the literature on the two visual channels discussed below.
  4. For a contrary opinion, see Neisser 2002.
  5. In a classic experiment, Sperling (1960) briefly presented human subjects with 12 letters arranged in three rows. They could usually report no more than four of the letters. But when prompted in advance to report on a specific row, they could report any row on demand even if the prompting cue was presented up to 250 ms *after* the offset of the target stimulus. It seems that the basic information needed for the construction of all 12 rows was represented in early processing, but

---

not brought to completion in full identification of the letters unless this information was specifically called for. In general, it is reasonable to suppose that there is a huge amount of natural information readily available to a human person through perception at any given time which that person is capable of translating into inner intentional representations, but that only a very small proportion immediately relevant to current concerns ever actually gets translated.

6. This is a point about what can be represented, not an epistemological point. I am not saying that it is always as easy to tell how far one thing is from another as it is to tell how far it is from me.