Trans/Form/Ação, São Paulo,
v. 14. p. 53-72, 1991.

RATIONALITY IN CHILDREN: THE FIRST STEPS

Andrew WOODFIELD*

ABSTRACT: Not all categorization is conceptual. Many of the experimental findings concerning infant and animal categorization invite the hypothesis that the subjects form abstract perceptual representations, mental models or cognitive maps that are not composed of concepts. The paper is a reflection upon the idea that conceptual *categorization involves the ability to make categorical judgements under the guidance of norms of rationality. These include a norm of truth-seeking and a norm of good evidence. Acceptance of these norms implies willingness to defer to cognitive authorities, unwillingness to commit oneself to contradictions, and knowledge of how to reorganize one's representational system upon discovering that one has made a mistake. It is proposed that the cognitive architecture required for basic rationality is similar to that which underlies pretend-play. The representational system must be able to make room for separate 'mental spaces' in which alternatives to the actual world are entertained. The same feature underlies the ability to understand modalities, time, the appearance-reality distinction, other minds, and ethics. Each area of understanding admits of degrees, and mastery (up to normal adult level) takes years. But rational concept-management, at least in its most rudimentary form, does not require a capacity to form second-order representations. It requires knowledge of how to operate upon, and compare, the contents of different mental spaces.*

KEYWORDS: The roots of rationality; conceptual categorization; perceptual representations; mental models; cognitive maps; common-sense psychology; cognitive capacities; representation of absent objects.

I. INTRODUCTION

Are children naturally rational, or do they have to learn to be rational? At what age do they start being rational? Questions like these were raised in the times of the Ancient Greeks and probably before. Every mother and father in human history has probably wondered about these matters, especially at times when the child was behaving badly. It is only in the twentieth century that people have begun to investigate these questions in a scientific way. I say 'scientific' with due caution. Although developmental psychologists aspire to this honorific title, it has to be recognized that their domain is not a hard science like chemistry. In some ways, our knowledge about the intellectual development of the child is pre-scientific; there are

* Department of Philosophy – University of Bristol.

philosophical and methodological problems which have not yet been solved. The notion of rationality is partly normative; so the standards of rationality need to be defined (and our choice of standards should be justified). Rationality is not a directly observable characteristic, so we need some operational tests of it. Interpreting the behaviour of the child can present problems. Given that the child moved her body in such and such a way, there is still the question of identifying her intention. If she produced apparent speech sounds, what did she mean by them? We should not simply take for granted that she meant what the words conventionally mean.

Rationality is not a single quality. It has many components. No one would claim to have a complete list of these. But I suppose that any list would include the following:

1. Means-end rationality. The capacity to act appropriately in the light one's desires and beliefs.

2. Rationality in belief-formation. Weighing appropriately the information provided by perceptions and by other people's utterances.

3. Inferential rationality. There are at least two types (deductive and inductive), and perhaps other (e. g. abductive, analogical). All have to do with transitions from premisses to conclusions. Deductive rationality includes drawing out the more obvious logical consequences of one's thoughts and noticing contradictions. Inductive rationality inciudes the ability to form generalizations and hypothesis that are well-founded in beliefs about individual cases.

4. Rationality in the Management of Stored Beliefs. Unwillingness to remain for very long with an incompatible set of beliefs. S tries to drop one (or more) of the beliefs, taking into account the balance of evidence and also the overall coherence with background beliefs.

In general, being rational is a good strategy for arriving at true beliefs, and for maintaining a store of beliefs that are true; both being worthwhile goals. If you form only empirical beliefs that are well supported, you will maximize the number of beliefs that correspond to empirical reality. However, success is not guaranteed. The relation between evidential justification and truth is complex; the former is not sufficient for the latter, nor is it necessary.

It is impossible to do justice to all aspects in one paper. I want to concentrate upon the roots of rationality, the very first signs of rational thinking, and the area I focus upon is conceptual categorization. On the whole, the literature on infant categorization fails to address the issue of rationality, and this reflects a widespread reluctance amongst psychologist to recognize how much weight is carried by the word 'concept'. Categorization at the conceptual level is a far more complex matter than mere stimulus-generalization. But may cognitive psychologists use the terms 'concept' and 'category' interchangeably, thus making it hard for themselves to keep in mind that not all categorization is conceptual, and at the same time obliterating the absolutely crucial distinction between ontological theory and psychological theory.

II. PERCEPTUAL CATEGORIZATION IN INFANTS AND ANIMALS

Developmental psychologists want to know how children of various ages classify things. The older the child, the more likely it is that she will group things in the same categories as her parents, because she will have had more time in which to learn what the parents do. A child who is old enough to speak and who knows a kind name 'K' must already possess some knowledge of the parents' use of that name. But the pre-verbal phase (from 0 to 12 months) is of great interest. If pre-verbal babies categorize their experiences at all, their classifications are more likely to be innate. Data about infants, and especially about new-born babies, will provide the best material to test for this. Accordingly, many researchers have speculated about whether the newborn's criteria of similarity coincide with adult criteria and whether babies see the world in the same way as grown-ups.

Until recently, good evidence of neonate classification was hard to obtain. Babies do not have sufficient motor control to indicate their cognitive responses by reliable limb movements. But in the late 70's and 80's, developmentalists made increasing use of the technique know as 'habituation-dishabituation by looking', and obtained results which were interpreted as evidence about the baby's ways of categorizing stimuli. Neonates can control where they look, and how long they look. The experimenter, by noting the baby's looking patterns, can tell what the baby finds interesting. When presented with a static display, the baby gets bored with it after a while and shifts her gaze to something else. If a novel stimulus is introduced, the directed babys's attention will be directed toward that stimulus. The technique exploits these facts. The baby is subjected to a series of habituation-trials in which an object or set of objects is presented and the baby is allowed to look for as long as she likes. If the displays on successive trial are perceived as similar, the length of time spent looking at them goes down. The baby becomes habituated to those objects. The experimenter decides upon a criterion of habituation, that is, a length of time such that, if the baby does not look at a stimulus for more than that time, the baby is said to have habituated to the stimulus. For example, the criterion may be half the looking-time that occurred on the first trial. After habituation comes a test trial. The baby is given a pair of objects to look at and the duration of her gaze at each object is measured. If she finds one of the objects novel, she will look at it for longer. If she gazes at an object for a relatively short time this is a sign that she assimilates it to the set of the objects which she became familiar on the habituation trials. Clearly, the experimenter can manipulate many variables within this framework, while controlling for individual differences and randomly fluctuating causes of looking or failing to look. Careful experimental designs yield results whose most plausible explanation is that the subjects perceive distinctions of kind and quality amongst the stimuli. The technique reveals whether the test-stimuli are expected or unexpected, against the background context of the training stimuli.

Thanks to this technique, many fascinating facts have been discovered about infants' perceptions of objects, events, and similarities. Very young children can and

do see objects as possessing quite 'deep' objective properties, as well as possessing superficial properties such as colour and shape. They see the external object as an external object, because they are sensitive to such properties as: the unity of the object (the fact that its parts stay together and move together), spatial boundedness (the fact that the object occupies a demarcated place in a three dimensional space), its substantialness (the fact that it impedes other objects from occupying the same place at the same time), and its spatio-temporal continuity (that is, the fact that it persists even though it may be temporarily hidden from view). Even the youngest infants see the world in terms of three-dimensional solid objects, and not as a series of fleeting two-dimensional colour-patterns. These discoveries, many of which are due to Elizabeth Spelke (and summarized in Spelke (12)), challenge the earlier doctrines of Piaget, who held that infants perceived the world egocentrically, as a sequence of evanescent sense-data. According to Piaget, children do not perceive independently existing solid objects in objective space until they have formed the *object-concept*, and it takes them at least two years to do so.

If properties such as depth, unity, substantialness, permanence and so on are not 'given' to the subject in the proximal stimuli that impinge on the retina, the subject must in some sense 'construct' the objective three-dimensional world by 'supplying' the missing properties. Piaget accepts this, and he then makes the further step of assuming that the *way* in which the mind objectifies the flux of experience is by subsuming the experiences under *concepts*. Piaget is, in this respect, a follower of Kant. He holds that the ability to perceive the environment as containing solid, substantial objects requires that the subject already possess certain concepts, such as the concepts *solid, substantial,* and whatever other concepts may be logically implied by the concept *object*.

Piaget saw no evidence that the infant under two years of age perceived the world thus. Indeed he had a great deal of evidence that seemed to show that infants fail to perceive the world oft enduring solid objects (e. g. the of replicated and reliable evidence of certain kinds of errors that are made by infants at various stages between 0 and 24 months). He was led, therefore, to hypothesize (a) that infants do not see the world as adults do, and (b) that the reason for *that* is that infants lack the necessary concepts.

If Piaget had been alive today, it would have been interesting to know his reaction to Spelke's experiments. Her results, like his, have been replicated many times. It is possible that Piaget would have accepted that infants perceive the world in terms of enduring, solid, three-dimensional objects right from the earliest days of life, and accepted that his stage 1 through stage 6 errors are probably due to non-perceptual in the infant. If he were to revise hypothesis (a) in the light of the new data, he would undoubtedly also reject (b) as well. He would draw the conclusion that Spelke herself draws, namely that children possess the *object* concept much earlier than the age of two, indeed that they probably possess it at birth.

However, there is a weak link in the reasoning, the link connecting hypothesis (a) with hypothesis (b). It is the assumption that a person must have the concept of

object in order to perceive something as an object. It depends what one means by 'concept', of course, but surely there exists the possibility that a subject might *experience* the depth, substantiality, and 'object-hood' of objects *without conceptualizing* them as three-dimensional, substantial objects. Not all mental representations are conceptual, after all. Some are images. And some are more like maps or models.

There are several reasons for favouring the alternative view that babies have representations of objects, and ways of categorizing objects into groups, which are nonconceptual.

In Spelke (12), no satisfactory criterion of the conceptual is given. Several criteria are hinted at, but none stands up too well. For example, she suggests that a representation of a quality Q is *perceptual* if quality Q is present in the proximal stimulus, *conceptual* if Q is not a property of the proximal stimulus. This is no good as a *definition*, for there are qualities such as colour and shape which are both present in the proximal stimulus and conceptualizable by adults. Since any quality can be represented in different formats, is impossible to define a form of representation (such as 'concept' or 'percept') just by specifying which qualities it represents. But even to use this as an operational criterion is unsatisfactory, because it begs the question against the hypothesis that depth, permanence, solidity, etc. can be represented in a purely perceptual way. Such a hypothesis taken seriously in the information pick-up' approach (see Gibson (4)), but it is compatible also with an 'information-processing' approach that assumes that such properties are neither present nor 'specified' in the proximal stimulus.

David Marr's theory of vision (11) and Fodor's theory of modular input-systems (1) both adopt the cognitivist 'information-processing' approach. For them, a perceptual system is a mechanism that takes proximal stimuli as input and produces as output representations of the distal stimulus (the external object). The system performs a series of computational transformations upon the input, guided by built-in 'assumptions' about the physical world and the properties of light. For example, if there is a sudden change in the texture gradient along a line in the retinal image, the visual system 'assumes' that there are two surfaces out there oriented in different planes with respect to the eye. This is not an assumption mode by the viewing subject. It is simply a hard-wired principle by which the mechanism operates. The fact that animals possess input-processing mechanisms that use such principles is due to natural selection; a mechanism sensitive to texture gradient-changes is well-adapted to the terrestrial environment in which such changes normally are correlated with changes in surface-orientation. The perceptions that we enjoy, the outputs of our special-purpose input-systems, are largely independent of our beliefs and thoughts. Even a solipsist who believed that the world was an illusion would still have visual experiences presenting an apparent world of solid objects in 3-D space. The solipsist cannot prevent his visual system from producing such experiences. Equally, it is probable that babies have visual experiences of a 3-D external world from the earliest age, as soon as their input-systems start to work.

Another criterion proposed by Spelke is that a representation is conceptual if it is *amodal*. She conducted a series of habituation experiments upon children aged 4 months, in which the infants were allowed to touch two rings, one in each hand, under a cloth which concealed the rings from view. One group was given two rings connected rigidly together by a metal bar, in fact, a single object shaped like a dumb-bell. The other group was given two independently movable rings joined together by a cord. Each infant was habituated to one of these two stimuli and was then shown alternating *visual* displays of a pair of rigidly connected rings and a pair of rings joined by string. Those who were habituated to the feel of the rigidly connected rings looked longer at the string-joined rings. This is evidence that they saw the latter as something different from the object they had touched. The infants who had been tactually habituated to the movable rings looked longer at the rigidly joined rings, again showing that they regarded the latter as being different from what they had touched. Spelke concludes that the child's expectations about the unity and boundaries of stimulus objects generalize from the haptic mode to the visual mode. The four month old child forms an amodal (or multi-modal) representation of the stimulus object.

This fascinating result fits in well with the theory that infants construct a cognitive map of their environment, locating bounded objects at positions within a spatial layout, and endowing those objects with many of the primary qualities. The experiment suggests that information obtained through different sensory channels gets pooled into a single, modality-independent map. But amodality is not sufficient to prove that the information is conceptualized. The common representational format which encodes information from different senses could be non-conceptual. And surely, if the subjects had been lower animals instead of human babies, the latter hypothesis would have seemed preferable.

The idea, perhaps originating with Tolman, that rats construct a geometric cognitive map of their local environment has been confirmed by a wealth of experiments (see Gallistel (3, chap. 6) for a review). Suppose it were shown that other species, much less intelligent than rats, also did so. We would be faced with a choice between saying that those animals possess concepts, and saying that concepts are not required in order for an animal to construct amodal representations of objects laid out in space. To justify our choice either way, it would be necessary to define more precisely what makes a representation count as conceptual.

Let us return to the categorization of visual stimuli. Herrnstein's work has shown beyond doubt that pigeons see the world in terms of a three-dimensional space containing objects, though their representations are not quite so rich in depth information as ours, and also that pigeons classify objects into groups, and that their groupings can be made to coincide with categories such as tree, car, human being, and so on. But it seems far-fetched to say that a pigeon, with its small central nervous system, can acquire concepts. Lloyd Morgan's canon ('Do not ascribe more mental apparatus to animals than is strictly necessary in order to account for their behaviour') recommends the second option: pigeons categorize things non-conceptually. Current cognitive science implicitly recognizes that

stimulus-generalization is not the same as conceptual categorization, for there exists a growing body of ethological and computational research perceptual categorization (or 'categorical perception') in animals and humans, which is *de facto* independent of the tradition of work in cognitive and developmental psychology. Its main concerns are to discover the discriminatory capacities of different species, and to propose models of possible mechanisms which would account for the discovered functional capacities.

In a well-known experiment, Hernstein (6) showed pigeons 80 colour slides each day. 40 were photographs of trees, trees in full view, trees partially hidden, under various lighting conditions and at various distances from the camera. These 40 were the positive instances. If a pigeon pecked at a key in response to such a slide, it received a food reward. The other 40 slides showed no trees, and pigeons received no reward if they pecked in response to them. Each slide was projected for 45 seconds. After 5 days even the slowest pigeon was discriminating the trees at a statistically significant level. The three fastest were discriminating by the second session, having seen the slides only once before. Evidently they found the task easy. Not only do pigeons have the capacity to see trees as similar to one another, they also readily exercise this capacity if given an incentive. They would readily do so in the wild, then, if nature provided them with an incentive. Indeed, they surely do have natural incentives for discriminating trees from other objects such as telegraph poles and chimneys, since trees are better places to perch, and feed, and build nests.

In another experiment (8), pigeons learned to sort underwater photographs of fishes (which were of various species taken from many different angles) from photographs of turtles, shrimp, starfish and scuba divers. In the wild, pigeons never encounter any underwater creatures, so grouping all fish together would not be naturally useful. Yet pigeons have the capacity to do this, and they can easily be induced to exercise the capacity, if any reward depends upon it. Other discriminations that pigeons can, and will make, in increasing order of difficulty, are: oak leaves vs leaves of other kinds (very easy); photographs of a particular woman in various orientations, contexts, and clothing-styles vs. photographs of other people (easy); pictures of Charlie Brown vs. pictures of other characters from the Peanuts cartoon (difficult); computer-generated line drawings of cubes and other solid forms vs. computer-generated distortions of such drawings which failed to represent solid forms (very difficult). The work is reviewed in Herrnstein (7).

From an evolutionary perspective it is plausible that a species highly dependent upon vision and widely dispersed across the globe should be innately endowed with a flexible capacity to learn to make discriminations on the basis of all sorts of visual cues. The discriminations that an individual pigeon actually learns are those that are useful to it in its particular environment. The set of useful discriminations amounts to a tiny subset of the set of possible discriminations that the pigeon *can* learn. Some will be harder to learn than others, of course (for an interesting discussion of the idea that animals are programmed to learn certain things easily, see Gould and Marler (5)).

It is also plausible that certain habits of grouping will be useful in *every* environment, for example, the tendency to classify perceptions of a single object viewed from different angles. Such tendencies might well be innate in pigeons.

All of the above remarks are probably broadly true of human infants, though the discriminations that come easily to birds may not be the same as the ones that come easily to human beings. However, facts about natural tendencies to generalize tell us evidence little or nothing about *concepts*. The evidence of perceptual categorization in infants is not of the right sort to show that infants possess the *object* concept, nor does it support the hypothesis that they have any specific classificatory concepts, such as *cat, human face, or Mama*. So what sorts of evidence would be relevant? I shall focus a small area, and try to establish a link between concept-possession and very rudimentary form of rationality. The starting point is common-sense psychology, and more specifically, the common-sense view of *judgements*.

III. JUDGING THAT AN OBJECT BELONGS TO A CATEGORY

Concepts are normally taken to be representations that figures in judgements and inferences. Consider a physical kind concept K, where K stands for a kind like cat, dog, or car. Among its various other roles, a K-concept plays a predicative role in categorical judgements. One important type of categorical judgement is of the form 'This is a K', where 'This' is a demonstrative referring to a currently perceived object. In order to make such a categorical judgement you need *two* representations. You need a percept of the object, and you need a general representation of the category K, and then you must combine the two predicatively to form a complete thought. As already mentioned, your percept of the object already categorizes it, in the sense that it represents it as having certain properties, including some quite deep properties. To judge that the perceived object is a K requires also that you mobilize a K-concept and apply it to that object. Every act of judging involves exercising some concept in the predicative part, and hence requires that the judge should already possess that concept. So judgements depend upon concepts. But I also want to claim that possession of the concept K is partly defined in terms of the ability to make rational K-judgements. The two are mutually interdependent, the concept K and the ability to make rational judgements about K-hood. There is a partial circularity here, but not a vicious circularity, because we can ground both theoretical terms simultaneously. Also it is not a closed circle, because concepts have other roles as well as their role in judgements, and they may be partly defined in terms of these other roles.

Another assumption made by common-sense is that judging is a mental act regulated by *norms*. Two important normative principles are:

(i) that judgements about empirical matters should aim at truth;

(ii) that empirical judgements should be justifiable by reasons.

In the case of a categorical judgement about a perceived object based on the object's appearance, S's perception should normally furnish good reason for the judgement.

A sophisticated judge makes a cognitive commitment, knowing that his or her act is potentially subject to evaluation and critical scrutiny in the light of norms (i) and (ii). A being who is capable of making judgements must understand not only that judgements can be either true or false, but also that false judgements are *incorrect* when assessed against principle (i). Such a being must appreciate also that hasty, groundless judgements are *bad* according to norm (ii), even they happen by luck to be true. I am talking here about a being who has fully mastered the art of judging, a competent participant in the game of rational inquiry. A young child who has not yet acquired full mastery will not fully appreciate that her categorical thoughts are evaluable by these two standards. But in order for her thoughts to count as judgements at all, she should be able to recognize in a rudimentary way the existence of intellectual values. She must appreciate that there is a difference between intellectual right and wrong.

Conceptual categorization, then, is a mental act governed by a norm of truth and a norm of evidential justification. Accepting that these norms guide one's mental activity, voluntarily submitting oneself to their authority, is a basic kind of rationality. Full concept-possession demands that the subject should possess at least that kind of rationality.

Let us examine what goes on in categorization tasks when the subject is a competent, rational concept-user. The proper description of the adult case is essential, before we consider the question of children. We need to describe the competences to be acquired, in order to identify the states of the child which approximate to those competences. In categorization research, there are several quite distinct paradigms. There is the 'free-sorting' task beloved of Inhelder and Piaget, the 'forced choice on triads' task, the 'discrimination-learning' task where stimuli are presented one at a time, and many other types of tasks. There might be no feedback about the result of a trial, or there might be feedback. In the latter case, feedback information can take various forms: it might be a reward, or the reply 'correct' or 'incorrect', or even an explanation of why one's action on the trial was correct or incorrect. It is important to specify the experimental paradigm, for the subject's strategies will be adapted to the demands of that particular paradigm as he perceives them.

In the paradigm I wish to consider, S is presented with one object on each trial, and there is a predetermined category K such that S has to decide whether the presented object is a K. S knows which category is the relevant one; S possesses the concept K, and S knows that he is supposed to decide whether the object is a K. S may choose between three types of response (e. g. there are three buttons to press): one means 'Yes', one means 'No', and the third means 'I don't know'. This task is far easier than the one in which the experimenter has a certain definite category in mind as the one which fixes the standards of correct and incorrect responses, but where S is

ignorant of which category that is. On the latter task, S tries to guess the category, using the information that gradually accumulates about the correctness or incorrectness of his responses so far. Under the 'ignorance' condition , the task is two-fold. The general problem across a whole series of trials is to find out which category fixes the criterion of correctness. The specific problem on each trial is to decide 'Is this object before me a member of category X?', where X is the category that S is currently supposing to be the relevant category. No algorithm exists for solving the general problem, because the information available to S, however long he sits at it, always underdetermines the choice of hypothesis as to which category is the right one. But although it is theoretically impossible to guarantee that a person or a machine will solve the problem, in fact people often succeed in guessing the category that the experimenter had in mind. S's homing-process is highly constrained; one hypothesis is sometimes much more salient than all its rivals.

But we shall focus upon the easier task, where S does know which category is in question. Suppose the category is *oak tree*. The task is, essentially, a test of S's skill at recognizing oak trees on the basis of their visual appearance.

There are many concepts that I possess which I cannot reliably or confidently apply to perceived objects. I do not know much about the appearance of instances of the following concepts: chlorophyll, cholesterol, maple tree, capibara. If you present me with furry animals that are not dogs, cats or any other kind that I am familiar with and ask me on a series of forced-choice trials if they are capibaras, my performance will be poor. But I do possess the concept *capibara* (I know that capibaras are South American rodents, and that they are the biggest rodents in the world), and I would exercise that concept on each trial, in an affirmative or a negative judgement. If I am forced to say definitely yes or no, even in cases where I prefer to say 'I don't know', I will *expect* a low success rate. Similarly, some people are poor at recognizing oak trees. Poor performance by S does not show that S was not trying to put the objects in the right category. On the contrary, on every trial S exercises the relevant concept *oak tree*. But an observer might not be able to tell that S is employing the concept *oak tree* at all.

Conversely, consider the response pattern of a different subject attempting to judge whether the same stimulus objects are beech trees. And suppose that this man is a poor judge of beech trees. His responses may coincide exactly with the responses produced by a good judge of oak trees. Even a 100% correct response profile for the oak tree task would not prove that this man had been employing the concept *oak tree* in his judgements. Perfect performance is not sufficient proof, just as poor performance is not sufficient disproof.

Another important point is that S's success-rate depends upon how difficult the stimulus-materials are. It depends, for instance, upon the degree of contrast between the positive stimuli and the negative stimuli. A blurred photograph of part of an oak tree at a great distance is more difficult to interpret as an oak tree than a clear picture of an oak tree in sunlight at 20 metres. If all the positive stimuli are blurred, then

even a skilled oak tree-recognizer will not score very high on a forced choice task. For similar reasons, the success-rate will be lower if the negative stimuli are similar in appearance to oak trees, higher if the negative stimuli look very unlike oak trees. A person who is not good at recognizing oak trees will score as high as an expert botanist, if the task is simply to see the difference between oak trees and cars.

Being good at recognizing oak trees is a comparative attribute. It means being better than the average person. Here, 'average' is relative to a contextually given reference class. The average farmer in Europe is better at recognizing oak trees than the average computer programmer in Hong Kong. But the Hong Kong programmers and the European farmers are all thinking about the same category, and their judgements are subject to the same standards of truth and falsity.

In a set-up like ours, where S has the option of saying 'I don't know', rational subjects will adapt their responses to suit their skill, in the light of how difficult the stimulus materials are. An expert botanist will say 'I don't know' less often than a novice. Of course, people differ in their temperaments. Confident subjects are happy to take risks. Others are cautious. A cautious person will say 'I don't know' more often than a risk-taker even when the two have the same level of expertise. However, a rational person realizes that there are limits to caution, and also limits to confidence. A person who knows very little about how oak trees look, and who knows that he knows very little, should not make firm commitments on every trial. He should realize that he cannot always make a definite commitment. Learning this is part of learning how to judge. And it is part of what it is to possess the concept *oak tree*, or any other kind-concept. Concept-possession in general entails knowing how to manage your concepts sensibly, in a variety of contexts.

Suppose S learns, after his response on each trial, whether the object presented is or is not an oaktree. How might this feedback after his performance over the medium to long term? Suppose that the stimuli are clear and easily discriminable, and suppose that at time t, after a number of trials, S becomes aware that his responses have been correct only 70% of the time. S may reflect as follows: 'My success rate is rather poor. I may be under some misapprehension as to the diagnostic features of oak trees. I shall therefore modify the criteria I am using, and see if my performance improves'. S takes a gamble. By altering his criteria he runs the risk of a decline in performance rather than an improvement. But if he does get worse, and takes note of the fact, he will be free to revert to the original criteria, or to experiment with another modification. Any improvement in performance will cause him to retain the changed criterion. Over a long run of trials, and with a bit of luck, S's performance should improve. The constant trial by trial feedback gives him the opportunity to learn how to recognize oak trees better.

Yet the identity of the concept does not change. It is still the concept *oak tree* after 500 trials, just as it was on the first trial. The changes that occur are changes in S's criteria for applying that concept observationally. Suppose that some criteria for visual recognition are included in S's concept. S has a partly observation-based

concept of oaktrees. Then it is true that S's concept undergoes internal development, if his criteria of recognition change. But its referential content stays the same: it is still the concept *oak tree*.

Improvement in performance can go only so far. There will be an asymptotic level at which S is exploiting all the visual features which it is within his power to exploit. But remember, some categories have very few distinctive visual characteristics, and so the asymptote will be quite low for these. For example, powdered chalk, salt and magnesium chloride all look pretty much the same. A rational person regulates his practice according to how much he knows about the appearance of K's, but also according to how distinctive in appearance he thinks the category K objectively is.

Enough has been said to establish that there exists no unique behavioural profile, no particular performance-pattern, such that everybody who possesses the concept *oak tree* will exhibit just that pattern profile, on a series of open choice trials with feedback. What one would expect is that a rational person faced with this task will adjust his or her responses in the light of multiple considerations tailored to his or her own personal knowledge and skill.

IV. COGNITIVE CAPACITIES REQUIRED FOR JUDGING

What underlying cognitive capacities must S have, in order to make rational categorical judgements? The question seems to demand a top-down analysis of a very global ability. The whole capacity gets subdivided into sub-capacities, the sub-capacities in their turn get subdivided, and so on, until we arrive at a set of capacities that are taken as primitives. A full top-down analysis gives the structure of a hierarchy of capacities and identifies every node in the hierarchy. It is beyond the scope of this paper to attempt a complete analysis, but let us at least make a start.

As a first step, we isolate two components involved in S's general understanding that a response may be either correct or incorrect. What does it mean to say that an act of pressing a button in response to a visual stimulus is correct? The act is not correct in virtue of its intrinsic properties. Nor is it correct in virtue of the fact that it leads to a reward or to any confirming feedback. The outward act elicits that feedback only because the act is interpreted as the sign of a correct judgement. The judgement is an internal act mediating between perception and outward behaviour. The subject needs to understand, therefore, that it is the mental state which is evaluated as correct, in the first instance. The behaviour is correct only in a derivative sense. So the first component is an ability to isolate, to identify, that particular part of the whole cognitive process which is evaluated. S must be understand that the response was rewarded only because the judgement was correct. This component may be called 'ability to locate the primary object of evaluation'.

The second component is understanding *why* the judgement was favourably evaluated. S must know that the criterion of evaluation is simultaneous with the

judgement, and not forward-looking. That is, the correctness of the judgement does not consist in its being instrumental in producing good consequences. Rather the converse is the case: the judgement produces good consequences because it was correct at the time when it was made. Why was it correct at that time? An adult can answer this question by saying 'Because it matched reality. It corresponded to the fact that the object *was* a member of category K. In short, the judgement was correct in the sense that it was *true*'. 'It is hardly likely that a young child would produce such a sophisticated answer. Yet the child must, in my view, have some inchoate grasp of the notion of truth. To appreciate fully that the judgement's correctness consists in its corresponding to an independently existing fact, S needs to be able to think thoughts of the forms 'My recent judgement that o is a K was correct because (as I now learn) o really is a K', and thougths of the form 'My recent judgement that o is a K was wrong because (as I now learn) o is not a K'. Thoughts of these two require S to compare his previous judgement with his current updated judgement and to give precedence to the latter. To do this, S must remember his own earlier judgement. That is, S must represents the content of his earlier judgement as a content that he judged earlier, but without committing himself to it the second time round.

The two components both imply that S needs to be able to reflect upon his own representations. This capacity, in turn, can be subdivided into components. Also it admits of degrees. Conscious, explicit representations of one's representations would be the highest degree, but, as I shall shortly show, there exist lower degrees of reflective ability.

Reflection is some form or other lies at the core of rational concept-management. It plays a role, for instance, in the process of deciding to respond 'I don't know'. A rational decision to abstain is the upshot of thinking 'I don't have enough evidence to judge that this object is a K, but equally I don't have enough evidence to judge that it is not a K'. This thought is about the *relationship* between current evidence and a possible future judgement, and hence it is not available to any being who cannot think about possible future judgements. A child of two perhaps cannot do this. But perhaps a child of two can think about his own actual past judgements, ones that he remembers making. This is an empirical question. At the moment, since we engaged in a top-down analysis of the adult capacity, our next step ought to be separate the various components of the ability to reflect critically upon one's own actual or possible judgements. This ability is probably one aspect of a more general ability to reflect upon one's own desires, intentions, hopes, fears, and other first-order mental states. And self-reflection is probably just a special case of an even more wide-ranging capacity to think about mental states and representations of all sorts, not only one's own but also those of other people.

Suppose we call this most general capacity 'meta-representation'. It contains at least two components: the ability to form second-order representations, and the ability to do things with them in thought, comparing a meta-representation with a first-order representation and drawing some conclusion from the comparison. To analyse these two components would be an extremely interesting, exceedingly

intricate exercise, which would take a whole book. So I propose to curtail my analysis at this point in order to take stock of the situation and relate it to the case of young children.

V. EVIDENCE OF SUCH CAPACITIES IN CHILDREN

Top-down analysis is like reverse engineering. You take a system that works and you dissect it to find out how it works, what its components are, how they fit together. A description of the structure of a fully developed system is synchronic. But it has diachronic implications regarding the process of manufacturing such a system. When you do forward engineering, you will be well advised to adopt Herbert Simon's 'watchmaker' principle: 'First take the smallest components and build some small, easy to handle, sub-systems. Next join the sub-systems together. Don't try to build the whole system directly out of the smallest components, or you will get lost'. Nature generally follows this principle. So an analysis of what is involved in adult conceptual categorization does have developmental implications. If a high-level ability presupposes a set of lower-level abilities, we may infer that the lower-level abilities must be established at an earlier point in time than the high ones. After the low ones are in place, it becomes possible to synthesize the higher ability. Before they are in place, synthesis is impossible.

At what age do children start reflecting upon their own thoughts, and what stages lead up to this? Instead of working backwards let us now adopt a more natural chronological perspective and look at some of the relevant phenomena in the order in which they appear.

Activating Representations of Absent Objects

By 9 months, children can update their memory of the location of absent objects. Mandler (10), cites the example of the little girl who goes directly to the top drawer where she last discovered her ribbons, instead of to the bottom drawer where she knew they had been previously kept. Children of this age have a large store of memories about familiar objects in their home which they can update as necessary. These representations, when mobilized in the course of planning, are not confused with perceptions. They are activated not in 'perceiving' mode, but in 'remembering' mode. 9 months old children can process representations that are detached or 'decoupled' from input-processing *while they are processing input*. This simple fact already implies that there are at least two levels of processing going on simultaneously. Both are on-line, but only one of them operates directly upon sensory input.

Constructing Possible Worlds

According to Forguson and Gopnik (2), among the first phrases that English-speaking children learn (at around 12-18 months) are 'All gone' and 'There'.

The former is used when an interesting scene disappears from view (e.g. when there is no food left on the plate). The exclamation 'There' is uttered when the child successfully carries out an action that she had antecedently intended. It is an expression of satisfaction that things went according to plan. An intention framed in advance of action is, of course, a representation of a non-actual state of affairs. It may never become actual, and may not have been actual in the past. So by this age, children can construct possibilities, as well as remember past actualities. The utterances 'All gone' and 'There' suggest that the child is making a comparison between the currently perceived state of affairs and another represented state of affairs. With 'All gone', the latter is the state in which an object is present. With 'There', it is the state of affairs envisaged and desired. In the former case, the child recognises that the two states are not the same: the world has changed. In the latter, she recognises that they are the same, though they used not be. Again, the world has changed. It seems, then, that a one and a half year old can two representations entertained in different modes. She can extract information from a comparison carried out in her head.

Pretend Play

By the age of two, most children engage in 'pretend-play'. They construct imaginary scenarios within which they carry out routines on play objects which are normally performed on serious objects. The toy telephone rings, and the child picks it up and pretends to listen. Or in Alan Leslie's example, a banana is help up to the ear as if it were a telephone.

Leslie, who has studied pretend play in depth, believes that it provides the earliest evidence of meta-representation. In an experiment described in Leslie (9), he and the child jointly construct a scenario in which toy animals pour 'water' into cups, 'drink' from the cups, and play around with the 'water'. In reality there is no water. The experimenter takes a jug and pretends to fill two cups with water. He then picks up one of these cups and turns it upside down. He replaces the cup on the table and asks the child which of the two cups is empty. The child points to the cup that was upturned. The experimenter upturns the other cup onto the head of an animal, and the child is asked to explain what has happened. The child replies that the animal is all wet. This behaviour shows that the child not only constructs an imaginary world using real objects as props, but she also makes inferences about what is happening in that imaginary world, using general knowledge derived from her experience of the real world. For example, the child keeps track of the consequences of actions in the pretend world (the animal is wet), and at the same time knows that the animal is not really wet. As Leslie puts it, the child must be able to create two mental spaces, one for the real world, one for the pretend world, and in each mental space there are representations related to one another in a coherent way. In both spaces, the same causal laws hold and the same rules of logical inference apply. The child will neither judge nor pretend that a cup is both full and empty at the same time. A contradiction

is impossible in any world. But the child can represent the cup as empty, while at the same time entertaining a representation of the empty cup as being full. She can keeps both in mind at the same time provided the second representation is activated in 'pretend' space.

Of course, this talk of mental spaces is metaphorical. Although it will eventually need to be cashed out, in the present state of theorizing this metaphor, like many others, can be a useful tool for describing the child's mind. I think it offers us a good way to characterize the child's inchoate grasp of the fact that she has categorized an object incorrectly. In pretend-play, the child represents real objects as having certain properties in a 'pretend world'. This possible world is created inside its own dedicated mental space, with the same representational building blocks, be they models, images, prototypes or symbols in the language of thought, as are used in the space where the S represents the actual world. In rational categorizing, the child keeps a record of real world commitments that she has previously made so as to create a 'previous world', that is, a world of the facts as she has conceived them from her earlier point of view. The 'previous world' is kept inside its own designated mental space.

The space which contains her 'previous' world is separate from the space in which she holds representations to which she is currently committed. So there is room in her mind for the content expressed by 'o is a K' to appear twice. One token of that representation-type figures in her 'Then' space and another token of the same type appears in her 'Now' space. And normally there will be two such tokens of any given type that she has committed herself to, for her standard procedure is to make a copy of every judgement she makes and insert it in her 'Then' space, while retaining the original in her 'Now' space.

The distinctness of the two spaces does make it possible, however, for S to represent that o is a K in her previous world *without* representing this in her current world. When this happens, she detaches her previous commitment from her current commitments.

The child approximates to the adult meta-representational state of thinking 'I have judge that o is a K', in so far as she represents o as being a K from her previous point of view. This is like representing the cup as containing water from a pretend point of view. Doing this is not the same as mentally *denying* that o is a K; it remains open whether she is or is not committed currently to o's being a K.

S is normally loyal to her previous judgements. But individual judgements can be overridden. This happens when she responds in a certain way to the signal 'No' or 'Incorrect' immediately following an outward manifestation of the individual judgement in question. S interprets the signal as an instruction to retract her judgement. Such understanding is a complex disposition to engage in the following operations.

> (a) S defers to the 'authority' of the feedback signal. She gets ready to carry out some energy-consuming internal reorganization, in such a way as to make her representations conform better to a rule imposed from outside.

(b) S deletes 'o is a K' from her 'Now' space while retaining it in her 'Then' space.

(c) S inserts 'o is not a K' in her 'Now' space.

(d) S compares the token of 'o is a K' which is in her 'Then' space with the token of 'o is not a K' which is in her 'Now' space, and she notes their logical incompatibility.

(e) As a result of (d), S blocks an operation that she would normally be disposed to make, namely, moving a token of 'o is a K' from 'Then' space down to her 'Now' space. She is not willing henceforth to judge that o is a K, unless some new experience leads her to override the cancellation.

As well as making these changes to her current representational state, S changes the implicit criteria she will use to decide whether future objects are or are not K's. The set of features possessed by o, which she regarded as a sign that o was a K, will no longer be taken as a reliable sign of K-ness. If an object qualitatively identical to o is presented to her in the future, S is likely to judge that this object is not a K. The mechanism by which her criteria get returned is a matter for empirical investigation. But one possible mechanism is that S compares the look of new objects with the appearances of previously seen objects that she has judged to be K's. This assumes, of course, that she remembers the appearances of those individual objects, and also that she keeps a record of whether or not her judgement about them were confirmed. If the new object is more like objects in the 'confirmed' class than like objects in the 'disconfirmed' class, S categorizes the new object as a K. This way her criteria will change automatically as a result of any corrections she makes, because part of what she does when she corrects is to bring about a change in the composition of the two remembered classes.

It is crucial to the picture just sketched there should be a causal relation between step (d) and step (e). S blocks a normal functional property of her previous judgement because she registers that it contradicts her current judgement. That is how she manifests that she regards the previous judgements as false. Indeed, it could be said that her primitive conception of the falsity of á proposition p *consists in* her being reluctant to accept p for the reason that p is incompatible with a proposition to which she is currently committed. Note that this primitive conception of falsity allows her to attribute falsity only to one of her previous beliefs. It does not apply to her current judgements and beliefs. Falsity of current beliefs is an undefined notion for S. A two years old does not understand, cannot entertain the thought, that his or her own current beliefs might be false. Ability to combine the notion of false belief with epistemic modalities such as 'might' and 'possible' comes a good deal later, perhaps around age four, if my picture is right, the stage at which the child begins to understand what mistaken judgements are comes fairly early, around the same time as the emergence of pretend-play (two to two and a half, according to Leslie).

A great advantage of 'mental spaces' talk is that it expressly refrains from ascribing second-order representations. My hypothesis, for example, does not claim

that a two years old has thoughts of the form 'I used to think that o was a K'. To ascribe such thoughts would, I think, be to exaggerate the intellectual capacity of such a young child. First-order representations in mental spaces are a substitute for, indeed are precursors of, representations that have other representations embedded within them. The child has an *implicit* understanding that she used to think that p, just as she implicitly grasps that she is only pretending that the cup is full. 'Implicit understanding' is explained in terms of the functional roles of representations in different mental spaces and in terms of procedures that the child is prepared to perform upon these.

Later, perhaps at age 3 but this is controversial, the child starts to meta-represent. A natural explanation of this development would be that the child, who already has separate mental spaces starts being able to embed one space inside another. She keeps them separate while also letting them intermingle and interpenetrate.

If I am right that rational concept-management requires at least these two mental spaces plus procedures for manipulating representations within and across them, it follows that run-of-the-mill conceptual categorization of the physical environment is intimately linked with other mental operations which depend upon the creation and comparison of mental spaces. And there are plenty of them. Juggling mental spaces is required not only for pretend-play but also for representing time, alethic possibility, epistemic possibility, counterfactuality, potentiality, powers, the appearance-reality distinction, moral right and wrong, as well as for understanding other minds. All are interconnected: the roots of rationality coincide with the roots of knowledge of objectivity, morality and psychology. The mark that distinguishes conceptual categorization from other kinds of categorization is that a conceptualizer interprets the actual in terms of at least one alternative possible world, the world as he or she has (once) conceived it. The structure of such a mind is very different from the functional structure of a connectionist learning system, even though both systems, after a respectively appropriate number of training trials, come to reflect objective patterns that are present in the world.

I shall end by summarizing the main points cited in support of my claim that conceptual thinking is intimately connected to a rudimentary kind of rationality:

(1) S must realize that the primary object of evaluation is a mental state. This requires an ability to monitor one's own cognitive processes.

(2) S must be trained to defer to outside authority. To accept correction involves an ability to operate upon already existing representations, to dismantle internal structures that have been previously set up.

(3) S must notice a logical incompatibility and cancel a previous representation because of it. She updates the contents of her 'Now' space in such a way as to maintain consistency amongst them.

(4) S modifies her criteria for making future judgements in the light of lessons she learns from feedback. She makes use of remembered knowledge in the service of future categorizing. The procedure is not as simple as

back-propagation, where error-signals modify the connection-strengths between the units responsible for the most recent response.

WOODFIELD, A. Racionalidade nas crianças: os primeiros passos. **Trans/Form/Ação,** São Paulo, v. 14, p. 53-72, 1991.

RESUMO: Nem toda categorização é conceitual. Muitas das descobertas experimentais sobre o processo de categorização nas crianças e animais sugerem a hipótese segundo a qual os sujeitos formam representações perceptuais abstratas, modelos mentais ou mapas cognitivos que não são compostos de conceitos. Este artigo é uma reflexão acerca da idéia de que categorização conceitual *envolve a habilidade de fazer julgamentos categoriais, tendo como guia as normas de racionalidade. Estas incluem uma norma de busca da verdade e uma norma de evidência adequada. A aceitação dessas normas implica boa vontade em respeitar as autoridades cognitivas, o desejo de evitar as contradições e o conhecimento de como reorganizar seu sistema representacional após descobrir que se cometeu um erro. Sugere-se que a arquitetura cognitiva requerida pela racionalidade básica é semelhante àquela subjacente ao jogo do "faz de conta". O sistema representacional deve ser capaz de arrumar lugar para "espaços mentais", nos quais alternativas para o mundo real são consideradas. A mesma característica subjaz à habilidade de compreender modalidades, tempo, a distinção entre aparência e realidade, outras mentes e éticas. Cada área de compreensão admite graus, e o seu domínio (alcançado pelo adulto normal) leva anos. Contudo a manipulação racional de conceitos, pelo menos na sua forma mais rudimentar, não requer a capacidade de formar representações de segunda ordem. Ela requer conhecimento do procedimento de como operar e comparar os conteúdos dos diferentes espaços mentais.*

UNITERMOS: As raízes da racionalidade; categorização conceitual e não-conceitual; representações perceptuais; modelos mentais; mapas cognitivos; psicologia do senso-comum; capacidades cognitivas; representação de objetos ausentes.

REFERENCES

1. FOOR, J. *The modularity of mind.* Cambridge. MA: MIT Press, 1983.
2. FORGUSON, L., GOPNIK, A. The ontogeny of common sense. In: ASTINGTON, J. W., HARRIS, P. L., OLSON, D. R., ed. *Theories of mind.* Cambridge: Cambridge University Press, 1988.
3. GALLISTEL, C. R. *The organization of learning.* Cambridge, MA: MIT Press, 1990.
4. GIBSON, J. J. *The senses considered as perceptual systems.* Boston: Houghton, 1966.
5. GOULD, J. L., MARLER, P. Learning by instinct. *American Scientific,* New York, v. 257, n. 1, p. 74-85, jan. 1987.
6. HERRNSTEIN, R. J. Acquisition, generalization and discrimination reversal of a natural concept. *Journal of Experimental Psychology:* Animal Behavior Processes, Washington, v. 5, p. 116-129, 1979.
7. HERRNSTEIN , R. J. Objects, categories and discriminative stimuli. In: ROIBLAT, H. L. et al. ed. *Animal cognition*: Proceedings of the Harry Frank Guggenheim conference, 1984.

8. HERRNSTEIN, R. J., DE VILLIERS, P. A. Fish as a natural category for people and pigeons. In: BOWER, G. H., ed. *The psychology of learning and motivation*: New York: Academic Press, 1980.

9. LESLIE, A. M. The necessity of illusion: perception and thought in infancy. In: WEISKRANTZ, L., ed. *Thought without language*. Oxford: Clarendon, 1988. p. 185-210.

10. MANDLER, J. M. How to build a baby: on the development of an accessible representational system. *Cognitive Development*, Norwood, v. 3, p. 113-136, 1988.

11. MARR, D. *Vision*. San Francisco: W. H. Freeman, 1982.

12. SPELKE, E. S. The origins of physical knowledge. In: WEISKRANTZ, L., ed. *Thought without language*. Oxford: Clarendon, 1988. p. 168-184.