and "para-mechanical" psychological hypotheses. Ryle {52*} seems to fear that to begin to give way to any such hypotheses is to step right onto the slippery slope to Cartesian dualism. However, in large part thanks to Ryle, dualism is no longer the "official" theory of mind, and avoiding it no longer calls for quite such exaggerated caution. Ishiguro {53} argues that mental acts can (and must) be allowed into our theory of the mind without losing the ground which Ryle has gained for us. She is thereby able to redeem his theory of the imagination.

§II.D.2. Images and Eye Movements.

There is one respect in which seeing certainly seems to involve action. This is the matter of eye movements. Our eyes are continually in motion. They not only 'track' moving objects, but jump rapidly from place to place in so called 'saccades', which usually occur several times per second {1}. Even when we seem to be staring fixedly at some particular point, experiments show that the eyes still move slightly in at least three different ways: there are slow 'drifts' of the point of fixation; rapid 'microsaccades'; and a continual very fast and low amplitude 'tremor' which is superimposed on all the other eye movements {2}. The amplitude of the microsaccades varies from about 2 to about 50 minutes of angle, whereas the amplitude of tremor is only about 20-40 seconds of

angle with a frequency of about 70-90 oscillations per second {3}. These 'fixation' movements, which are made quite unconsciously, and are not under voluntary control {4*}, seem to be absolutely necessary for normal seeing. If things are so arranged that an object does not move relative to the eye, so that its image is always projected onto the same spot on the retina, then it soon becomes quite invisible {5}.

We can easily have a conscious awareness and some degree of voluntary control over the large saccades between fixation points, but most of the time they occur without any deliberate volition (quite a good analogy might be drawn with the sort of conscious awareness and voluntary control which we have over our breathing). Nevertheless, these saccades are not made either regularly or at random. They are always under cognitive control even when we are not aware of them. The Soviet investigator Yarbus {6} has shown that when somebody looks at a picture most of the saccades are between points likely to provide particularly significant information, where there are subsequent brief fixations, and the eyes repeatedly return to these significant points. Given a picture of a face, most saccades go to one or other of the eyes, or to the mouth; given geometrical shapes most saccades go to the angles. Also Yarbus found that the pattern of saccades and fixations on viewing a picture varied noticeably when the subjects were asked different questions about its contents. More recently, Bozkov, Bohdanecky & Radil-Weiss {7},

studying eye movements made whilst looking at random irregular polygons, have found that if the subjects interpret the polygons as being crude pictures of something then the eye movement patterns differ in characteristic ways from when the shapes are seen as meaningless. This only goes to demonstrate the well attested fact that, although they largely occur unconsciously, our eye movements are under cognitive control and are affected by our conscious beliefs and fancies. As we have mentioned, we are not consciously aware of the microsaccades and 'drifts' which occur during fixations, and we do not have any significant voluntary control over them. However, they still seem to be under the control of our cognitive system. Gaarder {8} found that the pattern of his subjects' microsaccades varied in characteristic ways according to the position of a stimulus pattern close to a point which they were asked to fixate. Our eye movements thus seem to be responsive to what we see even when we are consciously trying to keep our eyes still. It thus seems reasonable to class all saccades (and perhaps fixation 'drifts' too) as actions. In the rest of this section, however, we shall be concerned only with the larger, more easily observable saccades between fixation points.

We do not experience the world as jumping about before us in time with our saccades (like a home movie when the camera has not been held steady). Likewise, we do not normally experience the world as if looking down a narrow searchlight beam, with blur or darkness all around,

although the fovea (the central part of the retina, where the cone cells are most densely packed and where our vision is thus at its most acute by far) is only a very small part of the retina and therefore takes in only a very small visual angle (about 1.3° {9}). What we usually experience is a fairly broad and stable visual scene laid out before us. Psychologists have frequently taken it, therefore, that our percepts must be somehow assembled from the various brief and narrow visual impressions we receive after each saccade. In his very influential text of 1967, **Cognitive Psychology**, Neisser refers to this process as the "integration" of "visual snapshots", and he concludes (for this and other reasons) that visual scenes are "constructed" or "synthesized" in our brains, over time, from a succession of such "snapshots" {10}. Like the movement of the eyes itself, this "synthesis" seems to be best construed as an activity, albeit one which is performed quite unconsciously. This is the activity which the Neisser of 1967 takes to underlie the experience of imagery. He treats imagery immediately after his discussion of "Visual Snapshots and their Integration" {11}, and announces that:

> images will be treated here as products of visual synthesis. If visual cognition is an active and constructive process in "perceiving," when there is much relevant information in the retinal image, it must continue to be so when stored information is primarily involved. Imagery is not a matter of opening old file drawers, but of building new models. {12}.

Very similar views were expressed at around the same time by Hochberg {13} and by Hebb {14}. Hochberg argues that:

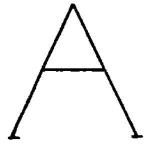
> The perceived structure of an object may consist of two separable components: (a) the features glimpsed in momentary glances, and (b) the integrative **schematic map** into which those features are fitted. {15},

and remarks that:

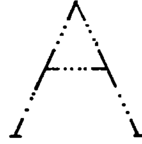
> If you try translating the above terms into the old words: momentary glance into "sensation," schematic map into "image," and perceptual structure into "perception," the fit is very good indeed. {16}.

Hebb states his theory in the context of his well known neurophysiological hypotheses {17*}, which need not detain us. But he is even more explicit than Neisser or Hochberg about the way in which percepts and images must be built up out of the contents of momentary fixations. If images were simply reinstatements of our original visual impressions then, according to Hebb, if our seeing of an object had involved, say, four fixation points, we would expect our image to consist of four impressions of parts of the object, superimposed on top of one another {18}. Clearly this is not what we experience. In fact, for Hebb, both our perception and our imagery should be construed as activities occuring over time, and "organized" by the "motor processes" of eye movements (saccades between fixation points):

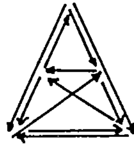
> When looking at the actual object each part-perception is accompanied by three motor excitations (assuming these four fixation points) produced by peripheral stimulation. One of them becomes liminal and the result is eye movement followed by another part perception. If the image is a reinstatement of the perceptual process it should include the eye movements (and in fact usually does); and if we can assume that the motor activity, implicit or overt, plays an active part we have an explanation of the way in which the part-images are



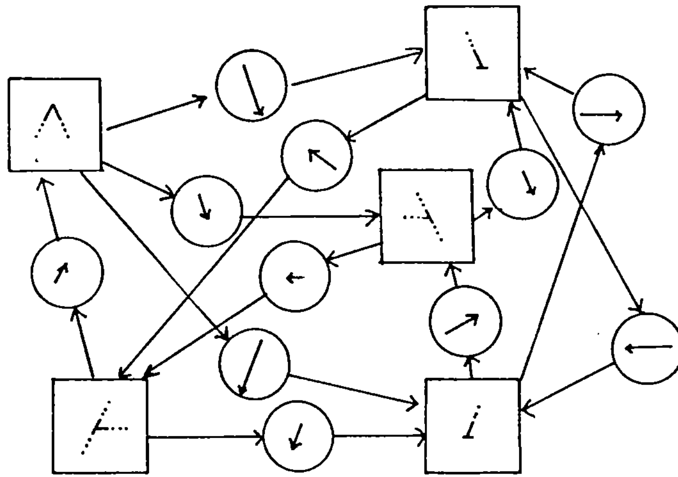
Simple pattern.



Features of pattern.



Some attention shifts between features of pattern.



Feature network for pattern.

Figure II.D.2_1
(Reproduced from Noton [1970].)

> integrated sequentially. In short, a part-image does not excite another directly, but excites the motor system, which in turn excites the next part-image. {19}.

A theory of perceptual recognition very much along the sort of lines which we have been discussing was proposed by Noton {20*}, and although he does not extend his speculations into the area of imagery, that short step has been taken for him by Janssen {21}. Noton's views are thus worth considering here, particularly because, working with Stark, he has been able to supply some rather direct experimental corroboration of his views. Noton's basic proposal is that a visual pattern is

> represented in memory as a network of memory traces recording the features of the pattern and the attention shifts required to pass from feature to feature across the visual field. {22}.

This notion of a "feature network" as the basis of pattern memory is illustrated in figure II.D.2_1, in which the circled arrows represent the directions of the "attention shifts". These shifts, Noton suggests, may or may not involve actual eye movements {23*}, but in either case he regards them as "a form of motor activity" {24}. Thus he is able to draw a strong analogy between seeing and ordinary behaviour:

> By comparing features with complete sensory situations and attention with conventional motor activity, it is seen that memorizing a pattern is analogous to memorizing a conventional sequence of behavior and recognizing a pattern is analogous to repeating this sequence of behavior. {25}.

In effect, seeing becomes a species of behaviour, an action.

Now in a pattern of any complexity, with more than a very few features, there are many possible patterns of attention shift, many orders in which we could successively examine the features. However, Noton {26} suggests that people are likely to establish specific "habits" in their 'visual behaviour', just as they do in ordinary behaviour. In particular, he suggests that they may establish particular personal "scanpaths" for each pattern which they learn, particular orders in which they traverse the network from feature to feature. This assumption, illustrated in figures II.D.2_2 and II.D.2_3, enables an experimental test of the theory to be made. If it is correct, and if visual attention shifts can be confined to eye movements, then it should be possible to record people's scanpaths and see if they do indeed become regular.

This experiment was undertaken by Noton & Stark {27*}. In order to force the subjects to move their eyes in their attention shifts, large, simple line drawings were used, which were illuminated rather dimly. Thus in order to discern a feature it would be necessary to bring it into focus right on the fovea, for which purpose the eyes would have to be turned. The results of the experiments were not entirely unequivocal, but they seem on the whole to support Noton's hypotheses. For most of the pictures presented most subjects seemed to establish a fairly regular scanpath consisting, on average, of about ten fixations. They would 'cycle' through this path for about 25% to 35% of the time

which they spent looking at the picture, but eye movements for the rest of the time were more irregular. When the subjects were later 'recognition tested' on the pictures the same scanpaths which they had previously established generally reappeared. As expected, the actual scanpaths established were specific to the subject being tested and the picture being used. These results are not unchallengeable - sameness of scanpath is rather a matter of judgement, and they appear less of the time than one might wish - but they do seem to provide some fairly direct evidence of the importance of 'motoric' attention shifts in vision and visual memory.

As already mentioned, the Dutch psychologist Janssen has extended Noton's theory to provide an explanation of imagery {28}, and support for such a move can be drawn from more recent experimental work done by Marks {29} and by Lai {30}, which seems to indicate that more vivid imagers not only remember pictures better {31}, but also, whilst performing such a picture memory task, they form more regular scanpaths, and with more consistency, than do less vivid imagers. Like Noton {32}, Janssen regards visual memories as composed of representations of particular, spatially localized features together with the directions of attentional shifts which have to be made between them - as a "feature network", as shown in figure II.D.2_1. Janssen's suggestion is that "the imagery **function** consists in traversing the feature network" {33}. What this seems to mean is that when we

imagine some remembered sight we somehow go through a sequence of attentional shifts similar to those which we found to be appropriate when the sight was actually before us. Of course, we will not find the right features in the relevant places. Presumably we must, as Ryle would have it, pretend, make-believe, that we find the features, and carry on going round the network regardless. Janssen is not very specific as to what he means by "traversing", but I take it he means something like this. Essentially this involves repeating the activity which took place during perception, but without the relevant sensory input which had originally produced that structure of activity. It should thus be apparent why I refer to theories of this type as "perceptual activity" theories of imagery.

Janssen points out that his theory makes it clear how we can rehearse images, maintain them before the 'mind's eye', by constantly going round the paths in the network. It is well known that we can retain verbal material in conscious short term memory by rehearsal, by repeating the words to ourselves, and that words quickly fade from consciousness if they are not thus rehearsed {34}. On Janssen's model we can see how images can likewise be rehearsed and thus retained in consciousness {35}. Watkins, Peynircioglu & Brems {36} have recently shown that such reahearsal of images (subjects were instructed to try "maintaining the picture before the mind's eye") improves the long term retention of pictorial material, just as verbal rehearsal improves retention of verbal material.

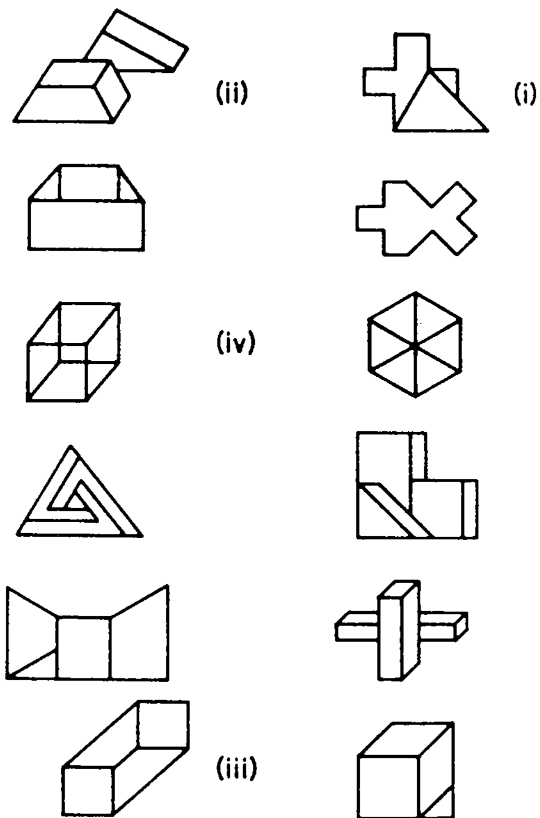
Janssen also notes that, although a particular habitual 'scanpath' may be favoured, there are actually many possible paths through a 'feature network', so that many different aspects of an image can potentially be experienced, and, if attention is confined to going round one particular region of the network, we may experience part images {37}.

At this point it may be appropriate to mention two schools of psychology which also treat perception and imaging as actions, and place a special stress on the relevance of eye movements. These are the schools of Piaget (sometimes called the Geneva school) {38} and of Soviet psychology {39}. I have neither the space nor the expertise to consider the views of these schools in anything like the detail they deserve, and I do not believe they have made a very great impact on the mainly American developments which have been the focus of our attention {40*}, but it seems relevant to at least mention them here. A useful summary of Piagetian and Soviet views on the rôle of perceptual activity in the formation of mental representations is provided by the French developmental psychologist Eliane Vurpillot {41}.

The work of Farley {42} has probably made even less impact on Anglophone imagery theories than has the Swiss and the Russian work. Nevertheless, it deserves our consideration because it fits right into the traditions of Artificial Intelligence research and 'information pro-

cessing' psychology with which we have been so much concerned. Farley was, in fact, a student of Alan Newell and Herbert Simon at Carnegie-Mellon University, and, like his fellow students Baylor and Moran {43}, he attempted a computer simulation of mental imagery as his doctoral research. Baylor's work, thanks largely to Pylyshyn {44}, became quite well known, and as we have seen {45}, can in some ways be considered as the Kuhnian exemplar for descriptional accounts of imagery. Moran's work, though less often discussed, is, if anything, even more paradigmatic of descriptional theories than Baylor's, since it firmly rejects the aspects of Dual Coding theory which Baylor had embraced. Farley's thesis {46} was the last of the three to appear, and where its existence has been noted at all it has been classified as simply another minor variant of the 'propositional' (i.e. 'descriptional') theory {47*}. I think this does it an injustice.

Inevitably, Farley's simulation does have strong affinities with those of his fellow students, and with the 'information processing' tradition in general. However, unlike Baylor or Moran, Farley is at pains to relate his imagery model to the processes of perception, and the theories of perception and imagery on which he draws are those of Hochberg and of Hebb {48}. In essence his program can be seen as a model of how eye movements are controlled so as to efficiently build up a perception of some pattern out of a succession of separate glances, and he regards the operation of these control mechanisms as equivalent to the



Stimuli of Experiment I, part 1. Four classes of line drawings: (i) two dimensional with overlay; (ii) three dimensional projections; (iii) three dimensional, impossible; (iv) three dimensional, reversing.

Figure II.D.2_4
(Reproduced from Farley [1976].)

having of a mental image. The simulations of both Baylor and Moran, it will be recalled, dealt with the solving of verbally presented puzzles, and relied on the introspective verbal protocols of people (actually the 'experimenters' themselves) solving the same puzzles in order to constrain the form to be taken by the computer's puzzle solving activity. No account at all was taken of perceptual processes, and really the only justification for calling these programs models of imagery at all was that human puzzle solvers typically do use imagery in these cases. Farley, for whom imagery derives from the control processes of perceptual activity, cannot be satisfied with such verbally specified tasks and protocols, and he thus turns to eye movements as the most accessible aspect of perceptual activity. The idea is that a machine which produces a similar pattern of eye movements to those made by a human when trying to identify a particular figure will have effectively simulated the imagery process.

However, it is beyond Farley's scope (as a computer programmer) to build and test such a machine. This would not only involve the fairly complex techniques of measuring human eye movements, it would also require the building of an artificial eye - a task which no-one has yet seriously dared to attempt. Farley thus settles for simulating a simplified, more manageable, model of eye movements {49*}. His subjects were presented with simple line diagrams such as those shown in figure II.D.2_4, but they could only view them through a small hole in a card which was laid in top

of them. At any one time, then, the subjects could see only one corner, junction or whatever of the diagram. However, the card could be slid about freely, to reveal fresh features through the hole. By doing this the subjects were soon able to determine the layout of the entire figure. This method of studying processes which are taken to be akin to ordinary eye movements was drawn from Hochberg {50}. The sequence of movements of the hole which the human subjects made whilst discovering the nature of the underlying diagram became the protocol which Farley's computer program was designed to simulate. Given the feature currently visible through the hole a successful simulation should be able to predict how a human subject would next move the card in order to efficiently ascertain the overall form of the hidden figure. In fact, there was no actual visual input to the computer program, and no actual movements were output. Instead the program operated on an internal description, a "vertex-feature list", of the relevant diagram, which was provided by the programmer {51}. Obviously this begs many of the most important questions surrounding vision and imaging. However, given that the diagrams used were so simple in any case, and especially that surrogate rather than real eye movements had been measured, perhaps this further simplification doesn't make things much worse. What we have here is not a model of a proper theory of imaging, but a sort of 'working illustration' of the general form which such a theory might take. Within these limitations Farley was able to get his program to predict human performances fairly closely. Of