II.B.

The Quasi-Pictorial Theory Today.

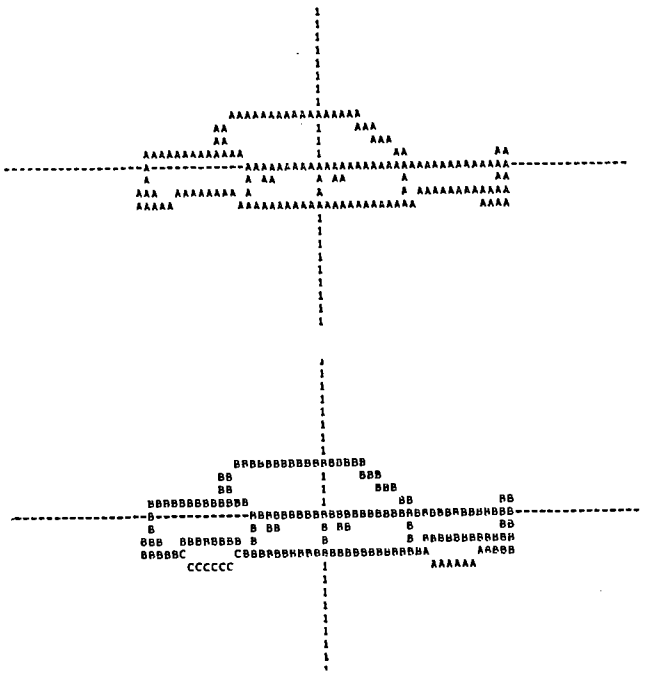
§II.B.1. Kosslyn's Model.

Stephen Kosslyn is by far the most important contemporary advocate of a quasi-pictorial theory of imagery - the theory in which an internal representation informationally equivalent to a picture is inspected and passed on to consciousness by a 'higher' process, a functional 'mind's eye' {1*}. Since 1973 he has published, with various collaborators, a very impressive amount of empirical research and theoretical discussion designed to support his version of this view. We have already discussed some of this experimental work in §s I.C.4 and 5. Here we shall outline the theoretical model which he takes his results to support, and then consider the empirical and conceptual problems which theories of this type seem to encounter. Consideration of his influential (and telling) criticisms of the rival 'descriptional' or 'propositional' theory will be postponed until the next chapter.

Kosslyn's model of mental imaging is perhaps the first, and certainly the most detailed, attempt to produce a quasi-pictorial theory consistent with materialism. As we saw above, Descartes's theory, which is probably the most detailed previous version of quasi-pictorialism, required an immaterial soul to inspect and be cognizant of the image

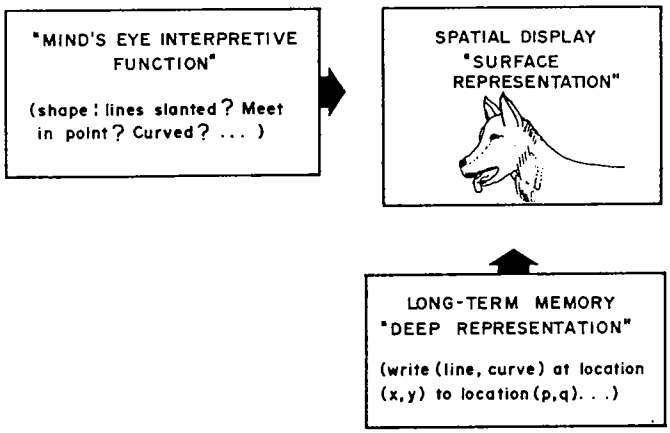
formed at the pineal gland. Kosslyn, however, is encouraged by modern researches in so called Artifical Intelligence (AI) to hope that such processes of consciousness will turn out to be explicable in computational terms; that is to say that once the "Artificial Intelligensia" have fulfilled their promise to show us how computers can have minds much as we do {2} then it will become plain just how neural tissue is also arranged to form a conscious mind {3*}. Kosslyn does not attempt to explain how this can be, of course - like the rest of us he is still waiting for the programers to deliver - but this hope, together with the formative influence it has had on the development of modern cognitive psychology as a whole, has determined that Kosslyn's theory of imagery shall be cast in computational terms. What Descartes explained in terms of strings and valves in the nerves, and the flow of animal spirits, Kosslyn explains not in terms of action potentials, synapses and neurotransmitters but in the more abstract and general terms of information flows and computations, so that the structure described can be claimed to be potentially realizable not only in a brain but also on a computing machine. And in fact Kosslyn has actually instantiated his theory as a running computer program {4*}.

In his doctoral thesis Kosslyn {5} suggested that imagery could be conceived of as analogous to the pictorial display produced by a computer on the cathode ray tube (CRT) at a graphics terminal, and he continues to use this analogy in his more recent work. Of course, he never meant



Two simulated images of a car: a "skeletal image" without added details (top) and an elaborated image (bottom).

Figure II.B.1_1
(Reproduced from Kosslyn [1980].)



A schematic representation of the cathode-ray tube (CRT) metaphor.

Figure II.B.1_2
(Reproduced from Kosslyn [1980].)

to imply that there is actually a lit up screen inside our heads, the analogue of the mental image should be taken to be not the actual display itself but the electronic information which specifies the display:

> The computer is able to interpret certain stored information as spatial images (whether or not it actually projects an image onto a CRT) because its CPU [Central Processing Unit] treats these data as if they were organized in a matrix; that is, these data function as if they were stored in a matrix, with some entries next to others, some diagonal from others, and so on. Thus, though the machine itself contains no actual screen, it can store and use material that is pictorial at the functional level. Our model suggests that the brain works this way too. {6}

Kosslyn's image is thus a 'functional image' in the sense of Rey {7} as, indeed, was that of Descartes (according, as we saw, to the **Dioptrics** {8}) {9*}. Such an image may quite fairly be considered as a two-dimensional picturelike layout, isomorphic to its objects projection, whether or not it is ever physically instantiated in this way. Kosslyn & Shwartz's {10} computer simulation treats image formation as consisting of the specification of light or dark points on a two dimensional grid or matrix {11*}, building up a picture somewhat in the manner of a newspaper photograph {12}.

Figure II.B.1_1 is an example of such a picture as produced by this program (it can also do a chair). The crudity of this picture should not, of course, be held against the underlying theory; with more computing time and more programing effort it could be made as naturalistic as you like. Furthermore, images are not

conceived of as being permanently stored in memory in this sort of form. Rather the information required to construct them is stored in long term memory (LTM) and the actual image, the "surface matrix" is constructed when required {13}. Again we can see a strong parallel to Descartes's theory from **L'Homme** {14}. Kosslyn has depicted the broad outlines of his model in the diagram of figure II.B.1_2.

Unfortunately this diagram fails to bring out the link between the imagery system and the associated perceptual system. In fact this link, which we noted in earlier theories, is quite properly maintained (it is, after all, more or less criterial of the having of mental imagery that it should be an experience in many respects very like actual perception). Kosslyn designates the locus of formation of his "surface images" as the "visual buffer", which is:

> a hard-wired, special-purpose, short-term, memory buffer that also is used in supporting the representations underlying the experience of seeing during perception proper. {15}.

Thus figure II.B.1_2 could perfectly well be amended to show an alternative possible input into the central box, the visual buffer, from the eyes. Kosslyn is attached to the widely influential "information processing" paradigm for psychology, in which perceptual information is conceived of as flowing inwards from the sense organs, passing through a sequence of computational "stages of processing" until it eventually reaches the consciousness in such a form that the things before us can be

recognised, and appropriate actions considered {16*}. Despite all the "processing" going on this can reasonably be called a passive theory of perception, in the sense that this computational processing is not itself subject to consciousness or the will. As Shiffrin points out, the widely accepted view of "visual information processing" is that:

> All or most of this process occurs automatically in accordance with the current structure of the processing system. {17}.

We may note the formal similarity of the "information processing" picture to the Cartesian theory of perception, with its similar passive {18} flow of information from the sense organs towards the centre. Of course, Descartes has no intervening 'computational' processing, but at least one contemporary "information processing" theorist of vision, Frisby {19} has acknowledged the general debt to Descartes. In the contemporary work the precise nature, number and order of the "stages of processing" are matters of much dispute, but perhaps the most sophisticated, and lately the most prestigious, visual theory of this type is that developed by the late David Marr {20} and his associates, to which Kosslyn {21} makes approving reference. Marr's model involves an informational structure, the "2-1/2-D sketch" which bears a loose resemblance to Kosslyn's "surface image". However, it is doubtful whether the differences between them could be eliminated without doing violence to the theoretical aims of one or other theory.

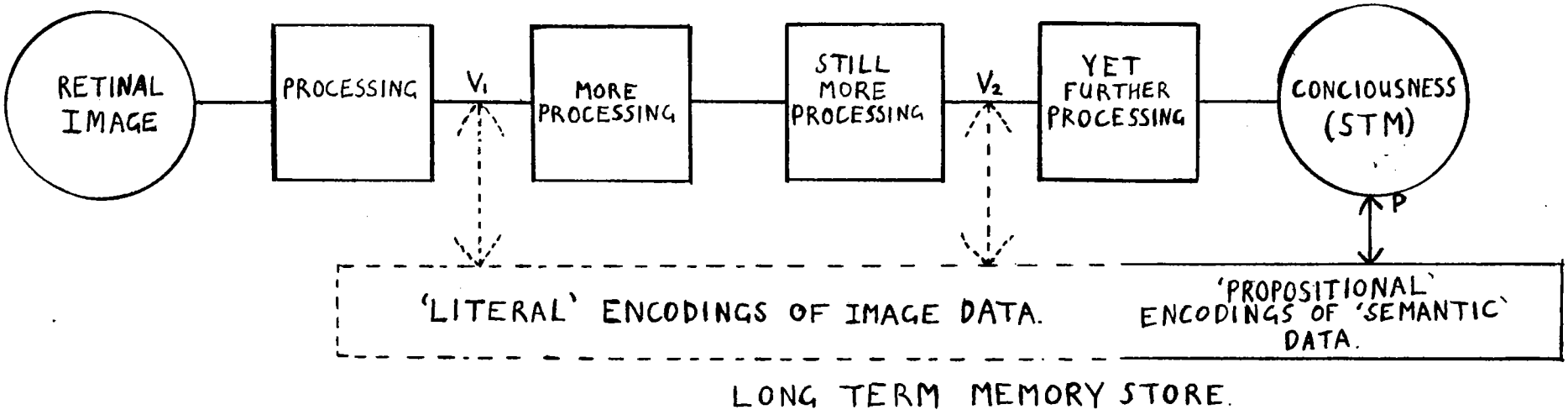


Figure II.B.1-3.
(loosely based on Neisser [1976 p.17].)

Solid parts indicate aspects more or less accepted by both modern 'quasi-pictorial' and 'descriptional' ('propositional') theories of imagery. Dotted parts are peculiar to 'quasi-pictorialism'. Arrows indicate directions of information flow. **V1** and **V2** are meant as **alternative** possibilities for the point where, on a 'quasi-pictorial' account, image specifying data might be abstracted from or inserted into the visual information processing stream, alternative sites for Kosslyn's "visual buffer". I believe that Kosslyn's theory is importantly ambiguous as to whether this is at an early or a late 'stage of processing'. P is the point at which descriptive 'propositions' are taken to store or reinserted into consciousness. According to 'descriptionist' theories their presence here is sufficient for experiencing imagery.

Kosslyn has in fact described the **essential** relation of his theory to "visual information processing" as follows:

> Consider the oversimple but not ridiculous view that vision is supported by a sequence of data structures starting with the pattern of receptor activation and culminating in a set of semantic or conceptual propositions describing what is seen. On this view, "pattern recognition processes" transform the more peripheral representations into increasingly central ones. The imagery debate {22*} (to oversimplify further) can be summarized as a disagreement over which of the data structures in the sequence support mental images (and consequently, which particular set of pattern recognition processes inspect images). At one extreme is the view (held by no one we can think of, and certainly not us) that images occur in the retina, and that processes such as contour enhancement and color normalization occur in imagery as well as in perception. At the other extreme is the propositionalist position that images are purely symbolic structures no different from those underlying abstract thought, and that **no** visual pattern recognition processes apply. Our position is that images are representations like those that occur in intermediate stages of visual processing, and that **some** visual pattern recognition processes can operate over them (e.g. detecting geometric shapes or parts of animals that do not receive explicit propositional encoding in long term memory). For example, the contents of the visual buffer may already be parsed into "Gestalt wholes" and interpreted in a perceptual sense, but not yet labeled or identified with semantic categories. {23}

Figure II.B.1_3 is a generalized diagram perceptual information processing theories adapted from Neisser {24*} but with additions (dotted) to incorporate related theories of imagery. Points V1 and V2 represent possible

alternative sites in in the processing sequence for the "visual buffer", whence partially processed visual data can be taken to store as a "literal encoding" {25}, and where, in imaging, this data can be reinserted into the normal 'visual information processing' flow. Alternatives are given because Kosslyn is not very clear - indeed, I shall argue he is fatally inconsistent - about at just how early or late a stage of processing mental imaging occurs. Point P shows where, according to the alternative "propositional" theories of imagery, fully processed (semantically interpreted) "propositional" information about some visual scene enters consciousness. We should note that Kosslyn entirely believes in this latter sort of representation and process {26}; what he does not believe is that this is sufficient to explain the subjective experience of, or the empirical findings on, imaging.

This should suffice us for the moment as a broad characterization of Kosslyn's quasi-pictorial theory of imaging. The actual Kosslyn & Shwartz {27} simulation (and its elaboration in later work) has numerous further features intended to model various empirical findings {28*} as to how images can be scanned, rotated, 'zoomed' in on, assembled from sub-images of parts, etc.. These effects are modeled in program modules with names like, respectively, SCAN, ROTATE, EXPAND, PUT and so forth {29}. Generation of complex images, such as the car of fig. II.B.1_1, are supposed to involve at least three such modules: PUT, which controls just where in the visual

buffer the image of a certain sub-part is to be placed; FIND, which keeps track of what is already in the buffer on PUT's behalf; and PICTURE, which actually translates the "literal encoding" of visual data in long term memory into the "surface image".

> These three procedures are coordinated by a procedure called IMAGE, which interfaces with the rest of the hypothesized cognitive system (i.e. a language comprehender, problem-solving apparatus etc.). In actuality, of course, IMAGE interfaces with the user [i.e. the computer operator], who specifies which images to generate and may indicate the size, orientation, location, and level of detail required. {30}.

In actuality also, the output of the program is printed pictures as in fig. II.B.1_1. The computer operator's fleshly eyes serve as that part of "the rest of the hypothesized cognitive system" which is the functional 'mind's eye'. There is no attempt to model this.

The most recent work of Kosslyn and his associates involves attempting to assign equivalent functions to the program modules of the computer module to particular sites in the brain {31}. Our concern, however, is with the viability of the basic conception of a functionally pictorial "surface image" formed in a "visual buffer" and inspected by a functional 'mind's eye'. If this were shown to be incoherent or empirically non-viable then the rest of the theoretical edifice would have to be abandoned or radically revised, and the empirical work it has inspired reassesed.