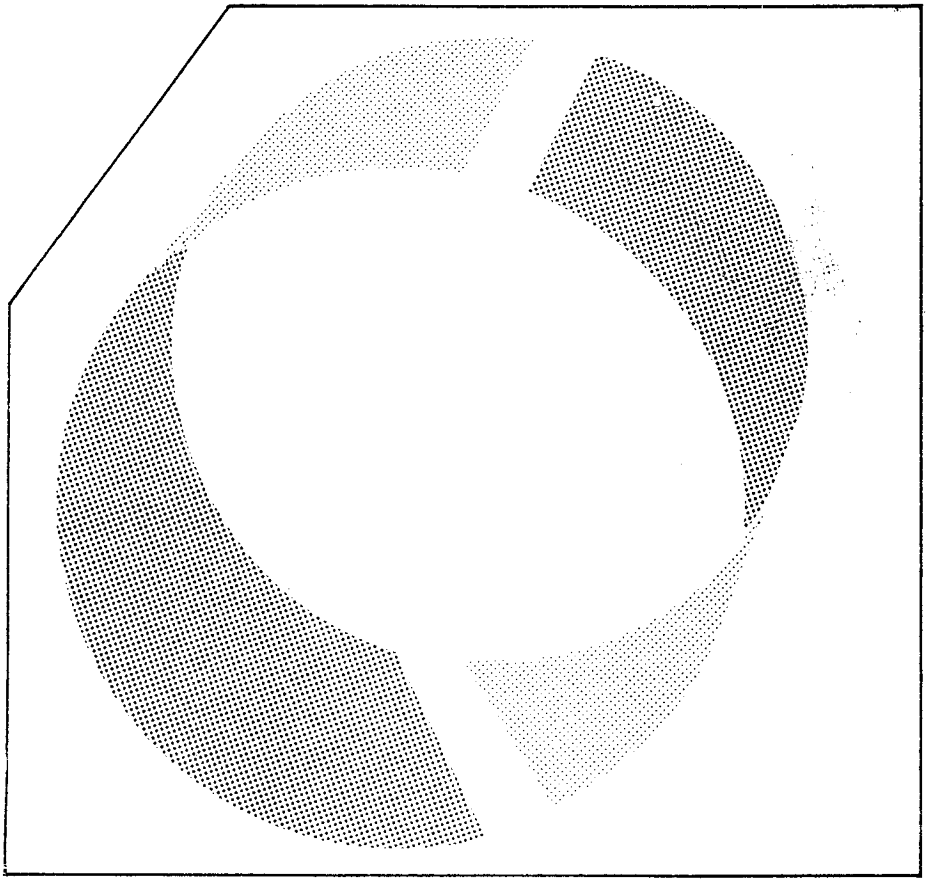
Volume 29, Number 1, 1996



EPISTEMOLOGY AND PSYCHOLOGY

Aleksandar Fatić (editor)

COMMUNICATION & COGNITION

Communication & Cognition
Vol.29, Nr.1 (1996), pp.37-64

AESTHETIC PERCEPTION

Jennifer A. McMahon
Australian National University

Abstract

In this paper I suggest ways in which vision theory and psychology of perception may illuminate our understanding of beauty. I identify beauty as a phenomenon which is (i) ineffable, (ii) subjectively universal (intersubjective), and (iii) manifested in objects as formal structure.[1] I present a model of perception in which I can identify a representation whose underlying principles of formation look like the sort of principles which might explain these features of beauty. The fact that these principles underlie the formation of the representation rather than constitute the content of the representation, provides an explanation for the ineffability and subjective universality of the perception of beauty in the form of objects. In part one, I set out two models of perception which I found to be suggestive of this approach. In part two, I suggest how such theories might offer a way of explaining and understanding beauty.

Part 1. Perception[2]

1.1. Introduction to Theories of Perception

According to the psychologist of perception, Irving Rock, 'perception can be distinguished from other modes of cognition such as imagination, dreaming or thought on the basis that perception 'is the mental representation of external objects and events that is based upon or in some way corresponds to the stimulation reaching our sense organs.'[3] The problem is, how do we mentally represent external objects and events? What are the processes involved? Rock sets the problem out succinctly in the following way:

> by either an implicit naive realism or an assumption that the proximal stimulus directly conveys to us the units that are isomorphic with phenomenal experience, we may fail to appreciate that perceptual

> organization represents an achievement based on some kind of internal processing.
>
> The fact is that the light rays from two separate points within an object have no more connection with one another than the rays from one such point with those from a point outside the object. Therefore, it is a problem why we usually group together into one phenomenal thing all the proximal stimuli related to one object rather than group some stimuli from the object with others from the surfaces between objects, or for that matter, it is a problem why such organization occurs at all.[4]

There are various theories and models of perception currently under debate. Two of the most dominant models of perception which are usually presented in opposition to each other are what are called symbol systems and connectionist networks. Both involve the idea that the perceptual apparatus in humans subtracts information from the distal array (objects in the world) which is then transformed by innately constrained mechanisms into the sort of information from which higher processes can infer knowledge about the world. In symbol systems, this process involves a series of ascending levels of representations which culminate in the information being subsumed under language related concepts (object recognition). In order to achieve such a model a symbol system needs to address: (i) the type of information processed by perception, (ii) the type of processes involved and (iii) the principles and assumptions underlying the processing of information. Now in a connectionist network there are no semantically analyzable ascending levels of representation. Instead according to such a model, perception is a spontaneous interaction of neural-like firings. Sensory units pick up information from the environment and compute a function that determines the output that is sent to higher layers of units. In some models the units are intended to function as a simplified neuron; in other models they are regarded as higher-level elements that do not correspond to neurons but are neuron-like or neurally-inspired.[5] The connections between units (the synapses in neural networks) are variously weighted by which means they determine the network's responses. That is, the weights are the means by which knowledge about a domain is retained in a network.[6] Connectionist models can provide new and illuminating ways to conceptualise certain aspects involved in perception. While I accept connectionist networks as more accurately capturing the workings of perception on a physiological level, and providing insight into how a particular perceptual process is implemented, I require a model of perception which provides an explanation of the function of perceptual processes. As connectionist models need the explanation of processes provided by symbol systems in order to construct their networks, they do not provide a symmetrical alternative to symbol systems but rather, it seems to me, straddle two levels of explanation, the physiological and functional. In this paper, I adopt the functional level of explanation of perception provided by symbol systems because it is at this level that the ideas about aesthetic perception which I set out in this paper have been conceived.

Now, there are many versions of symbols systems offered as models of perception. Some are designed with implementation on a computer in mind, while others are developed as psychological theories. Computational models of perception usually divide perception into low-level vision (often referred to as early vision) and high-level vision. Low-level representations are usually understood as solely the result of innately constrained rules operating on the incoming information. Higher representations in addition to such rules usually involve some degree of knowledge gained from past experience. Generally then, low-level vision addresses how sensory information is analysed while high-level vision involves the extent to which meaning and expectation influence and control that analysis or the product of that analysis.

Psychological theories concentrate more on the type of processes involved in perception, such as whether they can be said to be processes of description, problem-solving or some other type of process and what are the underlying principles and assumptions involved. As my aim in this paper includes establishing what could count as early visual representations which are accessible to consciousness, I adopt a model of vision largely based on two computational models of perception whose analysis of these early levels of visual representation suggest such a possibility. I have also drawn on the psychologist Irving Rock's theory to explain the principles and assumptions which underly many of the processes involved in perception. I do not set out Rock's theory in a systematic way as it closely resembles Marr's model as set out below. Instead Rock's theory is treated as a supplement to Marr's model, in some cases providing explanations as to what mechanisms or rules enable the perceptual system to achieve the grouping operations which are a key to Marr's model.

The models of perception discussed here while mainstream, are still controversial. However, they are plausible from the point of view of the neurological and anatomical evidence available and they have exhibited explanatory power from the point of view of cognitive psychology. The aspects of the following models which are most controversial are (i) whether we can speak of representations which are prior to the application or employment of language-related-concepts, (ii) whether such representations can be raised to consciousness[7], and (iii) to what degree the early levels of visual processing are uninfluenced by learning, expectations etc. Fodor points out that certain theories 'take the criterion of accessibility to be the availability for explicit report of the information that these representations encode.[8] The explanation of beauty I attempt to point to here both assumes and I hope may eventually illustrate the possibility that there are, to paraphrase Raffman, some sensory-perceptual states which have legitimate representational contents that are consciously accessible but not reportable,[9] and that some of these representations, particularly the ones that are of particular interest to me, have been formed without top-down influences.

1.2. Two Models of Perception: A Computational Analysis

1.2.(i) Fodor

I take as my starting point the model of perception developed by Jerry Fodor in his book *The Modularity of Mind*, in which he envisages the mind-brain as consisting of three types of system: transducers, input systems and central systems. In visual perception, the transducer is the retina which translates information from the environment into neural messages. Input systems translate these neural messages into a form which can be accessed by central systems. That is, input systems transform information about light intensity variations into information about depth, surface orientation, contours etc. As Fodor puts it:

> Whereas transducer outputs are most naturally interpreted as specifying the distribution of stimulations at the 'surfaces' (as it were) of the organism, the input systems deliver representations that are most naturally interpreted as characterizing the arrangement of *things in the world*.[10]

That is, the input systems would provide information about the 'layout' of distal stimuli[11]. Central systems manipulate the information furnished by input systems into meaningful constructs. Central systems, then, involve object recognition and refer to all the conscious and voluntary mental processes, and possibly to some unconscious and involuntary mental processes.

Input systems according to Fodor are strongly modular; that is, they are encapsulated (impenetrable to top-down or parallel flow of information), automatic and mandatory. Central systems, on the other hand, are global; that is, theoretically they can draw upon information from any domain for any particular task and 'the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system.'[12] Fodor acknowledges that in principle information which is specified at relatively high levels of information could be fed back to determine analyses at relatively lower levels. He points out that there are illustrations of the effects of information feedback in the psychological literature. For example, in the case of visual scotoma, where neurological disorders produce a "hole" in the subject's visual field, the evidence is that quite a lot of the visual input is masked without creating a phenomenal blind spot for the subject. According to Fodor: 'What happens is presumably that information about higher-level redundancies is fed back to "fill in" the missing sensory information.'[13] However, he argues that even if there are perceptual mechanisms whose operations involve feedback, there must be others which 'compute the structure of a percept largely, perhaps solely, in isolation from background information.[14] The reasoning which underpins this claim is that unless the input analysers are adequate to compute a representation of the stimulus from the information which transducers supply, then how else could we ever perceptually analyse unanticipated stimulus layouts ?[15] Fodor cites the Ames room and the Muller-Lyre illusions to illustrate that at least some of the background information at the subject's disposal is inaccessible to at least some of the perceptual mechanisms. Fodor argues that top-down flow of information which occurs within the input system module,[16] is not equivalent to top-down flow between systems. Furthermore, he points out that what is often interpreted as an example of context-driven prediction is, in fact, 'an effect of the biasing of post-perceptual decision processes.'[17] Hence the picture of perceptual processing we get from Fodor is bottom-up from distal stimulus to the interface between input processing and cognitive processing. At this interface top-down influences like beliefs etc, can and do determine how the 'primal sketches' furnished by input processing are interpreted.

My interest in Fodor's trichotomous architecture is that the isolation of perceptual analysis from the effects of background belief and top-down influence, has implications for the objectivity and hence universality of perceptually computed representations. Furthermore, if such representations are accessible, such a model could be employed in support of the notion of the ineffability of certain conscious representations. Fodor states that the rule seems to be that although

> perceptual processing goes from "bottom to top" (each level of representation of a stimulus computed being more abstractly related

> to transducer outputs than the one that immediately preceded), still *access* goes from top down (the further you get from transducer outputs, the more accessible the representations recovered are to central cognitive systems that presumably mediate conscious report).[18]

He adds that

> A plausible first approximation might be that only such representations as constitute the *final* consequences of input processing are fully and freely available to the cognitive processes that eventuate in the voluntary determination of overt behavior.[19]

Only those representations are *fully and freely* available which eventuate in the *voluntary* determination of *overt* behavior. Fodor, then, does not dismiss some sort of access to earlier levels altogether. In fact he qualifies the above further when he writes:

> it is inconceivable that such information [from low-level representations] is not registered somewhere in the comprehension process and, within limits, it is possible to enhance its recovery by the manipulation of instructional variables....
> it seem[s] plausible that the relative inaccessibility of lower levels of input analysis is at least in part a matter of how priorities are allocated in the transfer of representations from relatively short - to relatively long-term memory. The idea would be that only quite high-level representations are stored, earlier ones being discarded as soon as subsystems of the input analyzer get the goodness out of them. Or, more precisely, intermediate input representations, when not discarded, are retained only at a special cost in memory or attention, the existence of such charges-for-internal-access being itself a prototypical feature of modular systems.[20]

Given that the computational capacities of central cognitive systems are not inexhaustible in their ability to attend to impinging information, the special cost in memory or attention referred to by Fodor above could be the suspension of higher perceptual processes such as object recognition. Hence, Fodor's model presents the possibility of there being representations prior to the application of language concepts and yet accessible to consciousness.

Fodor suggests that the visual-input-system, bottom-up and modular though it is, also carries out the recognition of objects and events at a basic level, which

he calls, the basic-perceptual-categories.[21] He distinguishes these categories by their phenomenological salience which he defines as accessibility without sustained inspection.[22] These are the "middle levels" in implicational hierarchies. For example, the basic perceptual category in the hierarchy 'poodle, dog, mammal, animal, physical object, thing', would be 'dog'. Thus 'dog' is a basic categorization which is automatic and mandatory, and 'I see a dog' is the belief generated by the central systems. Whether this is or is not the case does not impinge on my enterprise here as long as such basic categories are not taken as the only representation from input processing which is accessible to us. The theory of aesthetic perception which I envisage requires that levels of representations prior to the subsumption of the data under basic categories or higher concepts are accessible, and Fodor's model generally does allow for this possibility.

There are other theories which explicitly treat high-level vision in humans as a higher operation involving the implementation of knowledge gained from past experience (learning), that is, beliefs etc. According to such theories, object-recognition which involves top-down flow of information is processed in high-level vision which is weakly modular, as distinct from low-level vision which is entirely bottom-up and strongly modular and this is the type of theory I will consider next.

1.2.(ii) Marr

To capture David Marr's model of perception[23], according to the issues raised in the previous section, one could envisage perception as consisting of a low-level visual module, which could be said to be strongly modular, and a high-level of vision which includes object-recognition and is penetrable to top-down flow.

Marr came to his belief in the modularity of vision through a consideration of the information provided by neurologists, physiologists, psychophysicists etc. For example, an account from clinical neurology which he discusses in his book, *Vision*, explains how the recognition of objects was affected in patients who had suffered left or right parietal lesions.[24] For those with lesions on the right side, the recognition of the object was possible if the view provided of the object showed its most characteristic angle. For example, such patients would be unable to recognize a water pail (bucket) seen end-on. When such patients did recognize an object they were able to name it, describe its function, its size etc. Patients with left parietal lesions, on the other hand, often had no language so could not state its name and function but they could convey that they correctly perceived its geometry (shape) even from the unconventional view. This suggested two things to Marr: first that a representation of the shape of an object was stored in a different place and was therefore quite a different kind of thing from the representation of its use and purpose; secondly, that 'vision alone can deliver an internal description of the shape of a viewed object, even when the object was not recognized in the conventional sense of understanding its use and purpose'.[25] According to Marr, vision tells about 'the illumination and reflectances of the surfaces that make the shapes — their brightnesses and colours and visual textures — and about their motion'. He emphasized, however, that vision's main function was 'to derive a representation of shape.'[26]

Marr adds to the definition of modularity as set out above in relation to Fodor's model of perception, when he writes that 'modularity states that any large computation should be split up into a collection of small, nearly independent, specialized subprocesses.'[27] Marr analyzes early vision into particular tasks according to the evidence provided by psychophysics and from the introspective evidence provided by everyday experience. He further argues in support of specialised subprocesses making up low-level vision on the grounds that

> if visual information processing is not organized in a modular way, incremental changes in its design, presumably an essential requirement for its evolutionary development, would be unable to improve one aspect of visual performance without simultaneously degrading the operations of many others.[28]

Now as the objective of many early visual computations is to describe the geometry of the visible surfaces, Marr concludes that the relevant information must be encoded in a shape's local surface properties. For example, a shape's local surface properties can provide information through stereopsis, shading, texture, contours and visual motion.[29] While the information processed by early vision is tied to the vantage point of the subject, the final step in vision is transforming this image into an objective representation of the three-dimensional shape and spatial arrangement. So Marr divides the derivation of shape information from images into three representational stages: (i) the primal sketch whose primitives are zero-crossings (virtual edges caused by intensity changes), blobs, terminations and discontinuities, edge segments, virtual lines, groups, curvilinear organization and boundaries, which all pertain to the geometrical distribution and organization of intensity changes and is therefore information about the two-dimensional image; (ii) the 2&1/2-D sketch whose primitives are local surface orientation, distance from viewer, discontinuities in depth and

discontinuities in surface orientation in a viewer-centered coordinate frame; and (iii) the 3-D model which uses a modular hierarchical representation whose primitives are volumetric and surface primitives, each one based on a spatial configuration of a few sticks or axes. This last representation describes shapes and their spatial organization in an object-centered coordinate frame.

According to Marr, early vision consists of two submodules: the primal sketch and the 2&1/2-D sketch. The output of the primal sketch feeds into the 2&1/2-D sketch. The primal sketch, itself a submodule, consists of two sub-submodules: the raw primal sketch and the full primal sketch. Each level throughout the hierarchy of representations is achieved by grouping operations formed on the previous level, capturing the spatial structure at that scale. These grouping operations entail selecting roughly similar elements from the raw material, grouping them together, 'forming lines, curves, larger blobs, groups, and small patches to the extent allowed by the inherent structure of the image.'[30] The first level at which these grouping operations are at work is the level derived most directly from the most peripheral level (the retinal image); the level referred to by Marr as the raw primal sketch. Jackendoff, in his summary of Marr's model, captures the status of the raw primal sketch. He writes:

> The full field of primitive elements derived from the retinal image constitutes what Marr calls the raw primal sketch. These elements are then subjected to computations that group them recursively into large scale elements to form the full primal sketch.[31]

The raw primal sketch filters the image at several spatial scales to find the intensity changes; these intensity changes furnish the bars and blobs (general patterns of lights and darks), and the zero-crossings (crossovers from light to dark) which provide information about edges. Marr writes that the

> raw primal sketch is a very rich description of an image, ... Its importance is that it is the first representation derived from an image whose primitives have a high probability of reflecting physical reality directly.
>
> Subjectively, you are aware of the raw primal sketch ... but you are not aware of the zero-crossings from which it is made.[32]

In normal perception you would not be aware of the raw primal sketch so presumably Marr means that such a sketch is available to consciousness given the right focus. For example, an artist might focus on aspects of a given scene in terms of the general pattern created by the cross-overs from lights to darks at a coarse-grained scale; that is, only noticing the virtual edges created by the more extreme shifts from dark to light. When shifting from this coarse-grained scale to a focus which involves a wider range of scales of intensity changes, it is as if the image comes into sharper focus just as it does when you focus with a camera, with the difference that as yet depth has not been registered so the shapes are just that rather than surfaces. The image as a pattern of broad flat shapes becomes a pattern of shapes within shapes. Thus the artist shifts from this overall coarse-grained view to one in which more detail (more edges) are discerned. At a level equivalent to the full primal sketch the artist might be focussing on the image as a pattern of detailed, but still flat shapes, arranged within a viewer-centred coordinate frame. Referring to this process as almost a recursive one, involving the build up of descriptive primitives, Marr explains that at each step, 'the primitives used are qualitatively similar symbols — edges, bars, blobs, and terminations or discontinuities — but they refer to increasingly abstract properties of the image.'[33] They are increasingly abstract in that the primitives are being increasingly categorized towards their subsumption under concepts. Imagined visually however, they are becoming less and less abstract. The full-primal sketch, then, carries out grouping operations on the raw primal sketch to create an image of edges at a higher scale, thus forming a representation which consists of a wider range of intensity changes. This amounts to more detail and more edges in the image. For example if you perceived a plant entwined around a post, at the raw primal sketch level you would have only the outer edges of the structure made up of the plant and the post. However, at the full primal sketch level which detects intensity changes over a wider range of scales, the different tone of the plant compared to the post would be registered and thus the outlines of the plant against the post would appear as well as the outline created by the post and plant against the background (the background not being registered as such but rather as just another shape juxtaposed to the plant and post on the same plane).

The 2&1/2-D sketch results from operations performed on the fullprimal sketch involving stereopsis (local depth), visual-motion (local depth) and shape-from-shading (local surface orientation). Marr's 2&1/2-D sketch explicitly represents three pieces of information about each point or small patch in the image: (i) the approximate distance of the point from the observer, (ii) the local orientation of the surface containing the point and (iii) the contours created by discontinuities in surface orientation or depth.[34] So at this level instead of the plant and post from our former example being read as flat shapes, the visual

system detects that the shapes of the plant and post represent surfaces which are closer to the subject than the surrounding area which is read as a more distant surface. According to Stillings summary of Marr's model, 'The 2&1/2-D sketch does not contain explicit representations of the shapes of surfaces or objects, nor does it contain descriptions or identifications of objects.'[35] The 2&1/2-D sketch represents the effect that judging the distance of each point from the viewer has on the full primal sketch. That is that instead of just outlines, the orientation of surfaces can be detected and thus the contours of surfaces which means that the image becomes not a pattern of flat shapes but a scene of variously oriented surfaces from the vantage point of the subject. In the artist's sketch then, this level would include the shading of the shapes which changes them from flat pattern into a scene like that just stated - of variously oriented but not necessarily unconnected shapes. The next level changes this subject-oriented view into an objectively-oriented view and in so doing, the image of variously oriented surfaces is transformed into a three-dimensional model of figure and ground.

In Marr's model, there is a one-way flow between early and high vision. Nothing contributes to the initial three-dimensional model apart from the internal operations of the early visual module. There is no influence from parallel or higher processes in any domain. The initial three-dimensional model is the interface between the strong module of early vision and the more global high vision. High-level vision draws upon various sources within, beyond and parallel to the visual processing, to recognize the three-dimensional model and in doing so fills in the details relating to specific shape and attaches meanings and associations to the object. This is the level of object recognition. According to Marr, object recognition is usually based on the formulation of a three-dimensional model from the information furnished from early visual processes. But this connection between the three-dimensional model and object recognition is not a one way connection. As Marr puts it, the

> two processes of construction [of the three-dimensional model] and matching [this model with a catalogue of stored models which is how he explains object recognition] cannot be rigorously separated because a natural aspect of constructing a three-dimensional representation may include the continual consultation of an increasingly specific catalogue of stored shapes.[36]

However, the initial three-dimensional model does precede object recognition providing information about an object's three-dimensional shape and its location in space. The sort of thing he has in mind for the three-dimensional model can be illustrated by a drawing of something reduced to a construction of cylinders, capturing the relative proportions and sizes of each part in a way which corresponds to the distal array. The main axis of each section is indicated also. In this way the three-dimensional model provides information which is not relative to any one viewpoint but rather is relative to objective coordinates in the distal array. From this basic model, in gradual steps, and through interaction with a stored catelogue of shapes, the visual system provides a recognized object. The psychologist Irving Rock posits as did Marr the interaction between these two levels (which he called the preferred percept and object recognition) to achieve a detailed description of the object. But Rock emphasized that top-down influences do not change the three-dimensional model but enriche it. He writes:

> Finally, the two- or three-dimensional description will generally lead to a description of what the object or event represents based on recognition, the interpretive mode of description. Here the interpretive description accompanies and enriches the object or event description but does not replace it.[37]

One could say then that up until and including this initial three-dimensional model, vision is modular in the sense that it does not require top-down flow of information.

Marr's theory of visual perception involves assumptions about the way the world is. For example, he reasons that assumptions like, "The world is made up of relatively few connected regions" and "Most objects are rigid" would be somehow built into the system presumably, I would suggest, at the three-dimensional model level. Consequently, we would not need to derive this information from each new stimulus.

Part 2: Aesthetic Perception

2.1. The Accessibility of Low-Level Visual Representations

The psychologist Irving Rock's theory of perception emphasizes the difference between object-recognition which he says is based on past experience, and two earlier organized percepts which correspond to Marr's 2&1/2-D sketch and his three-dimensional model. According to Rock's theory of perception, these earlier organized percepts are based on innate principles, innate assumptions and innate preferences,[38] and he claims they are both universal[39] and

accessible to consciousness.[40] Fodor, you will remember suggested that access to low-level representations was possible at some cost to the perceptual system. And according to Marr a level as early as the raw primal sketch was subjectively accessible. Interestingly, the processes used by an artist in representational drawing, from the general layout of pattern to developing three-dimensional objects to the intricacy of specific details, correspond to the ascending levels of representations entailed in normal vision as set out by Marr.

While the perceptual system according to the above models, utilizes the information available at a lower level and as in Rock's theory of perception 'engages in a sequence of hierarchically based levels of description each tending to supersede and dominate the lower levels of description'[41], according to Rock, this does not necessarily mean that these lower levels of description are inaccessible to consciousness. In fact Rock stated that while 'the highest level of description achieved is dominant or salient and the lower levels are in the background of awareness, the lower levels are nonetheless more or less simultaneously present.'[42] Presumably then, while earlier percepts are not ordinarily present to consciousness, we can access them. Later Rock reiterates the same claim in saying that

> the stimulus input that leads to a description of the object or event is in a certain sense introspectively available.... there are hierarchically based levels of description that occur beginning with one that is correlated with the stimulus, the proximal mode of description [Marr's 2&1/2-D sketch]. This suggests that the proximal stimulus information is available rather than being swallowed up, so to speak, by the processes that lead to constancy or world-mode levels of descriptions [three-dimensional models and object recognition] and that it can be internally scanned or examined.[43]

Rock explains that the proximal mode (the viewer-centred representations like Marr's primal and 2&1/2-D sketch) is difficult to access under ordinary circumstances because it is submerged by virtue of the dominant world mode.[44] Rock writes that the 'difficulty most people have in drawing three-dimensional objects or scenes, despite the fact that the task is simply to copy the retinal "picture", is precisely that that "picture" is relatively inaccessible because it is blocked by world mode or constancy perception.'[45]

2.2. Some Ideas Towards A Model of Aesthetic Perception

I wish to suggest that a model of the perception of beauty would involve locating a "percept" which is somehow dependent on the transformation of the proximal stimulus into a percept more related to the distal array, and yet located prior to object-recognition. Furthermore, the perception of beauty would involve not simply accessing such a representation but rather would involve the awareness of the principles underlying the formation of the percept. So the model of perception on which such a theory would be based must (i) include a level of perception which is accessible to consciousness but prior to the influence of top-down flow of information, and (ii) give some idea of the principles involved in this percept's formation.

A model of aesthetic perception might be developed along the following lines. We accept that low-level vision computes the type of information suggested by the Marrian model. That is, low-level vision converts information about intensity changes to information about edges to information about surfaces to information about three-dimensional models which can then be interpreted by higher systems as a particular object in a particular setting. This would mean that in most cases of normal vision top-down flow of information is only introduced at the level of object recognition. This is significant because according to our model, aesthetic perception is a corollary of normal vision.

Now the consciousness of any of the levels prior to object recognition requires the suspension of the conceptual knowledge (propositional knowledge) of the object. I suggest the suspension of higher processes rather than the arrested development of the normal perceptual process because aesthetic perception as it is developed here is a response to something in the stimulus and cannot, or is not usually, solely the result of a voluntary focus. As perception has evolved to perceive the function of objects and the spatial arrangement of objects in relation to the subject, it is probably mandatory that we recognize objects. For example, on initially looking (daydreaming usually excludes active looking) out the window onto a group of trees, say, we cannot help but perceive the tall, sinuous structure with somewhat brittle, feathery toppings as a tree. If we are going to attend to the sensuous features of the tree, this is necessarily after perceiving the configuration as a tree. Attending to the sensuous features of the tree (like colour, texture, quality of line etc.) does not constitute aesthetic perception in the model envisaged here. It may, however, be a necessary precursor to visual aesthetic perception. I will come back to this later.

The information provided by the early representations is the result of the interaction of innate mechanisms with the distal array (object out there). The processes involved in the formation of these representations involve, according to Rock, description, problem-solving and preferences for one solution rather than another. But as discussed earlier, all of these processes are nonconscious, nonverbal and based on innate categories and preferences. Marr's three-dimensional model automatically, then, arouses a matching type process in which the system looks for labels which match the percept. Once such a label is found, and this depends on information based on past experience, then the preferred percept is subsumed under a concept. Once this happens, we become aware of the details of the object which characterize the concept we have of it. In most cases the concept we have of an object is based on how the object functions in our lives. Thus, for example, we notice of a cup that it has a handle to pick it up with and is hollow so as to hold fluid, and has a smooth edge so that we can drink from it. Only in cases when the object displays features which arrest our attention away from this mode of recognition, or we have other distinct intentions in mind, do we deviate from this path. According to this model, then, the percept is universal until prior to the object-recognition level, because it is only then that the sort of personal experience which varies from person to person, influences the perceptual process. The only exception to this would be when not sufficient information is provided by low-level vision to form a three-dimensional model, in which case the system utilises higher information drawn from schemas of learnt information and assumptions about an object, in order to group the elements into figure and ground. In such unusual cases the formation of the three-dimensional model from the 2&1/2-D sketch may depend entirely on a top-down flow of information.

In normal cases of perception, certain principles and assumptions would seem to underpin each progressive level of representation. I am interested in those involved in the grouping operations which form the raw-primal sketch, the primal sketch, the 2&1/2-D sketch and the initial three-dimensional model respectively. According to Rock, up until the three-dimensional model (which he calls the prefered percept), the visual processes employed are simply a matter of description. However, to form the three-dimensional model from the 2&1/2D-sketch (which he calls the literal percept), the visual system introduces a problem-solving process which involves the application of innate preferences. These innate preferences determine how the information is formally structured, such as the relationship of the parts to the whole; for example, what constitutes figure and what ground. The principles presented by Rock as underpinning the formation of the representation at the three-dimensional model level are the following. (i) The Common-Cause Principle[46]. When the proximal stimulus is compatible either with a perception entailing no change or no event or with one entailing a change or an event, the former will be preferred. An example provided by Rock involves a single tilted rod rotating in depth about a vertical axis. The rod produces a sequence of images of varying length and orientation. After a while this sequence will produce the effect of a rigid rod of unchanging length rotating in depth, rather than the alternative of a rod spinning on its centre point on a flat plane, continuously shrinking then stretching then shrinking as it spins. (While this could explain the rule which allows the system to assume that objects are rigid, which is an assumption discussed by Marr, Rock points out that the Common Cause Principle covers more of the cases where this type of preference occurs.) (ii) The Coincidence Explanation[47], which applies the Common Cause Principle to stationary configurations rather than to events. This principle refers to the tendency to prefer and search for a common cause or single explanation of covarying stimulus transformations, correlated events, or apparent regularities in the organized pattern. That is, if there is a choice between two or more configurations, the principle is that the configuration which is based on the simplest explanation and an explanation which avoids accepting coincidence, will be preferred. This does not necessarily mean that the simplest configuration will be preferred, which is the idea of the good gestalt. For while the single explanation is preferred, the configuration resulting from the single or simplest explanation (avoiding coincidence) may be more complex than the configuration based on a complex explanation. Interestingly, according to Rock, 'the preference for the common explanation solution becomes greater as a function of the number of coincidental features' and I would suppose also greater as a function of the number of complex or contradictory features. Such organizing principles would most likely operate at the level of the three-dimensional model, when a coherent pattern is in the last stage of formation.

Now while the above principles are not by any means an exhaustive account of the principles underpinning the three-dimensional sketch, and in fact there are many more such preferences which the perceptual system executes according to Rock's and other psychologists' theories, I present them as an example of the sort of principles which might be involved in the perception of beauty. How this might work I will return to below.

In order to explain intellectual beauty one might need to assume that the rules and principles underlying thought are the same as those which underly vision. There are stories which could explain this. For example, the higher brain could have evolved largely through the visual interaction with the world. To

develop this line of argument is beyond the scope of this paper, suffice it to mention a few aspects of certain perceptual theories which suggest such an argument.

Rock, for example, argues that the units of description that constitute perception, that is, the basic units of perceptual knowledge, are based on the same categories as are the units of description of higher cognitive processes. Regarding the units of language or concepts used in perceptual description, Rock writes;

> These would have to be analogous to those we would employ were we to give the description in natural language. [For example, concepts such as "elongated" and "horizontal"] We must then assume that such concepts are therefore available in some internal lexicon. Whatever may be their origin and their nature, it seems plausible to me that they are the same as those that constitute the language of thought in general.... and which must preexist if a natural language is to be learnable at all - these same concepts may be those that are used in the process of describing forms.[48]

But, as he writes elsewhere;

> the kind of processing at issue [perceptual processing] is not verbal, not conscious, and not based on conceptually represented knowledge, ... the kind of intelligent processing we find in perception quite possibly has been the evolutionary link between phylogenetically earlier, less intelligent, more reflex-like sensory processes and the higher mental processes of thought.[49]

And also;

> Perception might be the evolutionary link between low-level sensory processes that mediated simple detection of environmental changes in phylogenetically primitive organisms and high-level cognitive processes in more complex forms of life. If the stimulus impinging upon sense organs such as the eye is at best an ambiguous and distortion-prone representation of the external object or object-event producing it, some mechanism had to evolve to yield reliable, veridical apprehension of such an object or event. One possible mechanism to achieve this end entails inferential processing. There-

fore, intelligent operations may have evolved in the service of perception. Once they emerged, they may have undergone further elaboration so as to become autonomous and no longer to be necessarily linked to sensory input.'[50]

While Rock clearly thinks the units of description are the same in perception and thought, the above quote suggests that the rules and principles underlying both perception and thought are also the same. Rock believes that such rules and principles are represented in some form other than natural language.[51]

Richard Gregory presents a similar argument but more specifically concerning vision. He writes that the eye is an organ of the brain and suggests that the mechanisms of the brain evolved based on how these mechanisms could aid perception. Gregory writes that, 'our most abstract thinking may be a direct development of the first attempts to interpret the patterns in primitive eyes in terms of external objects.'[52] A little later he writes that perhaps man's ability to respond to images (and absent imaginary situations) is the essential step towards 'the development of abstract thought'.[53] And elsewhere, 'The regions of the cerebral cortex concerned with thought are comparatively juvenile. They are self-opinionated by comparison with the ancient regions of the brain giving survival by seeing.'[54] Gregory claims that the brain would not have formed without an eye. An eye can give advance information and eyes freed the nervous system from the tyranny of reflexes.[55] Gregory's position regarding the relationship between perception and the higher centres is based on his understanding of the processes involved in perception. In a similar vein to Rock, Gregory believes that 'Perception involves a kind of inference from sensory data to object-reality. Furthermore, behaviour is not controlled directly by the data, but by the solutions to the perceptual inferences from the data.... So perception involves a kind of problemsolving - a kind of intelligence.'[56] Rock points out that such inferences need not involve learnt information etc., but as explained earlier, could rely on innate preferences and principles up until the level of object recognition. While Rock believes as does Gregory that the processes which underpin perception are the same type which underpin thinking, Rock makes an important distinction between perceptual knowledge and propositional knowledge which aligns his theory more with Marr's than with Gregory's. However, both argue for the same type of processes and principles underlying visual and intellectual processes.

Now to return to the type of organizing principles underpinning the three dimensional model. When one comes upon an object or intellectual construct

whose perceptual formulation involves the employment of the underlying principles in an unusual or unpredictable or unprecedented way, or in some way different to the average perception, and yet offers a configuration satisfying to the aim of such principles, then our attention might be arrested away from the conceptual grasp of the object, so that we become aware of the principles involved in the formation of the thing. The principles do not occur to us as such phenomenologically and cannot be reported because if we accept Rock's theory, they are stored in some form other than natural language.[57] Instead our inability to articulate what it is we experience leads us to the closest reportable claim which is that our response is due to the arrangement of the properties or qualities of the phenomena. In visual and musical beauty the phenomenological elements are the sensuous qualities and the effect of their particular arrangement. Intellectual beauty's phenomenology is perhaps the closest to reflecting its cause. The sort of criteria which physicists and mathematicians have attempted to draw up to define intellectual beauty, such critieria as symmetry, simplicity, order, coherence, unity, elegance and harmony[58], which while certainly not a list of sufficient and necessary conditions, do interestingly, on closer scrutiny of their meaning in an intellectual context, come closest to the sort of principles which psychologists of perception have developed to explain innate perceptual preferences.

Such an explanation of beauty might explain many of the characteristic features of our experience and understanding of beauty. For example, it might explain why logically 'beauty' is analyzed to be an emergent quality; that is, an indeterminate concept. Also, having assumed the universality, ineffability and association with formal structure as the characteristic features of beauty, such an explanation might vindicate such assumptions. For the apparent ineffability of beauty in all domains might be explained by the fact that beauty involves the principles which originate in processes prior to object-recognition; that is, prior to the subsumption of perceptual data under the labels which are the basis of language. Furthermore, beauty, when it is perceived, might be universal because the principles which give rise to the apprehension are universal. Phenomenologically, beauty often seems to be due to the formal structure of the object or other phenomenon possibly because the principles concerned are principles of organization. The fact that disturbing content is often taken as mutually exclusive to the apprehension of beauty does not contradict this model. Such content may stimulate the sort of engagement with the object which motivates a train of thought or action prohibitive of the suspension of conceptual engagement with the object. Furthermore this approach might explain why beauty can play an instrumental role in physics and mathematics if we consider perceptual mechanisms within an evolutionary framework. Perhaps the innate preferences embedded in our perceptual system reflect certain principles underlying nature. When a theory gives rise to the apprehension of beauty, the structure of the theory reflects a particular employment of innate principles which reflect principles in nature. In intellectual theories the actual ideas in the context of a particular construct can be experienced as having a sensuous quality; weighty, thin, intricate, delicate, robust etc. It is as though once the ideas are grasped intellectually, they can become an object of perception. Osborne offers an explanation for the instrumentality of intellectual beauty in the following.

> It has been argued with some plausibility that if certain conditions could be shown to have been necessary in the universe, in the solar system and on earth itself for the emergence of life and of human life in particular, then the fact that life did emerge is sufficient evidence that such conditions did in fact pertain. Similarly it might be argued that since human beings with powers of orderly thought have in fact evolved, and since the human brain is in close and intimate correspondence with thought, therefore the physiological organism must be well-ordered and for this to have been possible it must have evolved within a system manifesting physical order at least at an observable level.[59]

2.3. Conclusion

This is I admit a very rudimentary explanation of our experience of beauty. I was lead to this explanation of beauty by seeing whether a theory of perception could accommodate a universal, ineffable apprehension whose character explained why beauty phenomenologically appeared to be due to the structure of the object or construct. All I have attempted to do here is suggest how we might approach an explanation of our experience of beauty and what sort of processes might be involved. Hopefully there is enough here to show that cognitive science and psychology offer illuminating ways to approach the problem of beauty as set out in traditional aesthetics.

Notes

1. To illustrate the phenomenological salience of the above three features in our experience of beauty would be a paper in itself. Suffice it to say that I have derived these features from the issues arising from philosophical

discussions on beauty in nature, the arts and in science. The above were the most salient characteristics which were common to each type of beauty.

2. From hereon, 'perception' refers to visual perception.
3. Rock, 1983, p. 28.
4. Ibid, pp. 71-72.
5. Bechtel and Abrahamsen, 1991, p. 315.
6. Ibid.
7. For arguments against this possibility see the work of Daniel Dennett 1978, 1979, 1982, 1988.
8. Fodor, 1983, p. 56.
9. Raffman, 1993, p. 5.
10. Fodor, 1983, p. 42.
11. The proximal stimulus refers to the 2 dimensional arrangement of lines, shapes etc. which is how the outside world would be registered on the retina. The distal array refers to the three dimensional objects in the world which is the original stimulus for the proximal stimulus.
12. Fodor, 1983, p. 107.
13. Ibid, p. 63.
14. Ibid, p. 66.
15. Ibid, p. 68.
16. Ibid, p. 76.
17. Ibid, p. 77.
18. Ibid, p. 56 The italics used are Fodor's.

19. Ibid, p. 56.

20. Ibid, pp. 57-58.

21. Ibid, pp. 94-95.

22. Ibid, p. 96.

23. Marr, 1982.

24. Ibid, p. 35.

25. Ibid, p. 35.

26. Ibid, p. 36.

27. Ibid, p. 325.

28. Ibid, pp. 325-326.

29. Ibid, p. 36.

30. Ibid, p. 91.

31. Jackendoff, 1987, p. 171.

32. Marr, 1982, pp. 71-73.

33. Ibid, p. 91.

34. Ibid, p. 328.

35. Stillings in Garfield, 1987, p. 329.

36. Marr, 1982, p. 326.

37. Rock, 1983, p. 19.

38. Ibid, p. 91.

39. Ibid, p. 101.

40. Ibid, p. 92.

41. Ibid, p. 19.

42. Ibid, p. 19.

43. Ibid, pp. 39-40.

44. Ibid, p. 327.

45. Ibid, p. 328.

46. Ibid, pp. 134-138.

47. Ibid, pp. 138-164.

48. Ibid, p. 99.

49. Ibid, p. 338.

50. Ibid, pp. 1-2.

51. Ibid, p. 311.

52. Gregory, 1970, p. 31.

53. Ibid, p. 32.

54. Gregory, 1966, p. 224.

55. Gregory, 1970, p. 13.

56. Ibid, p. 30.

57. Rock, 1983, p. 311.

58. Engler, 1990, pp. 28-31.

59. Osborne, 1984, pp. 297-298.

References

ARBIB, Michael. 'Modularity and Interaction of Brain Regions Underlying Visuomotor Coordination', in Garfield, Jay L (ed), *Modularity in Knowledge Representation and Natural-Language Understanding*. Massachusetts; The MIT Press, 1987. pp. 333-364.

ARNHEIM, Rudolf. *Art and Visual Perception*. Berkeley and Los Angeles; University of California Press, 1974.

BAERGEN, Ralph. 'The Influence of Cognition upon Perception: The Empirical Story', in *Australasian Journal of Philosophy*. Vol.71, No.1, March, 1993. pp. 13-22

BECHTEL, W. & A. ABRAHAMSEN. *Connectionism and the Mind*. Massachusetts; Basil Blackwell. 1991.

BELL, Clive. 'Significant Form' in M. Rader (ed), *A Modern Book of Esthetics*. (fifth edition) New York; Holt, Rinehart and Winston, 1979. p. 287-297.

BODEN, Margaret. *Artificial Intelligence and Natural Man*. New York; Basic Books, 1977.

BURGE, Tyler. 'Marr's Theory of Vision' in Garfield, Jay L (ed), *Modularity in Knowledge Representation and Natural-Language Understanding*. Massachusetts; The MIT Press, 1987. pp. 365-382

CARMICHAEL, P.A. 'Aesthetic Knowledge' in *The Journal of Philosophy*, Vol.LVIII, No.14, July 6, 1961. pp. 378-387.

CHANDRASEKHAR, S. *Truth and Beauty*. (Aesthetics and Motivations in Science). Chicago and London; The University of Chicago Press, 1987.

DANTO, Arthur. *The Transfiguration of the Commonplace. A Philosophy of Art*. Massachusetts; Harvard University Press, 1981.

DEMPSTER, Douglas. 'Renaturalizing Aesthetics', in *The Journal of Aesthetics and Art Criticism*, Special Issue (Philosophy and the Histories of the Arts), Volume 51, No.3, Summer 1993. pp. 351-361.

DIRAC, P.A.M. 'The Evolution of the Physicist's Picture of Nature' in *Scientific American*. Vol.208, No.5, May 1963. pp. 45-53.

EKMAN, Paul. (ed.) *Darwin and Facial Expression* (A Century of Research in Review.) New York and London; Academic Press, 1973.

ENGLER, G. 'Aesthetics in Science and in Art.' in *British Journal of Aesthetics*, Vol.30, No.1, January 1990. pp. 24-34

FISHER, John. 'Beardsley on Aesthetic Experience' in J. Fisher (ed), *Essays on Aesthetics. Perspectives on the Work of Monroe C. Beardsley*. Philadelphia; Temple University Press, 1983. pp. 86-96.

FODOR, Jerry. *The Modularity of Mind*. An Essay on Faculty Psychology. Cambridge, Massachusetts; The MIT Press, 1983.

FODOR, J.A. & Z.W. PYLYSHYN. 'How Direct is Visual Perception? Some Reflections on Gibson's "Ecological Approach" ', in *Cognition*, 9 (1981). pp. 139-196

GHISELIN, Brewster (ed) *The Creative Process*. New York; The New American Library, 1952.

GOMBRICH, Ernst. *Art and Illusion*. (fifth edition). Oxford; Phaidon Press, 1977.

GOMBRICH, Ernst. *The Sense of Order*. (second edition). Oxford; Phaidon, 1984.

GOODMAN, Nelson. *Languages of Art*. (An Approach to a Theory of Symbols.) New York; The Bobbs-Merrill Company, Inc.,1968.

GREGORY, R.L. *Eye and Brain*. (third edition). London; Weidenfeld and Nicolson, 1977. First published in 1966.

GREGORY, R.L. *Intelligent Eye*. London; Weidenfeld and Nicolson, 1971.

GREGORY, R.L. (ed) *The Oxford Companion to the Mind*. Oxford, New York; Oxford University Press, 1987.

HARDY, G.H. *A Mathematician's Apology*. Cambridge; Cambridge University Press, 1948.

HEIL, John. *Perception and Cognition*. Berkeley; University of California Press, 1983.

JACKENDOFF, R. *Consciousness and the Computational Mind*. Cambridge; MIT Press/Bradford Books. 1987.

KANT, Immanuel. *Critique of Judgment*. (1790) Translated by Werner S.Pluhar. Indianapolis; Hackett, 1987.

KENNICK, W.E. 'Art and the Ineffable', in *The Journal of Philosophy*, Vol.LVIII, No.12, June 8, 1961. pp. 309-320.

KENNICK, W.E. 'Does Traditional Aesthetics Rest on a Mistake?', in F. Coleman (ed), *Contemporary Studies in Aesthetics*. New York; McGraw-Hill Book Company, 1968. pp. 411-427.

LANGER, Susanne. *Philosophy in a New Key* (A Study in the Symbolism of Reason, Rite and Art). London; Oxford University Press, 1951.

LEWIS-WILLIAMS, D. *Images of Another World*. BBC-TV Production, televised ABC: Melbourne, 5 August 1990.

MARR, David. *Vision*. New York; W.H. Freeman and Company, 1982.

MARTIN, M.G.F. 'Perception, Concepts, and Memory', in *The Philosophical Review*, Vol.101, No.4, October 1992. pp. 745-763.

MAYR, E. *Toward a New Philosophy of Biology; Observations of an Evolutionist*. Cambridge; The Belknap Press of Harvard University Press, 1988.

McALLISTER, J.W. 'Scientists' Aesthetic Judgements' in *British Journal of Aesthetics*, Vol.31, No.4, October 1991. pp. 332-341.

OSBORNE, H. 'Aesthetic and Other Forms of Order', in *The British Journal of Aesthetics*, Vol.22, No.1, Winter, 1982. pp. 3-16.

OSBORNE, H. 'Aesthetic Perception', in *The British Journal of Aesthetics*, Vol.18, No.4, Autumn 1978. pp. 307-316.

OSBORNE, H. 'Mathematical Beauty and Physical Science' in *The British Journal of Aesthetics*, Vol.24, No.4, Autumn 1984. pp. 291-300.

OSBORNE, H. 'Notes on the Aesthetics of Chess and the Concept of Intellectual Beauty', in *The British Journal of Aesthetics*, Vol.4, No.2, April 1964. pp. 160-163.

RAFFMAN, Diana. *Language. Music. and Mind*. Cambridge MA; A Bradford Book/The MIT Press. 1993.

ROCK, Irvin. *An Introduction to Perception*. New York; Macmillan Publishing Co., Inc., 1975.

ROCK, Irvin. *The Logic of Perception*. Massachusetts; The MIT Press, 1983.

SELFE, Lorna. *Nadia. A Case of Extraordinary Drawing Ability In An Autistic Child*. London; Academic Press, 1977.

SELFE, Lorna. *Normal and Anomalous Representational Drawing Ability in Children*. London; Academic Press, 1983.

SHARPE, R.A. *Contemporary Aesthetics. A Philosophical Analysis*. Great Britain; The Harvester Press, 1983.

SIBLEY, F.N. 'Aesthetic Concepts', in C.Barrett(ed), *Collected Papers on Aesthetics*. Oxford; Basil Blackwell, 1965, pp. 61-89.

SIBLEY, F.N. 'General Criteria and Reasons in Aesthetics' in J. Fisher (ed.), *Essays on Aesthetics. Perspectives on the Work of Monroe C. Beardsley*. Philadelphia; Temple University Press, 1983. pp. 3-20.

SNYDER, A.W. and M. THOMAS. 'Autistic Artists give Clues to Human Cognition.' Paper presented to the *Robertson Symposium*, ANU, 1993.

STILLINGS, Neil. 'Introduction' to The Visual Module, in Garfield, Jay L (ed), *Modularity in Knowledge Representation and Natural-Language Understanding.* Massachusetts; The MIT Press, 1987. pp. 325-332.

STILLINGS, Neil. 'Modularity and Naturalism in Theories of Vision', in Garfield, Jay L (ed), *Modularity in Knowledge Representation and Natural-Language Understanding.* Massachusetts; The MIT Press, 1987. pp. 383-402.

TOPPER, D. 'The Parallel Fallacy: On Comparing Art and Science.' in *British Journal of Aesthetics,* Vol.30, No.4, October 1990. pp. 311-318.

WEINBERG, Steven. *Dreams of a Final Theory.* London; Vintage, 1993.

CONTENTS