Visual Concepts

Mohan Matthen
University of Toronto

I. Sensory Classification

The visual system classifies objects: that is, it computes what (perceptual) kinds or classes objects in the environment belong to. For example, it processes retinal data to figure out the shape of a given object: whether it is square, diamond-shaped, elliptical, round, etc.1 This activity of the visual system is not merely a matter of its being differently affected by different kinds of things. Thus, the visual system is not a classifier simply in virtue of its states carrying *information* that differentiates one object from another. Rather, its classificatory activity manipulates visual information, and subserves specific knowledge-gathering activities in which organisms engage.

Think of a data-base maintained by an organization to guide some aspect of its operations – a telephone company's billing records, say. Let's say that the data-base is updated every time a customer makes a call. The signal received from the customer would contain many kinds of information: a voice stream, interference indicating the condition of the line, and so on, and embedded among all this information a code that identifies who is calling. This last item is what the phone company needs for billing purposes. Suppose that it is all that concerns the company. It would be highly inefficient to provide clerical staff with the whole signal, leaving it to them to figure out who called and when. So the company will use a customer interface that is designed to extract this information from the incoming signal, and disregard the rest. The billing data-base itself will be organized in terms of specific categories relevant to its own operations – who took what services, how they pay, where they live – but not what they say, whether their accents are domestic or foreign, and other irrelevant pieces of information that *could* be extracted from the incoming signal. In other words, the company's billing interface will ideally be designed to extract information relevant to billing, to ignore the rest, and to blend this in to information that it already has on record. On the other hand, a customer's interface – a telephone – will be designed to extract the voice-stream and ignore the rest.

Analogously – though with the difference that they tolerate a certain amount of error – *organisms* gather and keep records concerning their environments. These records do not consist merely of information that happens to impinge on their skins – of "impacts at our nerve endings," as Quine once put it. The knowledge-gathering operations of an organism are organized with reference to what the organism needs to do in its environment. The visual and other sensory systems are designed to analyse incoming information from this point of view. Just as the phone company's billing interface classifies incoming signals by source-customer, sensory systems classify the objects they confront for the organism's knowledge-gathering purposes. The records that an organism keeps are organized and deployed in terms of proprietary categories specific to its environment and style of life. Correspondingly, its sensory systems are designed to analyse incoming information in terms of these proprietary categories. A bird's environment and style of life are very different from those of a mouse. Consequently, the visual features processed by a bird are different from those processed by a mouse. In short, what a bird sees is not the same as what a mammal does (cf. Matthen 2005, chapter 7). Similarly, an animal's visual sub-systems process incoming information for special purposes: the human face-recognition system serves specific functions different from those served by the visual startle system.

II. Sensory Experience

Having classifying visible objects in this way, a sensory system enables epistemic operations pertaining to that object. In the case of shape, color, and other consciously available visual qualities, it does this in a rather simple way. Having determined what shape, what color, etc. an object is, the subpersonal2 visual system ensures that the visual perceiver has *personal* access to its determination. To put it in another way, it provides the organism news of how it has classified the object, leaving it to the perceiver to employ this "news" in further knowledge-gathering operations of its own. The news comes to the perceiver in the form of visual experience. Visual experience, or consciousness, or sensation, is, in other words, the "code" that the visual system uses to "tell" a visual perceiver how it has classified various objects.3 Sensory qualia encode sensory features. The square *look*, one might say, *encodes* the square kind, or *denotes* it: it indicates that *according to the visual system* the object in question is square. (Of course, the perceiver may not accept the determination of the visual system.)

The process described in this and the preceding section can be schematized as follows:

1. *Reception* Light from the object reaches eyes and affects receptors in the retina.
2. *Classification* By processing receptoral activity-arrays, the visual system sub-personally sorts objects into proprietary kinds (colors, shapes, etc.).
3. *Access* By issuing a characteristic visual experience, the visual system gives the perceiver access to the results of its classificatory activity.

In this schema, visual experience is the record of the visual system's classificatory activity – an object looks green because the system has determined it to be green and is informing the perceiver of this. Similarly, it looks round because the visual system has determined by a distinct process that it is round.

By combining various "atomistic" determinations of step 2, the visual system provides us with the rich visual phenomenology with which we are so familiar. The act of combination is a distinct and culminating operation:

4. *Binding* The perceiver's attentive visual experience of a scene or part of a scene is a composite of visual experiences that result from specialized sorting activities of step 2.

An object looks round *and* green because it looks round and it looks green for the above reasons, and the visual system has "bound" these qualitative states together.

This account is, leaving out the first step, virtually the reverse of that envisaged by Rudolf Carnap (1928), who took a Kantian/empiricist approach to visual phenomena. For Carnap claims that experience is indivisible except by an act of abstraction performed not sub-personally, but by the perceiver. He is assuming that visual experience is merely a trace of the environment's effect on a perceiver's sensory organs. Thus, Carnap *starts* with total experiences of the kind that step 4 yields, and gets the experience of color or shape by an act of abstraction therefrom. In his view, such experience is raw and unanalysed: the job of assigning it to visual kinds like *round* and *green* is left to the perceiver, as is the construction of the kinds themselves from past experiences.4 The present proposal is that, on the contrary, kinds like *blue* and *green* (as well as similarity-orderings) are sorting categories used by the underlying visual system: visual experience is news of classification, and presupposes these prior sorting categories, since without classification into these categories there would be no news to convey. Thus, the view articulated in this paper explicates the *green-look* in terms of *green*, the latter being a category generated by the visual system.5

The propositions articulated as 1-4 above are a version of what I have elsewhere called the *Sensory Classification Thesis* (Matthen 2005, Part I) . The Sensory Classification Thesis (*SCT*) is a natural extension of the commonly held views (a) that neuronal processes in the visual system should be regarded as a form of computation, or data-processing, and (b) that consciousness provides us with access to the results of underlying computational data-processing – (b) is the view implied by saying, for instance, that consciousness is a "global work-space" or that it is "higher-order" thought. However, it is an extension of these theses that has seemingly eluded most philosophers who work from this point of view. Most philosophers continue to define sense-features in terms of experience – this is the common element in dispositionalist, projectivist, and relationalist accounts. But *SCT* gives a reason for reversing the order of dependence. For if sensory experience is simply news of a preceding act of assigning objects to sense-features, it is an inversion to suppose that these features can be defined in terms of sensory experience.

In the next section, I sketch some evidence for *SCT*. I will then consider some indications against it brought to the table by those who argue that sensory experience has only "nonconceptual content". I take it that these philosophers would oppose *SCT* – at the very least, proposition 2 above – and I will attempt to respond to their concerns.

III. Sensory Specialization

1. The traditional view is that the visual image – i.e., that of which we are visually aware – is (a point-by-point optical projection of the scene beyond.[6] From the nature of this projection, the perceiver is able to infer certain features of this scene. However, the image itself has no representational properties, according to this view, no "semantic content" – it is not "accurate" or "truthful". The visual image is supposed to be similar in this way to a shadow cast by an object: it reflects properties of its source, and carries information about it, but is not a representation thereof. Suppose, for instance, that the shadow is round. One may infer the roundness of the object from the roundness of the shadow. But the shadow does not literally *mean* (cf. H. P. Grice's 1957 non-natural sense of 'mean') that the object is round, nor is it appropriate to call it "accurate" or "truthful" with regard to the object. If it should turn out, because of the curvature

of the surface on which the shadow falls, that the object is *not* round, it makes no sense to say that the *shadow* was in error. So also, on the traditional view, the visual image has no semantic value.

Classifying brings an object under a general intension. The traditional view interprets the visual image as wholly concrete; it claims that there is no room in visual experience for a classificatory component or, indeed, for anything like a conceptual (or for that matter, non-conceptual) *representation* of how the world is. *SCT*, on the other hand, claims that visual experience subsumes objects under general concepts: for example, one's visual experience of a table assigns it to certain color and shape classes. (Here, I am using the term 'concept' to mean something like 'universal'. A concept is simply something under which more than one individual can fall or be subsumed. It is *not* meant to imply the involvement of "judgement" or "reason".)

It is now widely (though not universally) accepted that the traditional view is mistaken,7 but the point is generally made by reference to philosophical intuition – it is claimed that contrary to the no-content view, it makes perfectly good sense to characterize sensory experience as true or false. Here I will try to make the point in a different way, by reference to sensory specialization. This phenomenon – sensory specialization – cannot, in my view, be interpreted in consonance with the no-content view. This is an example of how empirical phenomena can tell against philosophical theories (which proves that the latter cannot be wholly a priori).8

2. I begin with the consideration of a non-visual sensory phenomenon, the perception of communicative signals emitted by other members of the same species. There is a whole raft of issues that surrounds this phenomenon: the cost of producing signals, the "truthfulness" of these signals, the connection with "mind-reading", etc. (See Hauser 1996, chapter 2, and Sterelny 2003 for discussion.) Here, I am concerned just with one issue: the species-specialization of the perception of communication. My claim is that this is (at least) highly suggestive of sensory classification.

The crucial facts about *human* speech specialization are, very briefly, these.9 First, we utter a certain number of distinct units of speech, phonemes, which are strung together to make *words*. We recognize that two spoken utterances are different semantic units when they differ with respect to some constituent phoneme; conversely, the same sequence of phonemes marks a single word, even when the utterance differs in other respects – pitch, intonation, etc. Thus, even those who cannot speak English, can (at least in ideal circumstances) recognize that 'bit' and 'sit' are different words (and are thus available to a speaker or a linguistic community to denote different things) because they differ from one another by the substitution of a different phoneme. In this way, the comprehension of meaningful speech depends on phoneme recognition (just as the production of meaningful speech depends on the *production* of phonemes). Phoneme recognition is performed by a discrete part of the auditory system.

Many species have signalling systems. Generally speaking, the production and auditory recognition of communicative signals operates in a species-specific way: only members of the same species are able by instinct to produce or perceive the particular sounds that a given animal makes. So it is with human speech communication. Most humans learn in early infancy to recognize the phoneme set to which they have been exposed, generally a large subset of the phonemes that are produced in all human languages. Non-human species cannot distinguish or produce these phonemes10; at the same time, humans cannot (without a great deal of effort) recognize or produce the calls, cries, and songs of other species. Humans, on the other hand, are easily able to learn to distinguish the phonemes to which they are exposed as infants (even, as it happens, when they are exposed to these phonemes in visually encoded form, as in sign-language).

Now, here is a surprise. Phonemes are not *acoustic* patterns (especially not in sign-language, of course, but not in acoustically articulated languages either). For the features that characterize different utterances of a given phoneme are quite diverse from an acoustic point of view. For example, the pattern that marks the /b/ in 'bat' may have a quite different spectrographic profile than that in 'bottom' or 'both': crucially, *there is no initial segment in common*. Most likely, this has to do with how /b/ is *produced*: the vocal tract first produces a characteristic /b/ vibration, and then opens the mouth to release this. The vibration itself is *silent*; how it is released depends on the vowel that follows. (The mouth opens differently for different vowels.)

A syllable like 'ba' consists of three parts:

i. The resonance in the articulatory tract that corresponds to /b/. This is *silent* and forms no part of the acoustic signal that the speaker transmits.

ii. The transitional phase. A frequency slide from the /b/ frequency to the characteristic frequency of the vowel. This *is* part of the acoustic signal.

iii. The vowel frequency. This too is part of the acoustic signal that the speaker transmits.

Now, the important fact is that by itself, the transitional phase (ii above) is *not* heard as /b/. It is heard as a click or some other inarticulate sound. What must be true is that the phoneme-recognition system uses this as a diagnostic element, i.e., to figure out how the speaker made the sound in question. What is common to 'ba' and 'bo" is not any part of their acoustic profile, but *how the speaker releases the /b/.* Yet, the /b/ in 'bat' is *heard* as exactly the same as in 'bottom'. Not only this: it is heard as occurring before the vowel.

The continuous and smooth variation of the acoustic pattern along *acoustic* parameters – frequency and its transients – will result in sudden changes in what phoneme is heard by a human perceiver – a /b/ will turn into a /d/ or a /t/ suddenly as these parameters are varied.11 Again, such changes in perception correspond *not* to changes in the acoustic pattern that impinges on the ear, but on changes in the motor-sequence of the speaker who produces these sounds. Thus, consonants such as /b/ and /d/ have to be *extracted* from longer acoustic patterns, patterns in which they are combined with a vowel. Once they have been extracted, however, they sound the same as they do in other combinations; the /b/ in 'both' sounds just like that in 'bat'.

3. What happens, then, when you recognize the word that somebody has just uttered as 'bar' in your presence? First, the acoustic signal produced by the speaker has to excite your auditory receptors (cf *Reception* in section II above). Secondly, the speech-recognition system has to analyse the incoming signal and determine that it is one of the diverse spectrographic patterns that starts with /b/ (*Classification*). Since there is no initial part of these common to all of these syllabic units, the speech recognition system must first recognize a syllabic unit, and separate out the component phonemes – i.e., recognize the 'ba' sound that is produced by the articulatory tract when it emits an acoustic burst by means of the /b/-producing motion of the lips followed by the /a/-producing opening of the mouth.

Only after it has done recognized the 'ba'-sound, can the speech-recognition system separate out the /b/ and /a/ and ensure that you enjoy the characteristic auditory quale that you associate with /b/. The sound of /b/ in 'ball' has the same conscious quality as that in 'bile'. Moreover, it sounds as if it comes *before* the /a/, even though these are encoded together in the utterance. This sameness is *not* traceable to some common pattern in the acoustic signal that reaches the ear, i.e., by the detection of some common /b/ component in the spectrographic trace. It has to be attributed to the speech-recognition system recognizing two different patterns as both marked by an initial /b/. The characteristic /b/ quale and its place in a temporal sequence of phoneme-qualia are *consequent* upon this subpersonal act of recognition (*Access*).12

4. Consider what I earlier characterized as the traditional view of sense-perception. It holds that the *auditory* image corresponds to the activity of the auditory receptors. This view cannot account for the facts just adduced. Suppose that you have heard me utter /b/ as a part of the larger utterance 'ball'. Is it plausible to say that this event is *simply* the effect of your acoustic environment on your basilar membrane? No. In the first place, there is no part of the acoustic event that corresponds to /b/. Secondly, a dog or a cat would not hear the same thing, even if they were in the same environment. At least, they wouldn't hear it in just the way that we do: they would not experience the /b/ quale. They may achieve a coarse-grained phonetic recognition capacity by using other cues (such as voice-contour or inflection): thus, dogs can recognize their names up to the substitution of a few phonemes.13 Their experience would take in a number of things that you take in as a human. But it does not hear /b/ as a distinct feature of the spoken sounds that I emit independently of secondary cues. It would not identify this phoneme as sharply distinct in quality from /p/, and 'ball', consequently as a word that is different from 'pall'.

The communication-decoder that dogs possess is simply not attuned to the human phoneme code. The /b/ quale "encodes" a particular motor-sequence unit used to produce meaningful sounds. *Our* speech perception modules are able to identify otherwise disparate acoustic signals that contain this unit, and produce this particular quale to indicate that a speaker has performed this motor-sequence. Dogs can neither produce nor hear /b/ in a way that enables them to treat it as a distinct unit in spoken and heard speech.

5. The specialization of communication-perception makes it impossible to maintain that sensory content is merely the information-bearing effect of the environment on the organism's sensory system. When I utter a /b/, the information that I have done so is, of course, present in the acoustic signal that a human hearer receives. It is also present in the signal received by a bird in the same room. But the bird does not hear it the same way. By the same token, a territorial signal emitted by a bird is not readily recognized by other species: since I am not a trained ornithologist, it is not recognized by me.14 Conversely, such awareness as a bird might possess

of human phonemes is – presumably – learned and constructed. Thus, the *presence* of information in the acoustic signal does not tell the whole tale; one needs to say something about how that information is *extracted*.

I am urging that each social species has a communication recognition system. This sensory system actively processes environmental sounds, looking for the characteristics that would identify them as belonging to the various communicative signals used by the species. Once it has so classified the sounds that reach the ears, this sensory system subpersonally produces a consciously available *symbol* that indicates to other parts of the brain what it has found in the acoustic signal. How we qualitatively hear /b/ – what hearing /b/ is *like* for us – is such a symbol. This auditory quale is not a sound – it is an experience not a sound – nor does it indicate a sound. It stands for a phoneme. By producing such an experience, the phonetic comprehension module *tells* us – sometimes it is mistaken, so it does not merely *indicate* – that the sound just uttered belongs to the /b/ phoneme-category.

SCT urges a parallel model of visual experience, and of experiences provided by all the other senses (or perhaps just those with cortical involvement).

6. Consider another example, this time involving unconscious visual perception – again this is off to the side of our primary concern, conscious visual perception. It is well known that humans, including perhaps babies, can imitate the bodily actions of others – not perfectly, in the manner of a talented mime, but they are able to produce a functionally similar action. "Hold it like this," I may say to a child who is learning to write. The child may not have seen others holding a pencil, or seen that very grip before in any context. Yet it is able – slowly but surely – to imitate what I do. How?

One thing is clear. The child does *not* do this by manipulating her own hand in such a way as to create a duplicate of the optic array received from my hand. For one thing, its hand may not look anything like mine; for another, the angle of view of its own hand is different from that of mine; for yet another, a child "imitates" the actions of a dog or a baboon when these are more or less homologous with an action the child can perform. A child (or an adult, for that matter) will raise its right hand when told to mimic a dog raising it right front leg, even though the dog's leg neither looks like nor is positioned in the same way as the raised arm or hand. Similarly, a child will mimic biting or swallowing, even though its own actions are different in detail from those of the dog. This shows that the perception that subserves imitation is not based on similarity of receptoral activity.

Note that when an action performed by a dog is *outside* our basic repertoire, the ability to imitate is weakened. A child cannot imitate a dog walking or running – more often than not it will crawl on its knees when asked to do so, though this is obviously not what the dog is doing. Similarly, when a human observes a *robot* performing a homologous action – extending "finger" and "thumb," for example, or pointing – the imitation effect is much weaker. In the latter case, the *visual* similarity does not determine recognition: the robot's action may be strikingly similar to a human's, though clearly not "biologically" based. It is difficult to explain the phenomenon simply by reference to the information available in the optic array.15

Again, these results suggest that in order to perform certain instinctive actions – imitation, in this case – humans (and other animals) categorize environmental events into types. They are, moreover, able to translate the actions into motor commands – though this is incidental to present purposes except as evidence that sensory categorization corresponds to action types in the perceiving animal itself. Notice that *conscious* awareness plays a small, if any, role in this. Looking at somebody demonstrating different grips – a pen-grip, a hammer-grip, and so on – one may not experience a special quale attaching to each. This illustrates how sensory classification is separate from sensory awareness – the former can occur without the latter. Once this point is taken, it amplifies the role of sensory consciousness where it is available. Sensory consciousness provides us with a particular form of access to the results of sensory classification. Visually based imitation is subserved by a different form of access.16

The view that I am urging here is not *just* that sensory systems sort worldly objects into kinds, but that the kinds in question are often species-specific: humans tend to sort spoken sounds into phonemic categories and primate movements into categories constructed on the basis of basic behavioural gestures. How broadly does this point apply? Are sensory categories *always* species-specific in this way? Or do sensory classifications schemes more often correspond to physical categories such as temperature, reflectance (for color perception), physical shape, and so on?

Without wishing to exaggerate the scope of the specialization phenomenon, I would suggest that sensation in just about every modality and sub-modality is infected by the organism's point of view. Consider even a "primary quality" such as shape. It might be thought that there is no room here for species-variability. But even here there are perspectival influences. Christopher Peacocke points out, for instance, that a diamond is perceived differently from a square, even though they are the same shape. Why is this? Perhaps because they represent very different states of an object under gravitation. This suggests that what we are perceiving as "shape" is not an invariant property of objects, but a shape-related, perspectivally influenced object-state.17 Why do we perceive this rather than shape? Any answer to this question is bound to be speculative, but this kind of phenomenon suggests, contrary to some forms of "realism" with regard to perception (physicalism, for instance), that sensory classification takes the observer's point of view into account, and perhaps the actions that an observer might undertake from a particular point of view (cf. Akins 1996). Since different species have very different bodies and live in very different environments, they have different informational needs. This is why sensory classification schemes tend to vary across the animal kingdom. Sensory specialization is an outcropping, if you will, of the classificatory character of sensation.

Sensory specialization is readily comprehensible given *SCT*. If organisms of different species are sensitive to different environmental features, and if differences in what they can perceive does not originate in differences in their receptors, what they register has to be traced to post-receptoral processes, i.e., to what their sensory systems do with the receptoral array. *SCT* is a framework for understanding post-receptoral sensory function.

IV. Content

What exactly is involved in an act of classification? Fred Dretske (1981) has an illuminating and useful (and much quoted) insight:

> Until information has been lost, or discarded, an information-processing system has failed to treat *different* things as essentially the *same*. It has failed to classify or categorize, failed to generalize, failed to 'recognize' the input as being an instance (token) of a more general type. (*ibid.*, 141)

Dretske's thought amounts to this:

> To co-classify x and y as instances of the same type T, we must discard all information about x and y except what they have in common as instances of T.

What does 'discard' mean here? Presumably *not* that differentiating information concerning x and y is consigned to the wastebasket and is no longer accessible in any way. Just that the mental state that co-classifies them as Ts is itself arrived at by disregarding all such differentiating information. Thus:

> A classificatory state that subsumes individual x under type T retains only as much information about x as is relevant to its being an instance of T.

Dretske himself famously denies that sensory states are classificatory in this sense (1982, chapter 6). But this is a mistake. Look at the example of speech perception again. A /b/ uttered by three different individuals – a man, a woman, and a child, for instance – will be recognizably different in quality. Yet all are immediately recognizable for the phoneme they are. The process that decodes the speaker's *phonetic utterance* retains only the phoneme, and not the additional information about the age and sex of the speaker. *That* perceptual state is classificatory. Similarly, the 'b's in 'boat' and 'bat' are different acoustic patterns, but are heard as the same: the system is here extracting the common feature and suppressing, or "throwing away" the differences.

Of course, there may be other perceptual processes that use information emanating from the speaker while being insensitive to the phonetic content of the spoken message. For example, a child's voice may subconsciously evoke tender feelings even if you haven't taken in what it is saying: the process responsible for this is concerned with the child-like quality of the voice – to which, if you wish, you can pay attention independently of the phonemes uttered. This reflects the fact that within sensory systems, there are parallel "filtering" processes – computational processes, really – processes that retain only significant information. Some of these sort acoustic signals into kinds for the purposes of semantic decoding; others look for other characteristics. Our total experience of a spoken phoneme binds these together when we are paying attention, though it might disregard some elements in normal situations – one might hear what word was uttered without hearing who uttered it, and vice versa.

Let us pursue this line of thought a bit further. There is a certain amount of information available in the environment. Call this *ambient information*. Ambient information is a *trace* of what has happened – some condition of the environment that entails that certain events have occurred. The heat of the evening is ambient information: it indicates the heat of the day that has passed. Ambient radioactivity is information: it indicates earlier radioactive decay. Obviously, not all ambient information affects the visual receptors: some of it is in the form of sound, and though this is available to the perceiver, it does not enter the visual system; some ambient information is in forms not available to the perceiver at all – radio waves, gamma radiation, etc.

The cones and rods in the eyes are excited by incident light. There is information, therefore, in the state of these receptors, but this is only a small part of ambient information.

The information present in the visual receptors is what one might call *visually available information*. The visual receptors can be thought of as information filters: they throw away (or fail to accept) some ambient information. Thus, following Dretske, we may think of any given state of the two retinas – the activation levels of all light-sensitive cells in the two retinas – as indicating a "type". Two quite different states of affairs might result in the same retinal state, and hence yield the same visually available information. What this retinal state indicates about the environment is something these two states of affairs have in common. In this way, one might think of any retinal state as "encoding" a single type. It does this by "throwing away" all the information not relevant to something belonging to this type.

Christopher Peacocke (1992) defines a *spatial type* as follows:

> [O]ne basic form of representational content should be individuated by specifying which ways of filling out the space around the perceiver are consistent with the representational content's being correct. The idea is that the content involves a spatial *type*, the type being that under which fall precisely those ways of filling the space around the subject which are consistent with the correctness of the content. (105)

The state of my two retinas at any given moment constitutes a spatial type. It indicates a class of "ways of filling out the space around [a] perceiver": some ways of filling out the space around me are consistent with a given retinal state; some are not. A given retinal state *R* discards all information that differentiates ways of filling out the space around a perceiver that are consistent with *R*. It treats certain distinct ways of filling space as "essentially the same," to repeat Dretske's phrase. Thus, visually available information is a spatial type.

Other processes in the visual system result in type-content too. Within the retina itself, there is, for instance, a process that identifies sharp changes of illumination across the retinal field. Let *R1* be the retinal state that consists of a completely dark left hemifield and brightly illuminated right hemifield. Let *R2* be the state that consists of a completely dark *right* hemifield and brightly illuminated *left* hemifield. Both retinal states contain a vertical illumination "edge" at their centres, i.e., a centre-line that marks a sharp change of illumination. As retinal states *R1* and *R2* are as different from one another as two such states can be: the excitation level of each and every receptor is sharply different: the highly active ones in *R1* are quiet in *R2* and vice versa. Yet there is something common to the states, the vertical edge.

The point to which I should like to draw attention is that there are cells in the retina that respond to vertical edges of this kind. That is, there is a retinal cell *V* that "fires" for both *R1* and *R2*, but is silent when the retinal images do not incorporate a vertical edge at the centre. The on-state of *V* carries the information that there is a vertical edge in the centre of the retinal field. *V* carries no other information about the retinal field. In effect, it is the end-point of a process that discards all visually available information *except* this. It incorporates a *sensory* classification, and as such it contains less information than the retinal state it analyses – *V* throws away all the information that distinguishes *R1* from *R2*. From *V*'s firing, it is not possible to deduce whether the underlying visual information is more like *R1* or *R2* It is important to recognize that *V*'s on-state is in no way a picture of the retina: it has no pictorial function at all, but is simply an on-off, or Boolean-valued, indicator of some truth that has to be extracted from visual information.

Now, the visual system consists of layers of mutually interacting information-discarding – i.e., classificatory – units of this sort. The important point to realise is that right from the retinal receptors on, information is passed backwards and forwards from one classificatory unit to another. Assuming that the content of visual consciousness comes from the brain – and I assume that this is uncontroversial – it follows that one cannot identify visual consciousness with *visually available information*, i.e., with the information contained in the retinal type (much less with ambient information). *This* information is lost at the very start of the visual process. Rather, visual consciousness gives us news of classifications performed on visually available information, i.e., on the states of cells or modules that contain less information than the retinal type. Further, visual consciousness contains *nothing but* news of such classifications. There is, indeed, nothing else. As we saw earlier, receptoral cells are devices that filter ambient information. Every *other* cell that the receptors connect to are further classifiers, further filters, further throwers away of information. As visual processing proceeds, more and more information is thrown away and more and more classification is added. Assuming that visual consciousness emerges a great distance downstream of the receptors, it is the consequence of classification. What needs to be emphasized is this: in the terms that we have been treating of classification so far, i.e., in Dretske's terms, this is anything but a controversial claim from an empirical point of view. Dretske denied that sensation is classificatory: this is not because his notion of classification is different from mine, for I have been carefully adhering to his notion. His denial is based, rather, on a misconception of how perceptual processing works.

So far, I have been content to treat any state that results from throwing away information, and which thus treats different inputs as the same, as a classificatory state. But clearly the visual system doesn't just throw away information in a random fashion: it throws it away for a purpose. That is, it doesn't *just* treat distinct inputs as the same: it seeks significant similarities, and discards differences that are insignificant. I want now to suggest that classification is for action. This action may be something that the whole organism performs – a *bodily* action such as imitation, or an *epistemic* action such as induction or record-keeping or phoneme-recognition – or a further sub-personal data-processing step.

Consider some typical states of visual consciousness. A cup might look to be an arm's length away. This might prompt one to reach for it, rather than another cup that looks further away. Something looks green; this leads one to infer by induction that it has certain other properties: e.g., that it is unripe, if it is a tomato. That there is a vertical edge in a scene may aid in the segmentation of that scene into discrete objects. And so on. In each such case, a classificatory process guides further action. My claim is that the visual system's classification scheme evolved to enable and aid such instinctive epistemic operations.

It is this purposiveness in classification makes the visual system a device that produces representations rather than merely a system that throws away or loses information. It *extracts* a classification from visually available information, and preserves it so that the system it serves may choose appropriate action. To achieve this, the classification must be communicated by the data-processing system to an effector system, or to the "whole organism". The state or token by which this communicative act is effected is representational in that it betokens a situation beyond itself, a situation in which the effector is interested, either because it needs to respond to it behaviourally, or because it needs to keep a record of it. The classifications carried by such a data-transferring state or token do not belong to the categories of ambient or visually available information. *Action-relevant classification* is extracted from visually available information by the visual system and by other sensory systems. I claimed earlier that states of visual consciousness play the role of data-transferring token. It follows that states of visual consciousness represent action-relevant information.

I have been saying that the visual system extracts *information* from the retinal arrays. This does not exactly reflect the significance that classificatory states have for action. What the system does is, from the action-perspective, more like making a fallible *conjecture* about the scene on the basis of the visual array. The classificatory states of the visual system contain information alright, but the actions that are based on these states are better understood in terms of fallible conjectures, rather than in terms of the information carried.

Here is an example of what I mean when I distinguish information from conjecture. The visual array is two-dimensional. Its condition depends on the nature of light arriving at the retina and the state of the receptors themselves. But the action that the visual system subserves is ultimately about the three-dimensional world beyond the eyes and beyond the light. The shapes, motions, colors it is concerned with are object shapes, object motion, and object color, not (or not just) the transient and changing characteristics of illumination that is reflected from objects as the animal moves around the world. It has to infer distal states of affairs from visually available information, and such inferences are necessarily subject to error. Such error that is not merely due to imperfections of the system, but error due to necessary insufficiency of evidence: two-dimensional arrays created by light reflected from objects do not permit certain inferences about the external world.

Martin Godwyn (2006) introduces the term "para-information" to describe the background assumptions of such conjectural processing. An animal may be concerned with condition *C* in the external world. Retinal information *I* may reliably, but not infallibly, suggest *C*. Nevertheless, the visual system may prompt action appropriate to *C* when it has extracted this content from visually available information *I*. Implicitly, it is operating on the basis of the *false* (but generally reliable) conditional "If *I* then *C*". This conditional constitutes what Godwyn calls para-information: an assumption that lies behind a data-processing rule that the system incorporates for the purposes of reacting to environmental conditions on the basis of visually available information.18 I shall call the results of such conjectural processing *extracted content*.

Here is an example of extracted content: the visual system may be concerned with the condition in which a flying object is hurtling toward the animal. It detects this condition by the occurrence of a rapidly expanding coherent shape in the retinal image. When such a shape is detected, the system prompts the animal to duck or in some other way to evade the supposed projectile. Sometimes, it will do this even when there is no such object: a shadow or other such rapidly expanding retinal shape will provoke the same reaction. This behavioural pattern reflects the para-information that rapidly expanding shapes in the retinal array imply approaching objects in the environment. This para-information is used to extract content regarding objects on a collision course with the organism.

V. Similarity Orderings

Until now, I have been treating of sensory classification as if it yields discrete types. But this is a simplification. In actual fact, it more typically yields an *ordering*.19 As an example, consider color. Color is an ordering, not a collection of discrete "shades". A somewhat simplified view would have it that it is a position in a "quality-space" defined by three axes: black-white, red-green, and blue-yellow. The color vision system assigns a open-ended "value-range" along each of these three axes to surfaces and parts of surfaces in the visual field. A particular surface might, for instance, be coded as quite bright (more white than black), more red than green, and neutral between blue and yellow. (Such an object would be seen as pinkish.)

In this manner, color can be regarded as an inexact measurement rather than a simple true-false classification, a function from surfaces and their parts to an ordered sequence of comparative value-spreads with fuzzy boundaries. In general, sensory systems yield measurements of this sort. With some exceptions, they are not "categorical" in the manner of communication-perception systems, which in the case of humans at least, provide discrete values – classifying a particular consonant as an instance of /b/ rather than /p/, for instance, rather than as a sound that differs to a certain degree from others along certain definite dimensions.

Sensory ordering gives many philosophers the impression that the sensory systems simply react to physical variables. The thought is that such orderings reflect the excitation level of sensory receptors, that color is an ordering only because it reflects sensory receptor activity. In other words, it is claimed that sensation, or visual consciousness, corresponds to visually available information.[20] For example, many philosophers find it intuitive to hold that sensed *brightness* is a measure of the intensity of light at the retina, because the level of retinal activity reflects the intensity of light that falls on it. This is a mistake: cells in the retina do not simply respond to "quantum catch" or intensity. Because of lateral inhibition effects, it would be more accurate to say that they react to intensity *contrast*.

Intensity-contrast reflects the properties of environmental objects more than it does the condition of retinal receptors. For a white object placed in a scene with normal contrast will elicit a strong response from contrast-recording cells even in dim illumination, even though it reflects less energy to the eye than a darker object in bright light. Thus, a white object looks “bright” for the most part without regard to whether the light is bright or dim; this is because it reflects more light than colored objects in the same scene, and is brighter than them by contrast, whether or not it is bright in absolute terms. Thus, the message that the sensation of brightness carries is more indicative of the properties of external objects than is generally realized. This depends, however, on its being in scenes with “normal” contrast; when placed next to one another, white objects may look grey or even black in low light. Contrast informs us *fallibly* of features of the external world – and features chosen for their utility to the organism – as opposed to information concerning simply the condition of the receptoral array.21 Sensed brightness indicates extracted content, not visually available information.

Even very early in the visual process, then, the system is designed to extract content concerning external objects from receptoral activity rather than simply to record the latter. That this extracted content is “displayed” in terms of a sensory ordering rather than in categorical “on-off” terms should not blind us to the sorting activity of the sensory system using para-information.

VI. The Richness Argument

Similarity orderings have a characteristic that sometimes been taken to imply that visual content is “non-conceptual”. Richard Heck (2000) says, for example:

> My desk exhibits a whole host of shades of brown, *for which I have no names*. The speakers to the sides of my computer are not quite flat, but *I cannot begin to describe their shape* in anything like adequate terms . . . [M]y experience of these things represents them *far more precisely than . . . any characterization I could hope to formulate*. The problem is not lack of time, but *lack of descriptive resources*, that is, *lack of the appropriate concepts*. (489-90; emphasis added. Cf. Evans 1982, 229: “Do we really understand the proposal that we have as many colour concepts as there are shades of colour that we can sensibly discriminate?”)

This argument, which Heck calls the “Richness Argument”, is distinctly odd. Heck claims that the sensory color-ordering represents color “far more precisely” than a discrete system of “descriptive resources” would. He thinks that such precision is achievable only non-conceptually. Why? Is it because not every *point* in such an ordering can be named? If he were right, we should have to reject the possibility of a “conceptual” representational scheme for the real numbers. And while it is certain that we do not, and cannot, possess names or “appropriate concepts” for *every* real number, this can surely not be taken to imply that we mentally represent real numbers by a distinct *kind* of content, or (even more implausibly) by a distinct *kind* of faculty, than that by which we represent the rational numbers. Conversely, there are only a discrete number of phonemes that humans perceive: the obstacle of richness having been removed, are concepts admissible in *this* sensory domain?

Putting this consideration about the *validity* of Heck's argument aside, let us examine his premise. How many shades of brown *can* we see? A "shade" is a fully determinate color, one that does not admit of sub-division into more specific kinds. If we assumed that indivisible regions in color-ordering were like points on a continuum, the so-called Richness Argument would, at least, rest on an unassailable premise. But the notion of indivisibility has two quite different applications.

First, we have *physical* shades22:

> Physical objects *x* and *y* are of the same *physical* shade if they have exactly the same physical color-making properties (whatever these might be: reflectance, emittance, etc.)

Secondly, we have psychological shades:

> Physical objects *x* and *y* are of the same *psychological* shade to perceiver *P* if *P* cannot discriminate *x* and *y* with respect to color.

When Heck writes "My desk exhibits a whole host of shades of brown", he must be talking about *psychological* shades, since the fact that there are innumerable physical shades is of no relevance to how color is sensed or described. So one ought to ask what evidence there is that we lack the descriptive resources to capture the precise character of color-, or shape-, *experience*. It is clear that *physical* shades and shapes are as numerous as the real numbers. But this does not, of course, imply that there is even a large finite number of *psychological* shades and shapes, much less that these features are non-denumerably infinite. How many are there, then?

Switching to the example of shape for a moment, here is one consideration that might be thought to favor a very large number. Suppose that I am looking at a jagged city skyline – a more complex contour than that exhibited by Heck's computer speakers (see the quote above). Now, consider the following argument:

> (i) My experience of the shape of the skyline can be decomposed into a spatially extended series of experiences of indentations and protrusions (at least one such for every building that presents itself as a part of this skyline).

(ii) Each experienced indentation and protrusion is discriminable from one of sufficiently different elevation, size, and shape.

(iii) The permutations of all discriminable indentations and protrusions makes up a very large number.

(iv) Therefore, there is a very large number of discriminable experiences of shape that such skylines can present.

Arguments of this sort neglect the modularity of the perceptual process. Let it be true that I can visually discriminate indentations σ and σ' (at least when I attend to them). Let it also be true that shapes Σ and Σ' are the same but for a substitution of σ for σ' in one of them. Does it follow that Σ and Σ' are *visually* discriminable? No, because when I am attending to the larger shapes, I may not be capable of attending to the smaller one.

Suppose that I am presented with two photographs, taken a year apart, of Manhattan from the same distant view-point. I am challenged to spot the difference – a sort of Where-is-Waldo? game – but I cannot. Then my interlocutor points out a tiny sliver in the photograph taken later: it is a building that was erected during the interval between the two photographs. I can see and recognize the sliver. Now I can reliably tell the difference between the two photographs.

Does it follow that I can *visually* discriminate the two skyline contours? No. Of course, I can now discriminate them period. I can do this by looking for the sliver in question. But this involves (a) looking for the sliver, and (b) making an inference that cannot be performed by the visual system (namely that the skyline with the sliver is different from the one without). From the fact that vision is involved in step (a), it does not follow when I look at the photographs as wholes, I *visually* register the two skyline shapes as different. In fact, when I am not attending to any particular part of each contour, the two shapes may look pretty much the same – a jagged outline with peaks here, here, and here. What reason is there to think that they register differently? The argument, (i) to (iv) is therefore invalid: (iv) does not follow from (iii). From the fact that the experience of parts of shapes Σ and Σ' are different, it does not follow that the visual experience of Σ and Σ' are different.

Much the same sort of point can be made for color. When you look at a Munsell chart from a distance, trying to take it in as a whole, it seems to be composed of regions of different colors, much in the way that a rainbow appears banded. Look closely, however, at the supposed boundary between two adjacent color regions. The difference disappears; adjacent colors seem to blend smoothly into one another when you attend to them closely. Color differences you see from a distance are not visible up close.

Another argument runs like this.

(v) A visual image is fully determinate.

(vi) Therefore, no visual image can contain colors that are less determinate than shades.

The problem here is that, as we have just seen, our powers of discrimination may vary from context to context. Thus it is possible that a color that is fully determinate in one context is indeterminate in another. So the shades that one visual image exhibits might be a lot coarser than those that another exhibits. True, if visual images were like *physical* images, the above argument would lead us to conclude that there is a very large number of shades. But visual images are not nearly as determinate as physical images, and the shades they contain are not as finely grained as physical shades.

Now, it could be said that these are phenomena of attention, not of vision. We *see* the difference between the two contours, some might urge, but in certain contexts we are unable to attend to it.

Here is a thought-experiment that cuts against this suggestion. Suppose that you are presented with a split field with two very similar colors *C* and *C′* in the two halves. Suppose that you are successful in discriminating them. Now imagine two further tasks.

A. You are presented with a continuous expanse of color, smoothly varying from *C* at the left edge to *C′* at the right. You are asked to discriminate the left edge from the right.

B. You are asked to close your eyes and recall *C* rather than *C′*.

C and *C′* are discriminable in certain circumstances: that is what the split field presentation shows. But they might well not be discriminable in tasks A and B. In task A, the whole display might look uniform in color. In task B, it may not be determinate whether you are imaging *C* or *C’*.

Why would this be so? The explanation that some urge is that the color experiences in tasks A and B have the same fineness of grain as in the split field presentation – in other words, *C* and *C′* are discriminated just as well as in the split-field case – but that surrounding circumstances block or impede our noticing that they are different. The alternative that I would favor is that color experiences have different grain in these different contexts: the "shades" registered in tasks A and B are more coarsely individuated regions of color-space than in split-field task.

Try the following exercise. *Imagine* a color, say the saturated orange that is half-way between yellow and red. Now add a little black to it. How confident can you be that the color that you are now imagining is a shade as finely individuated as that which you see in the split field experiment above? If you were given several closely resembling Munsell chips, all satisfying the above description, how reliably would you pick the *same* one as most closely matching the shade that you are imagining? Pay as much attention as you can muster to your image of the dirty orange that I just asked you to envisage. Doesn't the discriminability of colors when you are actually looking at the chips outstrip their discriminability when you are merely imagining them? In visual imaging tasks, even introspection seems (to me, at least) to argue for the indeterminacy of the target shade, relative to the shades provided by direct vision. That is, I am not in the least confident that I can image a shade as definite as the ones that I can see, nor even that when I attempt to hold a visual image for a few seconds, it stays constant to the extent that it would have to in order to *look* constant to direct vision.

If imagined shades are relatively indeterminate in the way that I am suggesting, why should we think that in task B, you are recalling a definite shade relative to the shades presented in the split-field task? We might try the following as a control on B: after the display is removed, you are asked to pick from a selection including both *C* and *C′,* the Munsell chip that most closely matches *C* (the color displayed in the left hemi-field). Would the spread of chips you pick in successive repetitions of this task be as small as that which you pick when the chips are visible? In other words, are you less reliable in picking a shade when you are looking at it?

What about task A? Is *this* a case where *C* and *C′* register as distinct shades, but where we are blocked by the intermediate colors from noticing that they are distinct? It seems extremely unlikely. For one might ask: in this context, does the left edge look like *C′* (in the split field) or the right one like *C*? Neither option seems right. Both look the same shade, and hence the grain of psychological shades is different as between task A and the split-field task. Here at least, the correct answer must that the intermediate shades block *C* and *C′* from *registering* as distinct shades; the problem does not lie further downstream, i.e., with attention, or "noticing".

Fred Dretske once wrote "despite what [someone] *thought* he saw . . . the physical and physiological conditions were such that the object must have looked some [other] way to him"

(1969, 18). It is worth noting that this argument simply does not tell against the same physical shades being discriminable in one context and indiscriminable in another. The "physiology" of visual discrimination is sensitive to the context in which the targets are presented. It might well be, for example, that retinal contrast detectors will be active with the hemifield display, but inactive with the gradual variance display.

Gareth Evans once displayed a certain incredulity at the suggestion that we might have concepts for all the colors that we can sensibly discriminate. (See the comment parenthetically quoted just after the Heck block-quote above.) I find his point hard to understand: anybody who has visited a paint-store knows that there are more than enough "concepts" to go around, and there are, after all, a finite number of Munsell chips, and they all have numerical labels. Perhaps, Evans's hesitation had something to do with color-concepts we can carry around in our heads, rather than those that we can store in a book. Alright, we cannot carry all the Benjamin Morris paint-color names in our heads, but what does it take to have a color-concept that you can carry around in this way? Presumably, at the very least, some way of recalling it in visual imagination, or some way of applying it in the absence of an exemplar like a color chip. If, as I have suggested, visual imagination deals with coarser shades, then it would indeed be impossible to store in one's head as "as many colour concepts as there are shades of colour that we can sensibly discriminate." But this tells us nothing at all about the nature of the visual representation of color. It just tells us something about the coarseness of visual imagery and recall.

The concept of a psychological shade is notoriously problematic. According to the definition given above, two things will be characterized by the same shade only if they are indiscriminable. However, since indiscriminability is intransitive, shades so defined are not equivalence classes. Belonging to a shade does not therefore correspond to the logic of belonging to a kind (since kinds *are* equivalence classes). In addition to this, there is the problem that our response to the Richness Argument points to, namely that the shades are not determinately fine-grained. How are indeterminate shades to be represented? Not pictorially: an indeterminate feature cannot be pictorially coded.

Both problems can be addressed by abandoning the idea that sensory orderings provide us with direct knowledge of fully determinate classes. What they deliver, in the first instance, is precisely a multi-dimensional *ordering* of objects in the perceiver's environment. Such coding is not pictorial; it is propositional. It assigns each visual object to an open-ended range of values within a similarity ordering. Color vision, for example, does not present us with "fully determinate" shades (whatever these might be), but something more like: "This thing is roughly half-way between red and yellow, with quite a lot of black added". Such a description is normally indeterminate; it places the target within an open-ended region of color-space, rather than placing it precisely in a minimal region or pixel – though in optimal circumstances, it might accomplish the latter (provided that we understand that a pixel is a region, not a point). The seeming "richness" and "precision" of the visual representation of color masks this form of representation. Heck is right to suspect that the descriptive resources offered to us by a categorical (or discrete) scheme are inadequate to capture the richness of color perception; he goes astray when he suggests that no other resources are available to us.

VI. Perceptual Grasp

The argument given so far implies that visual states possess propositional content.23 They assign environmental objects to a kind, either to a discrete kind such as a phoneme or a named color (such as *red*) or a categorical shape (*square*, *round*: see note 1),24 or to a value-range within an ordering.25 If this is correct, the kind of content that vision provides is logically comparable to that provided by "thought". Visual content conforms, on this account, to Russell's conception of a proposition.

It is sometimes claimed that there is another kind of difference between perceptual content and thought content. Perceptual content is available to an organism even if that organism does not possess the concepts that constitute that content; however, thought content is not available to organisms that lack the contained concepts. I cannot *think* that a particular object is a cube if I do not understand what a cube is; however, something can look cubical to me, even if I don't understand what a cube is. To dramatize this point, some philosophers make the point that things can look cubical even to a dog or an infant, though these creatures lack the concept of a cube and can hence not *think* that the thing is cubical. Similarly, something might look taupe to me, though I lack the concept, *taupe*, and hence cannot *think* that the thing is taupe.

On the face of it, this argument is simply wrong-headed. The situation is symmetrical as between perception and thought: true, something can look taupe to me though I don't understand *taupe*, but equally I can believe or hypothesize that space is isotropic even though there is no such thing as the look of spatial isotropy. What is in question is surely not *inter*-faculty grasp –

understanding a visual concept or seeing an understood concept – but *intra*-faculty grasp. If something is to look *F* to me (in the phenomenal sense), then I should be able to grasp *F* perceptually. In both perception and discursive thought, the condition for entertaining a proposition containing a concept is that one should possess that concept in the manner appropriate to the faculty in question. And it might be thought that with *any* faculty, being able to entertain content is corequisite with grasping any concepts that partially constitute that content: when I see something as *F*, I am exercising my perceptual capacity to code something as *F*.

Now, though this challenge might sound initially plausible, it does demand further explication. What is it for one visually to grasp a feature? What exactly *is* perceptual grasp?26

I will be brief. Why is it not possible to entertain a thought without possessing the concepts that it contains? I would suggest that this has nothing to do with the nature of thinking as such. It is not, for example, that believing a proposition demands that one should be able to identify the application of the terms it contains. I might believe that Aristotle was a son of Nicomachus. This does not imply a capacity to identify either Aristotle or Nicomachus. It might not even imply a full grasp of the *son-of* relation. So one should not, I think, over-generalize from the perhaps justified observation that in order to believe that space is isotropic, one ought to understand what isotropy is. Not every thought requires that one should understand the constituent terms, be they referring terms or descriptive.

The condition on possession of constituent concepts is required, I would suggest, because without such a requirement, the identity of concepts in mentation cannot be assured. Suppose I believe both that Aristotle was a son of Nicomachus and that Aristotle studied with Plato. If Aristotle did not have a definite reference in both beliefs, I would not be able to use them to arrive at the further belief that a son of Nicomachus studied with Plato. For this inference can only be assured if the term is, so to speak, *mine* to move around from belief to belief. So *whatever* it might mean for me to possess the term 'Nicomachus' – however weak or non-descriptive or non-mental the conditions on possession might be for this particular term – I must possess it in this sense if beliefs containing it are to function as beliefs properly do. Now, one theory of proper names (the best such theory, in my view) requires that in order to possess the term, 'Nicomachus', I should stand in a certain causally uninterrupted passing-on relationship to its original use. So in order that I should perform the inference mentioned above – *really* perform it in the sense of detaching the consequent, as opposed to recognizing it as a valid inference – I should stand in this relationship to its original use. Possessing a proper name does not demand that I should be able to identify its bearer, or some such demanding condition of this sort, but merely that I should possess the term.

We need some such criterion for the *perceptual* possession of *perceptual* concepts. In order to arrive at an analogous test for *perceptual* grasp, let us note first of all that cognitive scientists implicitly distinguish between *responding* to a feature and sensing (or perceiving) it. A very pale person might burn in relatively low levels of ultra-violet radiation; a person with more melanin in her skin might not. This implies that the pale person's skin responds to a lower level of radiation; clearly, it does *not* imply that he *senses* ultra-violet radiation.27

With this kind of example in mind, researchers of comparative vision use *operant conditioning*, rather than merely differential response, as a test of sensory capacities. Suppose that some creature – a butterfly say – is rewarded with sugar-water when it tastes from a yellow dish, but not when it tastes from a blue dish (which contains unflavored water). Suppose that *as a consequence* it comes to try out yellow dishes in preference to blue dishes. Then scientists are inclined to say that it senses the difference between *yellow* and *blue*. They do not draw this conclusion when a creature simply responds differentially to these colours, but only in ways that cannot be modified in by operant conditioning.

The crucial point is that such a change in the creature's behaviour requires not only that it be affected by yellow things differently from how it is affected by blue things, but that it somehow preserve within itself a trace of earlier encounters with yellow things – a memory trace, if you like, though this probably wouldn't be a good use of the term 'memory'. This connection between perception and internal recording, the preservation of a trace that affects subsequent behavior, is, I am suggesting, a condition of perceptual grasp. Not all visually available information is graspable in this sense; clearly, features need to be extracted from visually available information for them to play a role in conditioning. Once these features have been extracted, they are available to be stored as part of a creature's record of the world.

Conditioning of this sort can be regarded an operational test of feature-representation in the sense that has been presented in the earlier part of this essay. The claim that I am making is that if an animal is in perceptual state *P* then it perceptually grasps all of the features that are represented in *P*. The possibility of conditioning is not, however, a *necessary* condition of representation. The grain in direct visual experience – i.e., when one is actually looking at something, as opposed to imaging it or remembering it – is much greater than in retained or anticipatory contexts. Here too, however, there are epistemic tasks that depend on this fineness of grain. For example, vision computes angle of view, spatial orientation, and object-identity on the basis of color-differences much finer than those that can be retained. Tests of representation could be based on these tasks, just as above I based them on conditioning. It should be noted that these tasks too depend on the system passing representational tokens from one sub-routine to another. A fuller discussion is not possible here, given constraints of space.

VII. Predicative Structure

The contemporary view that perceptual content is non-conceptual content is motivated by two threads of thought. The first is that perceptual content does not have logical structure, and in particular that it does not take subject-predicate form. If this is correct, it would not consist in the assignment of an object (as subject) to a class (as predicate). I have argued directly that it does take this form, and attempted to respond to two counter-considerations: the Richness Argument and the Conceptual Grasp argument.

In Gareth Evans (1982), we find another argumentative thread in favour of non-conceptual content in sensation. He argues that the sensory link between a perceiver and the thing that she is perceiving – the subject of perceptual content, not the predicate – consists in an "information-link" *not* in any descriptive concept such as color, shape, and so on. By contrast, he claims, *thought* about an object involves an identifying concept of that object. Thus, thought content is conceptual, while sensory or perceptual content is not.

While I will not pursue the point here, I believe this to be wrong. Evans is right, I think, to say that the relation between a perceiver and the thing is non-conceptually mediated. That is, he is right to say that the fact that I am perceiving *x* rather than *y* is not a matter of what properties I see these objects as having. However, I do not, for two reasons, think that this marks perceptual content out as different from thought content.

First, even though Evans's point about perceiving *objects* is correct, perceptual content might still have conceptual content in the predicate if not the subject place – that is, it still might have classificatory content. Evans's conclusion rests on the logical non-decomposability of perceptual content into object content and feature content, and it seems to me that this is a weak foundation for his view.

Secondly, I disagree with Evans's view about thought involving a identifying concept of its object. I believe that visual states enable human perceivers to *think* of visual objects demonstratively (Matthen 2005, chapter 13). Evans argues that demonstratively thinking about *x* demands that one should bring *x* under objective spatial concepts. I do not agree. But this is a topic for another occasion.

References

Akins, Kathleen (1996) "Of Sensory Systems and the 'Aboutness' of Mental States" *Journal of Philosophy* 93: 337-72.

Bergeron, Vincent and Matthen, Mohan (forthcoming) "Assembling the Emotions" in Luc Faucher and Christine Tappolet (ed.) *Canadian Journal of Philosophy* supplementary volume.

Brass, Marcel and Heyes, Cecilia (2005) "Imitation: Is Cognitive Neuroscience Solving the Correspondence Problem?" *Trends in Cognitive Science* 9: 489-495.

Campbell, John (2002) *Reference and Consciousness* Oxford: Clarendon Press.

Carnap, Rudolf (1928) *Der Logische Aufbau der Welt*, Berlin, Weltkreis. Translated by Rolf A. George (1967) in *The Logical Structure of the World & Pseudoproblems in Philosophy*: Berkeley and Los Angeles, University of California Press.

Coltheart, M. (2001). "Assumptions and Methods in Cognitive Neuropsychology" in B. Rapp (ed.), *Handbook of Cognitive Neuropsychology*) Psychology Press: 3-21.

Cussins, Adrian (1990) "Content, Conceptual Content, and Nonconceptual Content" in M. Boden (ed.) *The Philosophy of Artificial Intelligence* (Oxford: Oxford University Press): 380-400.

Dennett, Daniel C. (1991) *Consciousness Explained* New York, Touchstone Books, Simon and Schuster.

Dretske, Fred

(1969) *Seeing and Doing* Chicago: University of Chicago Press.

(1981) *Knowledge and the Flow of Information* Cambridge Mass: Bradford Books, MIT Press.

Evans, Gareth (1982) *The Varieties of Reference* Oxford: Clarendon Press.

Godwyn, Martin (2006) "A Defence of Extended Cognition." PhD Thesis, University of British Columbia.

Grice, H. P. (1957) "Meaning" *Philosophical Review*

Hamlyn, D. W. (1974) “Unconscious Inference and Judgement in Perception” in Nicolas (1974): 195-212.

Hatfield, Gary (1990) *The Natural and the Normative* Cambridge Mass: Bradford Books, MIT Press.

Hauser, Marc (1996) *The Evolution of Communication* Cambridge Mass: Bradford Books, MIT Press.

Heck, Richard G., Jr. (2000) “Nonconceptual Content and the ‘Space of Reasons’” *Philosophical Review* 109: 483-523.

Matthen, Mohan (2005) *Seeing, Doing, and Knowing: A Philosophical Theory of Sense-Perception* Oxford: Clarendon Press.

McDowell, John (1994) *Mind and World* Cambridge, MA: Harvard University Press.

Nicolas, J. M. (1974) *Images, Perception, and Knowledge* Dordrecht: Reidel, 1974.

Noë, Alva (2005) *Action in Perception* Cambridge MA: Bradford Books, MIT Press, 2004.

Peacocke, Christopher (1992) “Scenarios, Concepts, and Perception” in T. Crane (ed) *The Contents of Experience* Cambridge: Cambridge University Press: 105-30. Reprinted in Gunther (2003): 107-32.

Pylyshyn, Zenon (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science* Cambridge Mass: Bradford Books, MIT Press.

Raffman, Diana (2000) “Is Perceptual Indiscriminability Nontransitive?” *Philosophical Topics* 28:

Siegel, Susanna (2006) "The Contents of Perception", in E. N. Zalta (ed) *Stanford Encyclopedia of Philosophy* (Spring 2006 Edition): <http://plato.stanford.edu/archives/spr2006/entries/perception-contents/>

Sterelny, Kim (2003) *Thought in a Hostile World: The Evolution of Human Cognition* Oxford: Blackwell.

Zihl, J., von Cramon, D., and Mai, N. (1983) Selective Disturbance of Movement Vision after Bilateral Brain Damage. *Brain* 106: 313-40.

NOTES

---

[1]There is a simplification here that I will attempt to correct in section V: perception is not always "categorical." That is, it does not always present the world in terms of discrete properties such as *square* and *elliptical*, but in terms of similarity orderings. The example of shape is useful here because it *is* somewhat categorical: a roughly triangular squiggle or splash of paint presents itself as triangular, for instance. Marr (1982) takes advantage of this feature of shape-perception in his conception of 2 ½ D sketches, in which shape is reduced to a few canonical figures such as the cone, the cylinder, the ovoid, the cube, and so on.

[2] I treat the term 'subpersonal', which is in common use in philosophy of mind, as roughly equivalent to 'modular' in the sense intended by Jerry Fodor (1983). Indications that a data-process is modular include that it is (a) automatic, and thus not willed by the cognitive agent, and (b) "encapsulated", i.e., insulated from *willed* inputs from other data-processing operations that are occurring simultaneously. Personal operations are, by contrast, willed by the agent, and freely accessible to inputs and outputs from other operations. Fodor included other characteristics of modularity in his classic work, including – famously – domain-specificity. However, there is reason to think that these other tests are secondary (Coltheart 2001, Bergeron and Matthen, forthcoming). Note that this understanding of the subpersonal/personal distinction carries no overlay of the homunculus view of the personal, i.e., the Cartesian Theatre ridiculed by Dan Dennett (1991).

[3] It should be noted that we do not always have access to the results of sub-personal visual processing. The claim is that when we do, it is normally by means of visual consciousness.

[4]A version of this thesis is shared by philosophers such as Gareth Evans (1982) and Richard Heck (2000), who hold that judgements about the kinds to which objects belong – judgements of the kind 'My desk looks brown' and 'My desk is brown' – arise from the (personal) conceptualization of (subpersonally generated) sensory states. Evans says: "In moving from a perceptual experience to a judgement about the world (usually expressible in some verbal form), one will be exercising basic conceptual skills" (227). This is a position that traces back to Plato's *Philebus*, wherein it is argued that the mind imposes conceptual boundaries on the continuum of "the more and the less".

[5] It could be said that in this paragraph, I have confusedly counterposed Carnap's epistemological and normative thesis about the logical priority of total sensory experience with a scientific and naturalistic thesis about how the visual system processes information (cf. Hatfield 1990, Introduction). Not so, I think. Carnap assumes that total sensory experience is, as it were, *there* prior to our experience of color to provide a normative reason for our attributions of color to visual objects. The "naturalistic" facts show him to be wrong about this.

[6] D. W. Hamlyn (1974) argues that this dogma is historically a consequence of the "constancy hypothesis" – that there will always be a constant relation between how we visually perceive things and the pattern of retinal stimulation – together with H. J. Müller's docrine of "specific nervous energies" – that there is a functional independence of neurons. Together these imply that every visual sensation traces to a *local* stimulation in the homologous retinal location. No doubt, Hatfield (1990) would argue that Hamlyn is confusing the natural with the normative. But it is hard to see how philosophers of vision such as Berkeley could have arrived at this dogma concerning visual sensation without implicitly subscribing to these or equivalent doctrines.

[7] For a review of positions for and against the tradition, see Siegel 2006, especially section 2.

[8] Of course, the conditional statement that *if* there is sensory specialization, *then* there is sensory classification can be established a priori. Notice that the consequent of the just-mentioned conditional is a philosophical thesis, since it directly contradicts the no-content view. This shows how philosophical argumentation gets a foothold here. See also notes 5 and 6.

[9] A fuller account, tied more closely to a particular theory of speech perception, the Motor Theory, will be found in Matthen (2005), chapter 9, section I.

[10] Parrots can, of course, mimic human phonemes, though by a mechanism quite unlike ours. For them these phonemes belong to a much larger set of sounds that they can perceptually distinguish and produce. This might be regarded as controverting the perceptual specialization

alleged in the main text, but I believe that on a proper analysis it in fact illustrates it rather well – an excellent example of an exception that proves (i.e., tests) the rule. For the parrot is distinguishing these acoustic patterns for some purpose *other than* linguistic communication. I would therefore hypothesize that what the parrot hears has a different instinctive significance for it than it does for us, just as bird-calls have a different significance for us than for the birds. I will not pursue this topic here.

[11] This is the phenomenon known as "categorical perception" – see note 1.

[12] When I say that the act of phonetic recognition is subpersonal, I do *not* mean to imply that it is uninfluenced by the whole person's effortful epistemic condition. On the contrary, it is so influenced. For both the visual perception of the speaker's mouth and the semantic comprehension of her message will influence what we hear. In other words, there is feedback in phonetic processing as well as feed-forward. Thus, phonetic recognition does not simply "bubble up" from the subpersonal. There is, as John McDowell insists, a degree of "spontaneity" in this act.

[13] My father once acquired a dog that had unimaginatively been trained on the name 'Rover'. He (understandably) hated this name, and changed it to 'Babar', which, as he explained, was inflectionally similar. As far as I was able to observe, this never made a difference to the dog: he would respond similarly to either name being called.

[14] Even a trained ornithologist does not hear this call in the same way as a bird. For one thing, she would have to learn it in a quite different way. For another, it doesn't have the same instinctive significance for her as a terrritorial or sexual call. Perhaps one could say that the ornithologist's auditory experience of a bird's call is constructed from other qualitative units, whereas the bird's experience is simple and unconstructed.

[15] For documentation of these claims, see Brass and Heyes (2005) and the references therein.

[16] John Campbell (2002) claims that sensory consciousness provides us with knowledge of visual "reference". If he is right, this is not because of some special characteristic of consciousness as such. Consciousness provides us with epistemically special access to certain

features of the environment because it is the characteristic form of access for the human subpersonal systems that process these features. In cases where access is *un*conscious, knowledge is provided by whatever form of access we have in such cases. For example, the implicit form of knowledge we possess in the case of visual action-imitation is underwritten by whatever form of access normally underwrites it in human visuomotor system. It would less reliable, and less *appropriate* to the output, if it came in this instance out of conscious processing of the visual image.

[17] Adrian Cussins (1990) suggests that the perspectival character of sensation indicates an indexical element in its content. I do not think that this is correct. Something is indexical if its semantic value depends on the context of evaluation. Perspectival assessments are not always contextual in this way. Imagine that you are a diurnal creature, and that when your senses register a lot of light, your body enters into a wakeful state. This sensory determination is essentially perspectival: it characterizes the environment as one in which to be awake, rather than as merely one in which there is a lot of light. (A nocturnal creature would characterize the same environment as one in which to be asleep.) As such, the sensory determination is implicitly *relational*. But it is not indexical: it simply assigns the ambient lighting one of two values – <Sleep, me> or <Awake, me> – and in order to discover whether it is correct or incorrect, true or false, no context of evaluation needs to be provided. (Confounding the indexical and the relational is the error committed when 'here' is glossed as 'the place where the speaker is'. That, I take it, is Cussins' error.)

[18] It is not entirely clear that the system should in fact be taken as operating on false premises. It could be argued that it is operating not on the only approximately true conditional "If *R* then *C*" but on a more complicated premise such as "'If *R* then *C*' is generally reliable", or "The cost-benefit ratio of operating on the premise 'If *R* then *C*' is optimal". I do not find this plausible, since neither evolution nor the organism has measured the reliability of the conditional, or the cost-benefit ratio of using it over others.

[19] See Matthen 2005, Part II, for a more detailed treatment. That treatment is wrong in one small but important detail: there I treat sensory ordering as a detail that can be added to sensory classification thesis. Here, I treat it as a different, though related, perspective.

[20] Very few philosophers actually dissent from this thesis, which is a version of what Hamlyn (1974) identifies as "the constancy hypothesis" – see note 6. A recent adherent, though in a qualified form because he believes that this is only one aspect of visual consciousness, is Noë (2004). One indication of the strong appeal of the "constancy hypothesis" is that most philosophers claim that a white wall looks pinkish at sunset, even in situations where the perceiver herself may have no tendency at all to mistake the wall as pink, and may not even be aware of any pinkish cast.

[21] Here again, the "measurement" process is idiosyncratic. For as Hardin (1988) has argued, color similarity space bears little resemblance to the similarity space of physical quantities such as reflectance.

[22] I am very much indebted to Diana Raffman for intensive tutoring on the material that occupies the remainder of this section. See Raffman (2000) for more information about indiscriminability.

[23] Zenon Pylyshyn (1984) takes this view too.

[24] *Sounds* are the environmental objects assigned to phonemes, material objects those assigned to colors and shapes.

[25] Perceptual systems often assign objects to hierarachically ordered classes, so that a particular object may be seen simultaneously to belong to a color range as well as to a color category such as *red* or *orange*. The latter are broad ranges of values, typically more inclusive than the most determinate ranges delivered by even the dimmest visual image. These categories sit on top of the determinate, so to speak: they are simultaneously present in a hierarchical ordering.

[26] In Matthen 2005, the term “perceptual grasp” is infelicitously used in a weaker sense – to cover what I call “response” below.

[27] As will become plain in a moment, I do *not* take consciousness to be the mark of sensing. That is, my grounds for denying that the pale person senses ultra-violet are *not* grounded on his not being conscious of it.