HUMAN NATURE REVIEW
ISSN 1476-1084
http://human-nature.com/

Book Review

Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes edited by M. A. Peterson and G. Rhodes. 2003, Oxford University Press: Oxford, New York.

Reviewed by

David Dickins

This is a scholarly work that focuses on an important area in psychology and neuropsychology, which is of intrinsic interest, as I hope this review will to some degree make clear to a more general reader.

I should perhaps declare my slightly peripheral research interest in it, which may explain deficiencies of insight that a reviewer working directly in this field would not have exhibited. I was looking for background information to help me predict the way the brain handles visual stimuli of different categories so that I might be able to exploit this in a study of learning connections between stimuli, and connections between these connections.

The editors begin a useful introductory overview by pointing out that the distinction between "analytic" and "holistic" processes in visual perception is the contemporary version of the classic dispute between Structuralists and the Gestalt psychologists. It does not follow from this that the terms are used with much consistency, either between authors, or even within authors when referring to different domains of perception. Perhaps they are only as good as the techniques used to distinguish between them operationally, which in this book are legion.

The faces, objects, and scenes of the title distinguish extraordinary feats of perceptual recognition that may depend upon different mixes of holistic and analytical processes, and involve distinguishable regions of the brain. With faces the remarkable thing is that subtle conjunctions of a limited array of features – eyes, nose, mouth etc. – which are so alike across all people, particularly within a racial group with which one is most familiar, can be used for highly reliable identification of specific individuals. With objects our ability to generalise a

myriad angles and aspects of view into reliable categories into which an object can be classified is remarkable. The perception of scenes seems to work in yet another way….

"Holistic" is not the same as "configurational". Surely the classic case of configurational analysis is that of face perception. Faces are made of pretty standard components (which themselves of course – eyes, mouths, noses – may be the outcome of lower order configurational processes), yet we recognise those we know from an impressive personal databank of subtle individual variations in the precise arrangement of these features. [I doubt we would recognise the isolated nose of our beloved or even her eye, even though its beam had been twisted with ours during many courtship days.] Is this information never encoded in memory, or only unconsciously so, or are we only ever at least consciously aware of them in their natural context, because the very record of them is coded in this way?

Most of the chapters in this book review a large amount of detailed material and go thoroughly into what these studies imply. To do justice to each would make this review too long. I was impressed by all, and delighted by most of the chapters in this book, and if I pick out some and exclude others, the authors concerned should forgive me for rather randomly foraging in a bonanza situation before total satiation set in.

Thus Bartlett, Searcy, and Abdi (in Chapter 1) make a complex case for there being two distinct routes to facial recognition: configural/holistic versus componential or part-based processing. First they review a face recognition study. If asked to identify a famous face either from the upper half or the lower, subjects can do this, but they do it more slowly if the face they see is a carefully aligned synthetic composite of two different famous faces, one upper and the other lower, compared with the control in which the face of one famous person is split into laterally misaligned upper and lower halves. If however the faces are all upside down, the difference between composite and non-composite controls disappears.

They go on to consider a plenitude of other behavioural studies involving, besides recognition (as above), assessment of degree of grotesqueness, learning names of faces, reporting whether or not a face has been seen before, and matching faces. Even though it is implied that different tasks may call forth different mechanisms of perception, details of what subjects are required to do are irritatingly not made clear, forcing one, if one wants to follow the argument entirely through, to consult the studies cited. Thus the Garner selective attention task (dating back to 1974 and used in a cited PhD thesis) is a subtle use of differential reaction times (to make what response?), which plays off internal and external features of faces against each other as cues, and studies the interactive effects of inversion on this. The conclusion is that some subjects at least performed "truly holistic" processing (defined by these particular authors as extending across large regions of the face, not just internal or immediately

adjacent features).

Much is now made of the "conjunction effect". A series of faces are shown, and then, in testing, these faces are mixed on successive trials with new faces, and the task is to say whether they have been seen before or not. On the test however a third type of face is also presented in which features from several previously seen faces are seamlessly recombined. One can then measure hit rates for old faces (ones correctly identified as such) with false alarm (incorrect positive) rates for both new and composite faces. False alarms are typically higher for composite faces than for new, indicating that to some extent memory for faces is based on storage of features. Some authors (Reinitz et al) have argued that all memory storage of faces is done in these terms, since they say relational information is hard to encode accurately and soon fades. This requires new relations to be worked out on retrieval, some likely to be erroneous, of which some may resemble the artificial combinations in the experiments. Bartlett et al. however have further refined these empirical findings by comparing participants of different ages, with the apparent discovery that some but not all of the older participants showed evidence of posterior brain damage. At this point the importance of exactly what the subject is required is addressed: participants had to say whether they consciously recollected a face, 'knew' it to have been present but without such recollection, or judged it as new. Also the effects of the number of presentations (1 or 8) of old faces and features prior to the tests is factored in, together with whether the faces were the right way up or upside down.

Following all this the results of a computer simulation are presented (which I shall not aspire to condense here), and finally a study of concurrent brain function, using Single-Photon Emission Computed Tomography (SPECT), which, like fMRI (functional Magnetic Resonance Imaging), measures blood-flow as an index of localised neural activity. Statistical parametric mapping (SPM) is used, based on difference scores, in this case between O, (old faces), and C, (conjunction), together with N, (new), (in both directions, O > C&N and C&N > O), and between C>N and N>C. These contrasts generated a wealth of data summarised clearly in a series of "glass brains" more familiar to this reviewer. A sketchy interpretation of these would run as follows.

O > C&N. Old faces produce better configural matches (drawing on right-occipital regions), and the processing of non-novel stimuli is in keeping with greater activity in right-medial-temporal regions, and evoke conscious recollection in left parietal, and working-memory processes in right-anterior brain regions. But facial repetition and familiarity (in other studies) may be linked to decreases as well as increases in activation, which fits with the C&N > O patterns in certain right occipital and parahippocampal regions, so that the authors felt that differences in both directions made sense.

C > N produced areas in ventral occipital cortex on both L and R,

including the fusiform gyrus, but extending further forward on the left to include the fusiform face area of Kanwisher et al. suggesting enhanced analysis of features. Also C>N showed activity in the left middle temporal gyrus which has been linked with word recognition and object classification, suggesting more verbal/semantic coding for conjunctions of familiar features. But no evidence was found from C>N of stronger frontal and parietal activations associated with conscious recollection.

In another chapter (5), Schyns and Gosselin present some remarkable pairs of hybrid pictures in which what you see depends upon whether you switch in one filter excluding all but high spatial frequencies or another filter excluding all but low spatial frequencies. You can do this by looking at these pictures "normally" (high frequency) or squinting at them or holding them at a distance (low frequency), when you can see, for example in one pair, either a nonexpressive woman above and an angry man below, or vice versa, a startling demonstration. These are monochrome images, and the frequencies alluded to are those of the variations of luminance or grey levels. Now while there is evidence that the perceptual system may often work on a coarse to fine sequence, i.e. it begins with low frequencies and takes a little longer to analyse a picture in terms of higher frequencies, the authors have elegantly shown that subjects can strategically adopt an appropriate frequency level (from several naturally available) depending upon which level is most appropriate for the categorisation task involved. Furthermore they have developed an elegant technique of exploring in "three dimensions" (the usual two of the picture's area together with that of spatial frequency) which different aspects and combinations of features at different frequency levels are most salient in either identifying a face, or assessing the gender of the face, or assessing the emotional expression on that face. The technique involves generating random clusters of "bubbles", parts of the total scene at a particular level of analysis for each bubble, and repetitively asking, for many such random compositions, what the subject sees. The answer is usually determinate and the subject is unaware of alternatives, and the corpus of results permit analysis of what the key features are for various types of categorisation (of the same picture). Some combinatorial analyses, which these authors term "holistic", such as combinations of features, or spatial relations between them, emerge in these performances. This kind of technique is a sophisticated and potentially powerful way of studying the selective use of diagnostic information in perception, which may be applied to any of the types of stimuli discussed in this book, and doubtless other aspects of perception also. One of the first things to do would be to repeat the above experiments, adding in effects of inverting the stimuli.

Tarr's impressive chapter (7) has similarly enormous range. He produces a lot of evidence by rotational studies of specially designed shapes and objects, that recognition of objects may depend, contra Marr, on viewpoint-dependent rather

than object-dependent processes. He adduces evidence that this may apply not only to the so-called subordinate level of classification, referring to specific object recognition, but also to the basic level, enabling one to categorise a diversity of objects as belonging to the same basic type. He then introduces a third factor, that of experience, in which he marshals striking neuropsychological evidence that face-recognition is just a special case of expertise (that virtually all of us have acquired). One route into this is by his team's invention of "Greebles", totally artificial objects (though to me at least rather humanoid) that have both "family" and "gender" characteristics and can be recognised as unique individuals. [They remind me somewhat of the "smurfs" or "Schlumpfen" associated in the UK and Germany in the past with a certain brand of petrol, which my children and their cousins used to collect]. With practice people improve markedly at being able to recognise shared and individual characteristics built into Greebles.

Tarr and his colleagues (Gauthier et al) have also carried out parallel studies using fMRI. The first described explored level of classification. They point out that evidence in the literature e.g. (Kanwisher)[1] for special face-identification regions of the brain, e.g. particularly the fusiform gyrus of the visual cortex, is based upon contrasts between individual faces (e.g. "Bob") at the subordinate level of classification, while those for the non-face, "objects" (e.g. "bird"), were at the basic level. Tarr and colleagues compared activations when the same picture had to be classified either at the basic level (e.g., "bird"), or at the subordinate level (e.g. "pelican"). In one of two maps printed in monochrome only in the chapter, but reiterated elsewhere in the book as a colour insert, areas of activation are shown derived from this contrast that correspond to putative face-selective regions. Controls were applied to rule out differences due to purely semantic differences between the two levels. The other fMRI figure shows subtractions for 3 inexperienced subjects for faces minus objects and Greebles minus objects: at the middle fusiform gyrus only the former comparison shows differential activity. Alongside this are comparable results for 3 subjects trained as Greeble experts, for whom both subtractions show activation in this location. I wondered whether the anthropomorphic nature of Greebles predisposed them to analysis by the same pre-existing system in place for faces, but Tarr has other creations (he illustrates "Fribbles", "YUFOs" and "Pumpkins"), apparently less open to this criticism, waiting in the wings.

Continuing this argument, in Chapter 2 Tanaka and Farah however found no evidence for holistic representation analogous to that required in face-recognition in experts on biological cells, cars, or Rottweiler dogs. From their view, despite Tarr's Greebles etc., faces may still be distinct as the best example of holistic representation in the real world.

Tanaka and Farah go on to review, and largely dismiss, the idea that in childhood holistic representation only develops between 6 and 8 years, relying on

featural encoding before then. But there may be a critical age for the development of configurational discrimination: LeGrand et al found that at age 9 children who had had congenital cataracts which were removed shortly after birth were selectively impaired in their ability to detect changes in configuration of facial features, though not in the detection of features themselves or in other kinds of pattern perception.

Another well-known and dramatic line of evidence for specific face-processing brain regions is the existence of the neuropsychological deficit of prosopagnosia. Tarr wondered whether the low face-recognition accuracy of sufferers from this, compared with their relatively normal object-recognition, including the claimed good performance on subordinate-level classification of some subjects on discriminating chairs or spectacles, might not be misleading. Instead of relying on simple measures of accuracy, when Tarr and his colleagues controlled for time available to respond, and equated the sensitivity for faces and non-face objects in 2 prosopagnosic subjects he found they were far more sensitive to the manipulation of level of categorization compared with controls.

When it comes to object perception the major question is how can a variety of views of the same object be integrated so that it can be identified from any of them? As mentioned above David Marr [2] theorised that a single orientation-independent representation was formed, but others have adduced evidence for multiple viewpoint-dependent representations. Although this dispute seems a separate issue, in relation to the near consensus that object perception is analytic whilst face perception is (fairly uniquely) holistic, there is some tendency to regard viewpoint-independent representations as analytic in contrast with viewpoint-dependence that may entail holistic, template-like representations.

[Introspection is deemed of little use compared with objective experiments, but one wonders whether a 3D (or even 4D) representation of the brain is something expert brain surgeons, radiologists, and perhaps some neuropsychologists possess, and into they can integrate multiple triads of transverse, sagittal, and coronal slices (as in glass brains), or if they merely acquire the capacity to behave as if they had.]

In practice it is likely that fusion of multiple views of an object owes a lot to the fact that clusters of these will occur as we encounter objects in relative motion in the natural world. Bülthoff & Bülthoff (Chapter 6) describe their work with Sinha on point-light walkers, recordings of a man in a gait analysis laboratory walking 3 paces with 12 lights attached to his joints, and only the motions of the lights presented on a computer monitor. Speed of first recognising that a human figure was involved depended strongly upon familiarity of point of view, as this was systematically shifted from various initial orientations, some more natural than others. This applied with or without stereoscopic presentation, and was not simply a function of the amounts of information available in different

views of the object.

With the perception of scenes, the Bülthoffs used virtual environments through which participants could "move" within a restricted area. Passive or active movement were more helpful than being shown a series of still shots, and participants were able to generalise to views they had been restricted from seeing during their exploration, and to distinguish these from views of environments distorted in some way from the training environment, even including simple mirror inversions.

Is the temporal continuity between different views of an object as we turn it round or a scene as we walk around it important in bringing them all to bear on subsequent recognition? This would seem highly likely.

In another extraordinary experiment 3D morphs were made so that as a detailed computerised model of one actual individual's head was rotated it (literally) 'turned into' that taken from another real individual. Only the experience of sequentially seeing them in this way, and not simply the presence of such morphs, prevented subjects subsequently discriminating between single views of the real faces in a same/different matching task.

These authors end with enriching references to the world of art and science fiction, and wonder how we might contend with virtual worlds in which objects and scenes became incongruously different as we walked around them, and offer a link to their Tubingen website with video sequences you can sample on http://www.kyb.tuebingen.mpg.de/links/metamorphosis.html.

In object recognition, one feat is to recognise objects of a given class irrespective of viewpoint. This we can do, though more familiar "canonical" views enable faster recognition than those less often encountered. Tasks given in experiments, may be either explicit – have you seen this before in this experiment ?- are these two pictures the same or different ? - can you choose which of these matches the one you've just seen ? – or implicit tasks – what sort of animal would you say this is ? - could this object actually exist in 3D space?

Hummel (Chapter 8) presents a functional take on the perceptual system, contrasting the costs and benefits of holistic and analytic representations of shape. He clarifies the various levels of shape perception, from primitives such as pixels, or 2D lines and vertices, through the kind of reference frame used, and then the specification of the relations which specify how an object's features are arranged within this frame. This brings him to the final level of binding, which is concerned with how different features or parts that occur together are cross-related in perception, and connected also with their spatial address. Hummel makes an important distinction between static and dynamic binding. Static binding would be inherent in the visual system with a separate unit for every possible conjunction of properties, e.g. a unit just for binding shape and colour which might only recognise (say) blue squares. This could be done quickly, but

would entail a huge cost in terms of computational space upfront. Dynamic binding would involve units for each property, say one for blue and one for square, being integrated as they both occur over as short a time as possible, before going on to the next conjunction of further properties. Here less computational space would be required, but more time. It is possible to use electrophysiological measures to log what must be brief moments of synchronous firing corresponding to dynamic binding, and to correlate this with how quickly a monkey for example might identify a face, or study the temporal parameters of masking. In terms of costs and benefits, both static binding, which Hummel identifies with holistic representations, and dynamic (corresponding to analytic) are required, and will complement each other, if a system is to work effectively. Again a few closely argued pages contain a wealth of ideas and data.

Kimchi (Chapter 9) uses the Navon task to explore the global precedence hypothesis that when e.g. one large letter (say H) is composed on an array of smaller letters (say S) – or vice versa – under most conditions the larger letter is identified first. But by examining the timing of priming effects she finds that holistic effects, defined here as dependence upon interrelations between components such as simple lines (smaller for Kimchi than the components of facial features in face-recognition), may occur very early, meshing with physiological findings that even primary visual cortex may show the effects of grouping and segregation. The classic findings of Hubel and Wiesel, and others, that the activity of simple cells in visual cortex was due to the convergence of inputs from bar and edge detectors and the like in the visual pathway are made more complex by the discovery in the cortex of horizontal interactions and back projections from higher to lower centres of the system. Again, the perceptual system seems to be highly interactive.

In her own chapter (10) Peterson goes through her work which has shown that figure/ground processes do not necessarily precede retrieval of object memories, but may be influenced by them. This fits with physiological work on the ventral cortical stream where cells have been found which respond to feature conjunctions and are sensitive to the presence and relative position of more than one simple feature. Simultaneous firing of cells of this kind would solve the binding problem mentioned above.

Behrmann (Chapter 11) describes some neuropsychological patients described as "integrative agnosic" (or in one place, agnostic – perhaps an overbusy spellcheck editor) who seem to differentially lack holistic configural abilities. They are unable to assign contours to the appropriate object when there are several overlapping objects, though they can perceive the contours per se. They are impaired on face recognition, and as Peterson says, they deserve to be tested in many of the situations described in this book, since they may support the view that there is a fundamental distinction between holistic and analytic

processing in perception generally.

As for the background against which faces and objects are commonly perceived, [to cursorily summarise the last 3 chapters of the book], the landscape seems only to be sketchily coded in most perception of it. One way to show this (Behrmann, Chapter 11) is to switch, with an interpolated gap, from one backdrop to another. Subjects often miss many aspects of the scene unless they happen to be concentrating on the particular region that is altered for the test, although there are signs from measures of eye movements (Simons, Mitroff, and Franconeri, Chapter 12) for example suggesting some implicit memory for landscape features. There is some suggestion, (Henderson and Hollingworth, Chapter 13), that we reinvent the landscape background from one occasion to another, generating familiarity rather than basing it on well-documented memories of its actual components.

The presentation of this book is to a high standard, and the graphs and other illustrations mostly fine, though some of the black-and-white photos and drawings are only just adequate (supplemented by some useful colour inserts). Altogether this is an extraordinarily rich and informative volume, despite its apparent small size (it actually has 393 pages including references and index), and the high price is perhaps a fair enough reflection of the value it is likely to serve to collate and encourage research in this progressive field, even though it may deter individuals from buying it. It may not have solved my problem, but has certainly made me more aware of myriad further problems, and this enhanced awareness will, I hope, inform my own work.

Anyone researching on human perception, or taking an interest in this from nearby areas, should persuade their library to add this to its collection, or somehow get to read this admirable collection of articles.

David Dickins, Honorary Senior Fellow, School of Psychology, University of Liverpool, UK. Email: D.W.Dickins@liverpool.ac.uk.

1. Kanwisher, N., J. McDermott, and M.M. Chun, The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 1997. **17**: p. 4302-4311.
2. Marr, D., *Vision*. 1982, New York: W. H. Freeman.