Rock Art Research **2013** - *Volume 30, Number 1, pp. 75-90. L. DOBREZ and P. DOBREZ*

KEYWORDS: ***Animal representation*** *–* ***Recognition*** *–* ***Profile*** *–* ***Salience*** *–* ***Neurophysiology***

ROCK ART ANIMALS IN PROFILE: VISUAL RECOGNITION AND THE PRINCIPLES OF CANONICAL FORM

Livio Dobrez and Patricia Dobrez

Abstract. The article examines factors involved in rapid and easy visual identification of animals in life and art. It gives an account of what we term canonical form in connection with 'basic level' recognition, profile depiction and the concept of salience. In the course of this it introduces discussion of part/whole relations and saccadic eye motion. Overall it offers a critical assessment of the literature on the subject of recognition and suggests likely neurophysiological correlates for the perception of real and depicted profile animals.

How do we visually identify or recognise objects in the world? Visual recognition is a matter of primary evolutionary importance and an understanding of the principles by which it operates helps explain the way in which we depict things. In this article we examine the limited but key example of animals in profile as depicted in rock art. Our interest is not in precise identification per se, that is, we do not ask questions such as 'is this a picture of a kangaroo?' or 'what species of kangaroo is it?' Rather we are concerned with *principles* of recognition, which we discuss in terms of the idea of canonical form. The term 'canonical form' is used by Hochberg (1972, 1978), though in a way that differs from ours. Similar terms are used by other psychologists ('canonical view': Rosch et al. 1976; 'typical contour': Deręgowski 1995) and by art theorists ('norm image': Arnheim 1969, 1974). In a computational context Marr and Nishihara (1978) refer to a 'canonical description'. There are other examples, including a mathematical usage of 'canonical form' we take to be broadly in line with ours. By our definition the expression designates those visual factors involved in easy, rapid and reliable identification of an object, in this case an animal.

Reality and depiction

For present purposes we require no distinction between a real and a depicted animal (Fig. 1a, b). This not because we think perception of real and of depicted objects entails an identical operation. Experience suggests it does not, and experienced difference has to have correlates at the neuronal level. At the same time we do not accept Gibson's celebrated but, as we understand it, minority thesis that we see what is real and what is depicted in radically different ways (Gibson 1966, 1971). Perceptual psychologists and neuroscientists implicitly run reality and depiction together when they use pictures instead of actual objects in experiments. Relevantly, there is a general postulate that perception, imagination and memory have much in common. Of course to view a picture of a horse and to imagine a horse are not experienced as identical and art theorists are aware of it (Wollheim 1998: 224–225). At the same time imagery-based perceptual psychology which takes account of neurophysiology (Kosslyn 1994) suggests to us some connection between imagining and viewing depictions. If this is so, it is reasonable to hypothesise that similar (if not identical) neural operations are involved in seeing, imagining, remembering — and viewing depictions. This is the approach taken in this article, and for a (mostly) psychology-oriented discussion relating to the matter we refer the reader to Deręgowski and commentary (1989); for an art history discussion we refer the reader to the disagreement between Gombrich and Wollheim on the perception of picture-surface and picture-content (Gombrich 1960; Wollheim 1973, 1998). To return to our argument about canonical form: that there is an evolutionary necessity for neuronal ensembles to fire rapidly and reliably is more evident in the situation in which we encounter a real animal than one in which we view a depicted animal; with the real animal we need to make a quick decision: either to attack something that is good to eat before it gets away or to climb the nearest tree to escape something that will eat us. However, the fundamental perceptual principles by which we recognise a real object as easily and rapidly as possible may also be

Figure 1a, b. *Road sign and red-necked wallaby (*Macropus rufogriseus*).*

expected to operate in situations of pictured reality.

Seeing-as

It is important to realise that when we perceive, say, a kangaroo, we do not first perceive — and subsequently recognise. We do not see X and then identify it as a kangaroo. Of course in bad light or without our spectacles we may be unsure about what it is we perceive. Likewise, if we have never seen a kangaroo, identification will rely on what we already know. Thus a seventeenth-century Dutch sailor will recognise, e.g. a 'large rat'. In all cases, however, there is no raw data, an X, subsequently turned into a 'kangaroo'. What we see is, in the first instance, a kangaroo. Even if, given confusing circumstances, we mistake a kangaroo for a dog, we see, in the first instance, a dog — then adjust: it was, after all, a kangaroo. What we always do is to see and recognise in the one perceptual act. The philosopher Wittgenstein (1968) referred to this as 'seeing-as': all seeing is seeing-as.

Basic level recognition

There is a connection between issues relating to recognition and questions of categorisation. If we may postulate universal levels of categorisation, at which level do we normally recognise objects, i.e. see them *as* something? Rosch et al. (1976), followed by Tversky and Hemenway (1984), postulated three such levels — superordinate, basic and subordinate — putting the case that we recognise at basic level. This argument has obvious relevance for the (usually polemical) discussion of motif identification in rock art — and we are not the first to refer to it in the context of rock art: Halverson (1992) has done so. At any rate for Rosch et al. the superordinate is the most abstract of the three levels. At this level we cannot recognise a shape as something specific, and when we try to imagine superordinately, i.e. to visualise a 'vehicle' (non-biological) or an 'animal' (biological), in other words 'vehicle' or 'animal' in general, we fail to form a concrete image. We would add here that depiction is unlikely to work at this level: even the most schematic rendering of a vehicle or an animal will presumably tend in a particular direction — a vehicle will take the form of a car or a boat or an aeroplane etc.; an animal, the form of a kangaroo or eland or bison etc. This shift takes us from the superordinate to basic level. Perception or imagining at this level is (relatively) concrete. For this reason it is operative when we first recognise an object. Likewise this explains why children first learn to discriminate objects at basic level. With respect to the subordinate, the Rosch et al. case is that it is less perceptually important, for reasons explained below. In summary, then, the thesis may be stated as follows: initial recognition is of a car (basic level) and not a vehicle (superordinate) or a Nissan (subordinate), a kangaroo (basic level) and not an animal (superordinate) or an eastern grey (subordinate).

While we regard the Rosch and Mervis (1975) analysis of the internal structure of categories (according to Wittgenstein's 'family resemblance' model rather than a 'criterial features' one) as problematical, we are inclined to agree with the general Rosch et al. (1976) superordinate/basic/subordinate thesis and consequently to posit that what we term canonical form operates at basic level. When we recognise, with ease and speed, that the object is a kangaroo, we are recognising at that level. By contrast, while it may, in some circumstances, be possible to make a superordinate identification (e.g. when we see something in the distance or in poor light and take it to be an animal without further distinguishing characteristics) we nonetheless accept that in such a case recognition is limited and provisional. What would we be relying on for the canonical form of an 'animal'? Movement? Clearly this is a situation of indeterminacy. Rosch et al. argue for an equal but different limitation on recognition at subordinate level. If they applied their thesis to diverse kangaroo species (eastern and western grey, euro, red etc.) they would say that these have more attributes in common across species than within a single species, that is, an eastern grey has more in common with a euro than with other eastern greys. In short, subordinate level categorisation generates informational diminishing returns: kangaroos are more easily recognised as kangaroos than as particular kinds

Figure 2. *Mt Borradaile, Arnhem Land, Australia.*

of kangaroos.

There is a necessary complication in this, and one which Rosch et al. address: the evident fact that one person's basic level may be another's subordinate and vice versa. The greater someone's expertise in an area, the more 'subordinate' their recognition in the eyes of non-experts. If we know about cars we begin by seeing a Nissan and not a car — and probably a Nissan Patrol or even (though this might be countered by a law of diminishing returns) a 1991 Nissan Patrol. If we know about kangaroos we will, despite what has been said above, identify an eastern grey or (Fig. 1b) a red-necked wallaby from the start. Knowledge of animals that may be expected of hunter-gatherers will very likely elicit recognition of greater detail than that indicated by 'eastern grey' or 'red-necked wallaby'. What, then, is that basic level at which we register, in our terminology, canonical form? Clearly it may be located at different points on a sliding scale, depending on expertise. But is such an elastic concept able to claim universal validity? Rosch et al. put a fair case for just such universality. The *content* of a category is individually, intra-culturally and cross-culturally variable, but the *principle* of basic level categorisation, as defined, remains constant: all perceptual recognition has *a* basic level at which it operates. If there is a difficulty in the thesis, it is that it seems a little less comfortable with biological categories, and that this is reflected in substantial experimental reliance on artificial objects and consequently sketchy treatment of animals. But the biological is what experiment might most usefully focus on, since it is there that evolutionary factors have been at work.

We have given space to the basic level proposition not only because it has some currency but because it bears on our own concern with *principles* of recognition. Moreover the sliding scale of basic and subordinate categorisation obtains in the perception of pictures: as in life, everything in art depends on the expertise of the viewer. The Rosch subjects, mostly Berkeley undergraduates in psychology and more familiar with cars than with fauna, might identify an Arnhem Land rock art depiction as being of a kangaroo (Fig. 2). Someone more familiar with Arnhem Land fauna might identify by sliding the categorial scale such that what for non-experts would be subordinate takes centre stage as basic. A certain kind of expertise, that of rock art researchers, might prompt cautious use of the family name 'macropod', indicating a very broad category. Still greater caution prompts identification of a 'zoomorph'. We recall those debates about uncritical identification of motifs, as well as proposed solutions to the problem, not least the (surely intentionally teasing) clicking exclamation prefix — for a witty overview of which we refer to Clegg (1991). The observation we make here in connection with such debates is that to the (varying) extent that 'macropod' and 'zoomorph' imply a situation of perceptual indeterminacy they are superordinate and therefore do not imply proper

Figure 3. *Picasso centenary postage stamp featuring* Guernica.

Figure 6a, b. *Llamas, Peru.*

Figure 4. *George Stubbs, British (1724–1806).* Lord Grosvenor's Arabian stallion with a groom, *c. 1765. Oil on canvas, 99.3 × 83.5 cm. AP 1981.03.*

Figure 5. *Linton panel B, Iziko Museums of Cape Town Social History Collections (Archaeology).*

recognition. What happens in the situation of viewing rock art is either that we suspend judgement because expertise or lack of it suggests an image is indeterminate — or that we perceive a particular object (a kangaroo), i.e. identify at basic level, but have non-perceptual reasons for not committing ourselves to a particular name.

Iconicity and realism

Principles of recognition, seeing something 'as' at basic level should determine iconicity in the Peircean sense, i.e. pictorial representation that we recognise reasonably concretely as a depicted kangaroo or horse or eland or llama etc. This would amount to recognising in the representation a particular canonical form, that of a kangaroo or horse etc. We must be firm in stating that it does *not* amount to a conflation of either the term canonical form or iconicity with 'realism'. A picture of a horse may be perfectly recognisable as such, i.e. iconic and exhibiting the canonical form of a horse, without reference to the specific representational rules of post-Renaissance European art or the camera. To grasp the force of this statement we need only compare Picasso's horse (Fig. 3) from *Guernica* (1937), which breaks all the rules of post-Renaissance art and the camera, with a Realist (not 'realistic', i.e. approximating reality, but in a Realist style) picture of a horse by the celebrated English painter of horses, George Stubbs. We suggest as an example (Fig. 4) *Lord Grosvenor's Arabian stallion with a groom* (c. 1765). *Both* depicted horses look like horses. But do both equally exhibit the canonical form of a horse? We say they do not — but *not* because one looks more like a real horse (i.e. is 'realistic'). Rather, because one facilitates ease and speed of recognition where the other does not.

Profile depiction

The fact is that a given object may be recognised in different postures and from diverse angles. So a variety of views may access canonical form: there is not simply one canonical aspect for the object. But are all views equally canonical? Is an animal recognisable with equal ease and speed from any viewpoint? Clearly not. Whether in life or art, some views will be dominant over others. So for example we recognise standing profile eland in one of the famous Linton panels in the South African Museum, Cape Town, more readily

Figure 7. *Toro Muerto, Peru.*

than eland facing away from us (Fig. 5). We recognise a real llama best in profile and not by its rump (Fig. 6a, b). Unsurprisingly, Andean camelids are depicted in profile (Fig. 7), as are North American bison, Franco-Cantabrian bovines and horses, Scandinavian moose and reindeer, Indian bovines and cervids, Australian macropods and the range of African fauna. With the exception of plan-view for a minority of animals and rare frontal or rear-view depictions, as in southern Africa, animals in rock art worldwide are overwhelmingly shown in profile, sometimes modified by so-called twisted perspective. We will not labour this point, as it is a matter of common knowledge among rock art scholars — simply noting that, while culture certainly influences the way we represent animals (see, for example, preference for 3/4 views in post-Renaissance European art), long-term historical preference has been for dominant views, i.e. ready perceptual recognition. Such recognition relates to canonical form and may be taken as perceptually universal. We have looked for experiments on the perception of animals in life and art, but psychology experiments do not necessarily focus on animals or on depiction (though they frequently substitute pictures for real objects). Results may be skewed by culture (e.g. off-axis preferences recorded by Blanz et al. 1999); when animal images are used, orientation may not figure (Thorpe et al. 1996, 2001; Van Rullen and Thorpe 2001); where orientation *does* figure, it may be of artificial objects (Edelman and Bülthoff 1992). Interestingly, Rosch et al. (1976) note, by way of an aside, that when they showed their students animal images they were surprised at the preference for profile views, concluding that these were information-rich in the same way as basic level categories. We would like to add to this that dominant views (for most animals, profile) best express canonical form. But why precisely is this so?

Salience and typical contour

One way of answering the question may be to focus on the phenomenon of salience. Salient features or even a single salient feature facilitate easy recognition, and for most animals these features are best exhibited in the laterally-viewed cervico-dorsal. In some cases, notably in Franco-Cantabrian examples from Cougnac, Pech-Merle, Niaux, Ekain etc., recognition is satisfied with a partial outline of the animal or even simply an outline of the cervico-dorsal. This suggests a principle of economy: that even within the general rule that depiction is bound to schematise its original, depicted canonical form requires minimum visual detail. We may certainly opt to depict with more detail than required, as in the case of Stubbs' horse and a great deal of recent European (Realist) art, but the historical trend, as illustrated by rock art, is towards economy. We recall the 'fixed action pattern' (FAP) identified in Tinbergen's experiments with gull chicks: the chick pecks at a red spot on the parent's beak which will cause the parent to regurgitate food, and when the parent and its beak are replaced by a stick painted so as to highlight the colour red, the chick continues to peck (Tinbergen 1960). In short, salience operates *pars pro toto*: the cervico-dorsal suffices for identification of a Palaeolithic mammoth in the same part-for-whole way that, for the gull chick, colour on a piece of wood suffices to stand in for its parent.

Salience is put forward by Rosch et al. (1976) as a determining factor in the formation of tendentious categories by appeal to a 'prototype' — the implication being that real categories are of the 'family resemblance' sort specifically outlined in Rosch and Mervis (1975). We have difficulty understanding how an animal *without* a trunk would fit into the category of 'elephant' through a family resemblance with elephants, i.e. by having features in common with some but not all elephants. The trunk would seem essential here, and it suggests the 'criterial features' model of categorisation. 'Criterial features' as the determinant of a category is readily compatible with the concept of salience, while 'family resemblance' may not allow for it, except, as pointed out above, when we opt for tendentious 'prototypes'. Unfortunately salience is infrequently discussed in the context of rock art. Cheyne et al. (2009) compared images of European Palaeolithic horses and bison with images of presumed similar real horses and bison and concluded that Palaeolithic artists had

Figure 8a, b, c. *QANTAS and Australian Made logos, and Australian coat of arms.*

exaggerated salient features, viz. the cervico-dorsal profile. Deręgowski investigated what amounts to salience under the rubric of 'typicality' in an article in *Rock Art Research* (1995) which develops ideas put forward in *Distortion in art* (1984). Believing (we think mistakenly) that the argument for 'typical views' of objects contradicts the Gibson account of perception, he nonetheless fruitfully extends the notion of a typical view to outline art, especially rock art. Outline depiction reproduces 'typical contours' in the real object which would seem to be perceptually universal — and these are discussed using Attneave's (1954) Information Theory scheme. Deręgowski argues that typical outlines (chiefly in profile depiction for animals) work because they exploit the fact that information is concentrated at points of change, i.e. where a line changes direction. We accept this, with the proviso given by Halverson in his comment on Deręgowski's article (Halverson 1995: 15) that the concept of salience, implicit in Deręgowski's argument, needs to be made explicit. As we see it, an object in life and art, let us say a kangaroo, may be visually described, but only in part, by reference to variations in its outline shape. The point is that some of these variations will be common to animals other than a kangaroo and so will not be typical of a kangaroo. In order to pinpoint that typicality which we term canonical form we require variations that are information-rich in a *particular* way, viz. salient features for the object in question.

Thus, modifying Deręgowski and choosing our own examples, we would say that those information-rich points of salience needed for easy and fast recognition of a kangaroo or a wallaby might especially come into play when the animal is hopping (Fig. 1a, b). Diagnostic features might include the line of arched back, extended, with upward tilt, to the prominent tail, and the 'V' shape of forward-propelling back legs. The result is more or less what is chosen for road signs and QANTAS logos and Australian Made —though not the Australian coat of arms (Fig. 8a, b, c). In fact macropods may exhibit salient features, i.e. canonical form, when in a number of characteristic poses, e.g. moving on all fours, standing upright etc. Still, the hopping posture probably has compelling primacy. It may be that not all animals are equally liable to variations on the canonical theme: the bison's salience is probably mostly in the hump, the giraffe's in the neck, the American camelid's in the right angle of neck and back, the cervid's in the horns etc. These examples need be no more than conjectural for the purpose of this article which concerns itself, as explained, with perceptual principles rather than specific cases. But what we are saying should be compatible with Deręgowski: the salient points for the hopping macropod would be points at which the lines of cervico-dorsal/tail and back legs markedly change direction. What is between these salient points would, in Information Theory terms, be characterised as 'redundant', i.e. predictable, whereas sharp shifts at areas of salience are unpredictable.

Now Deręgowski, in line with his general interest in perspective, nicely uses the notion of typical contour (in art, typical outline) to explain Breuil's anachronistic and Eurocentric invention of 'twisted' perspective. So-called *perspective tordue* simply illustrates the artist's wish to show a combination of typical views in the one image. We need only add here that 'twisted' perspective reveals a clear preference for the depiction of salience, indeed for a complex of canonical forms. It goes without saying that 'twisted' perspective is no such thing and exists solely in the minds of European observers unaware of the cultural relativity of post-Renaissance Realism or the camera image (for a lucid account of which the reader is referred to Hagen 1986). The limitation of Deręgowski is that, while offering a perfect deconstruction of Breuil's *perspective tordue*, he continues to manifest the bias towards modern European forms of perspectival depiction, constantly conflating the (undeniable) everyday reality of perspectival *perception* with the 'realism' of certain kinds of art and the camera, both of which unjustifiably and culture-specifically assume a single subjective station point as the criterion for correct depictive projection. On the positive side, we think that Deręgowski's reading of 'twisted' perspective as a way of combining typical views is very much to the point. It might be added, however, that there is a limit to the views the artist may usefully combine in the one image — and this explains why Stubbs' horse is taken as embodying the canonical more than Picasso's. Stubbs' horse is in simple profile, whereas Picasso's multiplies viewpoints, some of them uncanonical (profile belly, neck, chest and three legs — with the fourth hoof seen from below! — plan-view rump, frontal as well as

3/4 head). This playfully complicates canonical form, precisely countering the requirement of economy or part-for-whole for rapid recognition — but not quite to the point where we no longer recognise the horse. Interestingly, in this image (Fig. 9), taken from the present authors' Canberra market, the artist, who has chosen to depict all but one animal in profile, chose the deer so as to combine dorsal salience with the salience of horns. Where in some societies, past or present, this might entail a choice for Breuil's supposed 'twist', this contemporary Western artist has his cake and eats it by following post-Renaissance rules and turning the deer's head in the viewer's direction!

Figure 9. *Fyshwick Market, Australian Capital Territory.*

Additive vs holistic

Let us conclude this critique of Deręgowski by stressing that we see diagnostic salience, as the canonical form of a given object, in the context of the object as a whole. We recognise the bison by its hump and the giraffe by its neck, but only in relation to the rest of the bison or giraffe. This implies a particular reading of the *pars pro toto* aspect of the phenomenon of salience, viz. that in an immediate way the giraffe's neck activates perception of the entire giraffe, the bison's hump activates perception of the entire bison. The whole, in short, is able to be contained in the part, such that, for perceptual purposes, neck *equals* giraffe, hump *equals* bison. It must be said that there are very different ways of understanding the relation of part and whole. We take the view that a holistic model best explains the part-for-whole aspect of salience. But, to begin at the beginning, let us note that part and whole are themselves not straightforward notions. In theory, any unit may constitute a part *or* a whole, such that parts and wholes would be entirely arbitrary. In this view, any part may become a whole if detached from its larger unit, and any whole become a part if integrated into a still more inclusive unit. Thus a brick on its own is a whole; in relation to a house, a part; and the house, whole in relation to is bricks, becomes a part in relation to a town. In practice, however, that is, in the real world, parts and wholes may well be non-arbitrary. Psychologists as diverse as Arnheim (1974) and Biederman (1987) agree on this. The difference and the continuing debate, now extended to neurophysiology, is in the way we understand the relation of part to whole. The holistic or whole-first argument is found in highly original form in Gibson (1979). In classical form it is associated with the Gestalt school, more recently represented by Arnheim (1974: 42–95). Here the postulate is that of the whole as greater than the sum of its parts — not as an extra over and above its parts but as a given arrangement of those parts such that change to the arrangement alters every part, and vice versa. It is of course easier to see how this applies to a picture than a Mix Master, but then our concern here is with pictures. The parts-first argument postulates a principle of addition: add X, Y and Z and the result will be, say, a kangaroo or giraffe. This approach seems inescapable in computational modelling (Marr and Nishihara 1978), though, if we understand aright, there have been attempts to find a middle ground between atomistic bottom-up and holistic top-down visual processing models (Lowe 1987). An atomistic approach is critical for Hochberg (1972: 59–60, 1987: 290), who appeals to Penrose's 'impossible pictures' which purportedly must first be seen in their parts and only subsequently attempted (impossibly) to be perceived as wholes. The trouble with the Penrose examples is that, like Escher's impossible perspectives, they are not encountered in the world, but have been cunningly and artificially constructed precisely to provide a stumbling-block for the whole-first argument. Biederman's thesis goes further than Hochberg's: it has it not only that we see parts *first*, but that we *only* see parts. This is the 'recognition by components' (RBC) idea, which seeks to explain Gestalt phenomena by bottom-up perceptual organisation. RBC postulates that perceptual recognition decomposes its objects into primitive volumetric-geometric units ('geons'), thirty-six of which, in all possible combinations, suffice to characterise the visual world. Strictly speaking we register geons rather than objects — and we can, contra Arnheim, change the arrangement of geons without altering the geons. It may be objected that this formulation shifts the whole from the object (kangaroo or giraffe) to its constituent parts (geons), i.e. treats geons as fundamental wholes. But it cannot be the case that a geon is a whole, for the simple reason that it is not itself constituted by parts. Geons are irreducibly parts, which may be arranged in an external additive relation to each other.

To what extent this is a correct prediction of the way our visual system actually processes information remains to be seen, and we return to the point below, observing at this stage that, just as Biederman and others (e.g. Baylis and Driver 1995) have sought to challenge the whole-first position, so a number of experimenters have wanted either to modify or challenge the parts-first thesis. Palmer (1977) offers support for the position we adopt in connection with Deręgowski's presentation of 'typical contour', viz. our argument that Deręgowski fails to consider the importance of contextualising typical (read 'salient') features as elements of a larger

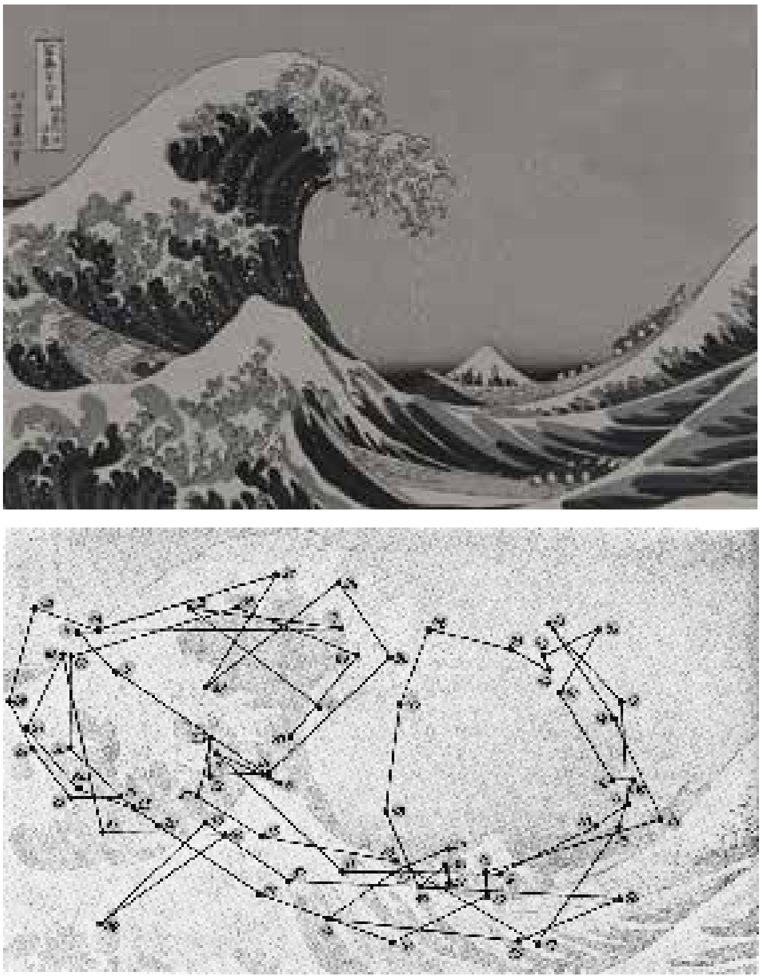

Figure 10a, b. *Hokusai,* Stormy Sea off Kanagawa*; Buswell 1935, Plate XXXII.*

whole. In fact Palmer sees his project as reconciling atomistic and holistic approaches. Navon (1977) is of the same mind, putting the question simply in his title: 'Forest before trees'. Working, for example, with large letters formed by small ones (e.g. an H made up of small Os) he concludes that global recognition is prior to local recognition: we see the H *before*, not after the Os. The point is taken up, contra Biederman and Marr/Nishihara, by Backer Cave and Kosslyn (1993), who argue that perception initially accesses the overall shape of objects, such that recognition occurs before any parsing, i.e. without need for analysis of object parts. Analysis may well be activated *subsequently*, as a 'fallback strategy' (1993: 230, 245), when the initial identification proves, for whatever reason, unsatisfactory. We find this a convincing account of visual requirement for *both* synthetic and analytic processes, with priority given to whole-first perception.

Figure 11. *Twyfelfontein, Namibia.*

This is the logic we would wish to apply to Hochberg's work on saccades. Saccades are eye movements at a highly variable rate by which we shift our fixation, i.e. bring an object from indistinct peripheral vision to distinct foveal vision. A pioneering example is Buswell's (1935) analysis (taken up by Gibson 1950 and Hochberg 1972) of a particular subject's successive fixations on the print of Hokusai's boats riding that celebrated wave, with Mt Fuji in the off-centre background (Fig. 10a, b). Fixations, numbered 1–70, begin just below the threatening tip of the great wave at the left of the picture, subsequently working their way up, down and around it, then up the slope on the right — and doing so, though not identically, several times, before resting at a point slightly below the start. In his first book (1950) Gibson sought to reconcile the phenomenon of discrete fixations with the seamless continuity of perception; in his last (1979: 212–213), he insisted that fixation is merely a laboratory phenomenon: in reality fixation and movement cannot be separated. Gibson may well be right, but we would like, if only as a heuristic exercise, to apply the Buswell model. In the sense in which we may speak of successive fixations, the process has to allow for perception of salient features, these being, in general if not inevitably, the ones we notice and direct attention to first. Accordingly, in the case of this Namibian giraffe (Fig. 11) we might begin, as with many animals, with the cervico-dorsal line, here greatly emphasised, and in a particular way, by the length of the neck. Thus initial fixation would be on a point in the neck, with subsequent fixations following the neck up to the small head or, alternatively, down to the body, then to the long non-salient legs which, however, might gain prominence if perceived as counter-balancing the long neck — after which the eye might retrace its steps all the way to the head. We do not suggest anything like an inevitable order in this, merely likely tendencies on the basis of salience. Now the temptation of an empiricist like Hochberg, contra Gestalt organicism and in a different way contra Gibson's idea of perceptual invariants, is to envisage recognition of the giraffe as an analytical, part-by-part process of adding up fixations so as to arrive at some stage at the conclusion that it is a 'giraffe'. The empiricist's problem is to postulate how discrete fixations on the giraffe are integrated so as to form the giraffe — and the answer is usually an appeal to memory. We accept that short-term memory has a role in integrating saccades, but not that mere addition of saccades generates a perceived giraffe. Rather we see no reason against supposing that the initial fixation, in combination with peripheral vision and activating the *pars pro toto* effect of salient giraffe neck, itself generates the entire giraffe, i.e. registers prior to subsequent analytic fixations — which of course may well *clarify* the image or, even, if it happened to be very indistinct to begin with (say by reason of poor light), be a *prerequisite* for recognition. This last would constitute the 'fallback' strategy mentioned by Backer Cave and Kosslyn above. Actually much of what Hochberg says may be read as supporting a more holistic understanding of saccadic movements, given that, following Buswell, he stresses their purposive quality (as guided by e.g. the design cues of Hokusai's picture), and also allows a role for peripheral vision in anticipating what we see. Despite which, however, he understands the operation as additive. Of course we regard his reference to anticipation as especially relevant, since in our terms seeing 'in advance' might be regarded as normal, given *pars pro toto* salience — and this is a matter to which we return below.

It must be added that there are experiments which would seem to contradict the saccadic, parts-first model of recognition. Results of work on ultra-rapid recognition by Thorpe et al. (1996, 1999, 2001), Delorme et al. (2000), Van Rullen and Thorpe (2001) and Van Rullen and Koch (2003) showed object recognition so rapid as to preclude that first fixation. In fact ultra-rapid recognition registered 'something there' rather than a specific object, this being read as an animal, since the instructions were to 'spot the animal'. While we do not believe these experiments undermined the basic-level thesis, as, in some cases, they were intended to do, they certainly undermine the Hochberg idea of a saccadic building up to a point of recognition. Rather they suggest something like the following: in that first ultra-brief exposure we note a peripheral-vision 'something' which is indeed superordinate but does not amount to recognition — and we do this with speed adequate to the evolutionary imperative of survival; that triggers fixation and, in the initial interplay of foveal and peripheral vision, we see a whole rather than parts of the given object (to a degree even if the object is partly occluded). This naturally assumes an object that is not too large. It is only subsequent to this operation, which should engage salience, e.g. the giraffe's neck, that saccadic eye movement comes into the picture as required. In more usual perceptual situations, which will be less demanding than those in ultra-rapid recognition experiments, that initial peripheral 'something' may already be specified by foveal vision, as we suggest in the previous paragraph. That would ensure that we did indeed see-as, following Wittgenstein's formulation.

We end this discussion of salience as the key to understanding the nature of recognition, i.e. of accessing canonical form, not least that of animals, by reaffirming our position — outlined in connection with comments on Deręgowski and, after that, on parts-first and whole-first approaches to recognition. Discussion of salience raises the most fundamental issues, viz. the nature of the relation of part and whole, and this relation comes into play in every instance of seeing an object *as* a giraffe, a kangaroo etc. Our view is that to make sense of the phenomenon of visual salience, we require some version of the holistic argument, the idea of a perceived overall organisation or pattern triggered by salient features such that, contra Biederman, change to the whole changes a part and vice versa. Here we suppose ourselves to be in general line with the Gibson idea of invariants, and also to be in agreement with Arnheim, though with reservations, since we do not espouse Arnheim's Gestalt position as such. In which connection we think that, while Gestalt's philosophical orientation, derived from Phenomenology, remains valid, a number of its fundamental premises were questionable from the start, while others are now superannuated in the light of developments in neurophysiology — to which we turn presently. Before that, however, we offer a clarifying addendum to what has so far been said.

Ambiguities and complications

It must be understood that different objects may have very similar, even, in highly unusual circumstances, identical canonical forms. This image (Fig. 12) was identified for us as a white rhino (F. Prins, pers. comm.

Figure 12. *Injasuthi, Kwazulu Natal, South Africa.*

Figure 13. *Rabbit-duck (Die Fliegende Blätter, 1892).*

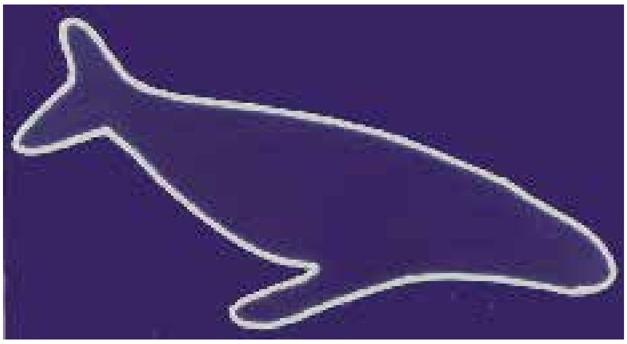

Figure 14. *Arizona whale-kangaroo (courtesy J. F. Kilhlstrom and E. Vezey).*

2009). At the time we canvassed the option that it might represent a warthog and indeed it is given as a bush pig in a pamphlet written by Bert Woodhouse (n.d.: 6), the famous South African amateur rock art recorder. Ambiguity centres on whether the figure features horns or tusks, and is probably unresolvable by visual means alone. Something of the sort could be said about Australian images which may be read as deer or humans on all fours and wearing a headdress resembling antlers (Welch 2012); or the so-called Woy Woy 'rabbits' (Stanbury and Clegg 1990), also probable humans with headgear (though other options have been canvassed). In these cases ambiguity is presumably quickly resolved, 'deer' and 'rabbits' being unlikely candidates. Still, perception registers ambiguity, i.e. varied objects with similar canonical forms. In the case of real-life ambiguity, say the stick/snake in your path, you may have to wait for it to move — or not. Evolution has taken advantage of canonical form similarities in varieties of animal camouflage. A well-known illustration of *identical* form for entirely different objects is the puzzle of the rabbit-duck (Fig. 13). Since form remains exactly the same for rabbit's head and duck's, we are forced to make something comparable to a figure/ground switch — here prompted by choice of right orientation (rabbit) or left orientation (duck). These comments apply to a more recent and equally amusingly concocted image (Fig. 14): the 'Arizona whale-kangaroo' or AWK (Kihlstrom 2006). What about symbolic representations? Do they exhibit canonical form? As symbols, they clearly do not: there is no such form for concepts of 'love' or 'justice'. There may be iconic representations of these (Cupid or a blindfolded female with scales). But these are only recognised *culture-specifically* as 'love' or 'justice'. As iconic, they simply represent a boy with a bow or a blindfolded female. It is possible for geometric forms to exhibit canonical form. Thus a cube is viewed canonically at an angle, since, taken frontally, it has the canonical form of a square. Finally, it is also possible for invented, non-iconic signs, mathematical or alphabetical, to exhibit canonical form, it being possible to obscure such form with 'bad' handwriting.

The visual system

We come to a consideration of issues relating to a possible neural basis for the argument presented above. If we understand correctly, it is now generally believed that the brain's processing of objects — such as, in this case, animals, real or depicted — is not primarily a linguistic function. Rather it is perceptual (Logothetis and Sheinberg 1996: 579). Focussing attention on the visual system — just as, thus far, we have focussed on visual perception — we turn specifically to its so-called 'what' pathway. The visual system at the back of the brain is broadly sketched in Fig.15. A double stream of visual information coming from the retina via intermediate midbrain areas (the faster stream associated with fuzzy peripheral vision and the slower stream associated with sharp foveal vision) is initially processed in the early visual areas

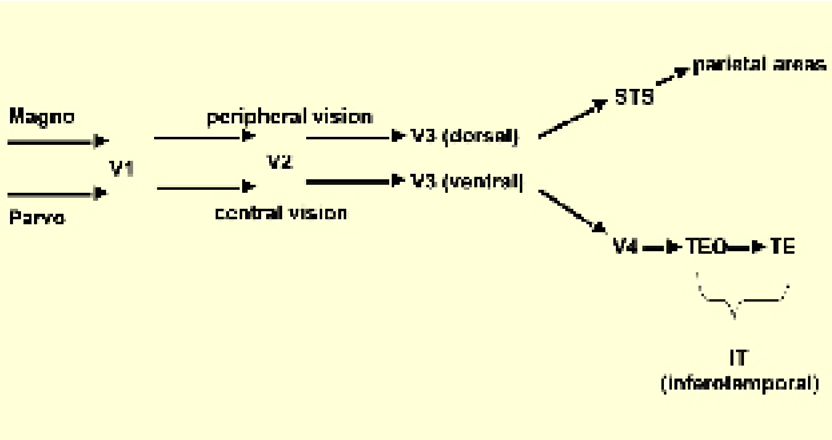

Figure 15. *The visual system.*

first examined by Hubel and Wiesel (1962), viz. V1 (the striate), followed by the extrastriate areas — all these being in the occipital lobe. At V3 there is a divergence identified by Ungerleider and Mishkin (1982), with a 'where' dorsal path through the superior temporal en route to the parietal lobe (this being concerned *inter alia* with the fast processing of perceived motion), and the above-mentioned 'what' ventral path to the inferior temporal via V4 (this being concerned *inter alia* with the slower processing of perceived objects). The inferior temporal (IT) has various divisions, in particular an area referred to as TEO and a final destination for visuals: TE. Whereas the early section of the visual system (V1) processes simple features of objects such as edges, IT is characterised by very large receptive neuronal fields, geared to the fine processing of the most complex objects (Logothetis and Sheinberg 1996: 605). To summarise: 'at each successive tier of processing mechanism, single neurons respond to increasingly more specific visual stimuli falling on an increasingly wider area of the retina' (Gross et al. 1972: 96). A clear account of the inferotemporal is given by Logothetis and Sheinberg (1996: 598–604), with stress on the determination of object shape (a response to 'optimal stimulus', p. 603) and, of special relevance here, orientation.

Faces and bodies: FFA, EBA and FBA

Work on object areas in the inferotemporal has concentrated on the processing of faces (for details of which we refer to the account of the fusiform gyrus and FFA or 'fusiform face area' in Dobrez 2012). Importantly for the present article, the FFA has also been interpreted as a '*flexible* fusiform area' able to process varieties of objects (Palmeri and Tarr 2008: 178ff.). Equally importantly for us was the finding that IT neuronal ensembles included neurons specialised to process face parts, and other neurons specialised for the entire face, that is, some neurons firing *only* for a given part, some *only* for the whole (Perrett et al. 1982; Desimone 1991; Rolls 1992; Tong et al. 2000). This may be presumed to apply to IT processing of any object. Some neuronal preference for parts is explicable in evolutionary terms: it is vital to identify an object that is partially occluded, i.e. to make a determination on the basis of a part (Wachsmuth et al. 1994: 519). This foreshadows our *pars pro toto* observations and discussion of salience below. We might expect, thinking along evolutionary lines, that the most useful parts will be those most salient and for which neurons will fire not merely at the TE terminal but from the start, at V1, as noted by Hodgson (2003: 6). Predictably, mouths (especially when active) and noses will feature in connection with the face. Tong et al. (2000: 271) specify eyes, nose, mouths, chin; combinations of these; as well as the 'overall configuration of the face'. Comparable commentary is found in Logothetis and Sheinberg (1996: 606). Neural preference for overall configurations indicates more holistic processes. As Logothetis et al. (1995: 561) put it, for objects of some complexity and biological importance, 'holistic representations may be the only ones possible' (see also Logothetis and Sheinberg 1996: 604ff. on 'combination encoding'). Significantly, we should not assume an additive process in the relation between part-only and whole-only neurons, i.e. we should not assume the brain 'adds up' neurons for eyes, mouths etc. to arrive at neurons for the whole face (Tong et al. 2000: 267). The same point is made by McCarthy et al. (1999: 441), with the observation, highly relevant here, that integration may occur by obtaining the whole by a mechanism of anticipation. Certainly some kind of holistic mechanism seems necessary in view of the time required for additive processes. At any rate we may say that TE cell groupings, with their varied specialisations, register maximum visual detail and integrate resultant complexity within their large receptive fields.

All this, however, does not take us beyond brain processing of faces. The fact is, less work has focused on body-processing and it has understandably concentrated on humans rather than animals. Nonetheless, when Wachsmuth et al. (1994) studied body-processing, they found that what was true for face-processing equally applied, viz. that there were neurons responsive to parts of the body, and neurons responsive to the body as a whole. Subsequently, an area dubbed the EBA or 'extrastriate body area' located in the right lateral occipito-temporal was identified by Downing et al. (2001) using fMRI scan. Cells in this area fired for living rather than inanimate objects, and for body parts and whole bodies presented as photographs, line drawings, stick figures or silhouettes. Face selectivity was low. It seems we have here a mechanism for dealing with the perception of other humans in situations in which the face is hard to make out. So the EBA would complement IT areas such as the fusiform gyrus which specialise for the identification of faces. At the same time, however, there is an argument on the basis of fMRI scan (Peelen and Downing 2005) that the mid fusiform gyrus almost equally selects for bodies (as against non-living objects). Following Schwarzlose et al. (2005), this area has been dubbed the 'fusiform body area' (FBA). So there would be a fusiform face area (FFA) and an FBA, both located away from Downing's extrastriate body area (EBA). It is worth noting in connection with face and body processing that Tsao et al. (2003) indicate comparable areas in the monkey superior temporal sulcus (STS). That the monkey STS processes faces like the human FFA is less important for us here than the fact that it also processes objects. So the neural evidence for the visual significance of biological bodies grows. But, suggestive as this may be, it does not specifically address the issue of animal bodies. As already pointed out, Downing et al. (2001) used animal images and showed a hierarchy of responses, highest for humans, intermediate for biological non-humans, lowest for inanimate objects. In the context of a discussion of animal depictions in palaeoart, Hodgson (2003) has alerted the present authors to the work of Damasio et al. (1996), highlighting, via PET scan, an area in the left IT selective for animals, and likewise to the work of Tanaka (1996: 111), which has some relevance to the perception of animals. Unfortunately, however, details remain sketchy, though there can be little doubt that neurophysiological research should confirm what we know about visual priorities: those of recognising living beings, human (faces and bodies) and animal. One last study is worth mentioning here: Tong et al. (2000) monitored FFA responses to both human and cat faces. The latter registered as strongly as the former. Other animal images were used, but add nothing to the present discussion, since the focus of the experiment was on face perception.

Orientation: object-centred vs viewer-centred cells

This brings us to a neural substrate for the processing of profile. The significance of object orientation in object perception has been much studied. The earliest experiments of Hubel and Wiesel (1962) showed orientation registering from the start, in the primary visual cortex (V1). (For an account see also Perrett et al. 1987: 358.) Celebrini et al. (1993) found that of 259 monkey cells selected in V1, about 2/3 responded to orientation, i.e. 2/3 of cells fired on the basis of the orientation of the visual object. At the other end of the process, the terminal point of IT, orientation is still selected for by cells — in addition, of course, to shape and other object features (Logothetis and Sheinberg 1996: 603). The seminal Gross et al. (1972) study indicated that cells firing for orientation of objects were equally sensitive to direction of movement. Of course it makes maximum difference if the carnivore is oriented towards us — and in motion! But what role might perception of objects in profile have in all this? Experiments have revealed that there are neurons in the object-processing area responsive regardless of orientation — and other neurons responsive only to particular orientations. In the case of face perception, and with orientation-sensitive neurons, frontal and profile figure prominently (Desimone 1991: 5). Researchers refer to the first class of neurons (those which fire for *any* object orientation) as 'object-centred' and to the second (which fire for *particular* orientations) as 'viewer-centred' (Gross 1992: 5). What is important for the present argument is that the temporal distinguishes *different views* of objects with extreme specialisation. Wachsmuth et al. (1994: 514) found 90% of cells in the monkey temporal to be orientation-responsive — and to the (human) body the monkeys were shown — with some of these cells specialising for profile, including *either* left *or* right profile. (One wonders whether exchange between such neurons is behind the perceptual switch in the rabbit-duck teaser!) Two facts are critical: that both object- and viewer-centred neurons operate in the visual system (Rolls 1992: 15; Logothetis and Sheinberg 1996: 579) — and that viewer-centred cells are considerably in the majority, with frontals and profiles probably at the fore.

In fact there has been debate on the question of object- vs viewer-centred models of visual processing, expressed as the distinction between 'viewpoint-invariance' (or 'viewpoint-independence') and 'viewpoint-dependence'. Clearly any argument for the role of salience in perception and for canonical views or, in our terminology, canonical form, more plausibly relies on viewer-centred responses or the idea of viewer-dependence. The Marr/Nishihara and the Biederman models postulate viewpoint-invariant structures in the visual system, something which cannot evidently account for the role of salience and canonical form outlined above. But, as noted, the viewpoint-invariance model appears to fit only a minority of cells. Unsurprisingly, then, despite the impact of Marr's computational modelling work and Biederman's geons thesis, many researchers have opted, with varying emphases, for viewpoint-dependence — while allowing for the operation of two distinct types of neurons. We refer to Tarr and Pinker (1989), Backer Cave and Kosslyn (1993), Celebrini et al. (1993), Wachsmuth et al. (1994), Edelman and Bülthoff (1992), Bülthoff et al. (1994), Logothetis et al. (1995), and Logothetis and Sheinberg (1996). What follows from the viewpoint-dependence thesis is that we may have begun to identify neuronal substrates for salience and canonical form. In this vein Wachsmuth et al. (1994: 511) make mention of 'optimal views' and Perrett et al. (1992: 24) of 'characteristic views'. Indeed it is the 1992 Perrett et al. article which prioritises frontals and profiles over other views. We are not far from our own argument here, though the evidence remains partial since it does not (to our knowledge) zero in on animal profiles. Given its evolutionary relevance, this is neurophysiological work that should be done, in tandem with the psychology work we have also been referencing. At the same time, what is emerging is a *framework* for a neurophysiological account calculated to support our argument. Manifestly the orientation of objects we see is registered by the brain as being — what else? — of great importance. Moreover there is plenty of evidence for neuronal preference for particular views, chief of these, presumably, being frontals and profiles. (Of course the fact that we are able to perceive — if not quickly identify — even the *least* canonical views means that there are neurons processing these as well!) At any rate it seems not unreasonable to apply the above to profile animals as, broadly speaking, a preferred view. However, while there would seem to be neuronal evidence for our argument, we have not as yet nominated a specific neural substrate for the phenomenon of *salience* which, as emphasised above, underpins the concept of canonical form.

Fast-track processing

Given that recognition via salient features operates by way of *pars pro toto*, the major neural requirement would be speed of processing. Now the faster visual pathway is the dorsal to the parietal lobe which passes through the superior temporal. However, we note that this pathway is geared, among other things, to the perception of movement rather than objects. For the slower route — the ventral to the inferotemporal object-processing area — we take it that one way to generate speed might be to skip some processing stages. Tanaka (1996: 110) speaks of 'jumping projections' (a leap from V2 to TEO, bypassing V4; another from V4 to TE, bypassing TEO). But he also points out that 'step-by-step projections are more numerous'. In other words, processing more usually involves *all* the stages of the pathway. There is another option, one we foreshadowed when underlining the phenomenon of 'anticipation' above in reference both to Hochberg and to McCarthy et al. Athletes anticipate motion, with heart-rate changes prior to actual movement. A psychological example of this might be the lecturer mentally preparing to deliver a lecture. In each case body chemistry is altered in preparation for coming activity. Now in neural terms, signals proceeding step-by-step along a given visual path are constantly affected by feedback. There is a two-way passage of information which allows the system to confirm, or modify, adjust or correct, incoming signals. Thus a perception is able to be fine-tuned, or to be reconsidered ('yes, a kangaroo, in fact an eastern grey' — or, 'it seemed a dog, but was in fact a kangaroo'). All this, if unclear at the start, would call for saccadic shifts allowing foveal fixation on different aspects of the object. However, this neural feedback requires a minimum of time to come into play. What about speeding things up? This would involve *feed-forward*, in this case producing recognition so rapid as to *forestall* feedback.

We have already alluded to researchers experimenting on feed-forward using human subjects (Thorpe et al. 1996, 1999, 2001; Van Rullen and Thorpe 2001; Van Rullen and Koch 2003; Delorme et al. 2000) in connection with debate about categorisation and saccadic eye motion. The supposed findings were that recognition can occur ultra-rapidly, i.e. prior to foveal fixation on the target object. With ultra-rapid recognition saccadic motion would be pre-empted. When experimenters showed pictures including animals for fractions of time, they found subjects registering with remarkable speed. In the event we have already expressed the view that these experiments probably do not relate to object-recognition, properly speaking. Rather they indicate a registering of 'something there' on the basis of prompting. At this point in our argument, however, this is less relevant than the likelihood that neural feed-forward mechanisms are involved in visual processing. Even if these mechanisms activate so early as to precede object-recognition proper, they may explain instances of great perceptual speed and so provide a possible neural underpinning for *pars pro toto* perception — seeing the whole in a given salient part. So on present evidence we have 'jumping projections' and also feed-forward as candidates for understanding the neural basis for the fact that salience prompts us to 'anticipate'. Indeed feed-forward researchers allow for salience in objects and accept the idea of anticipation (Thorpe et al. 1996: 522; Celebrini et al. 1993: 823). Celebrini et al. refer to 'optimal stimuli' which prioritise pertinent information. Oram and Perrett (1992), like Celebrini et al. working on monkey responses, showed discrimination between diverse views of faces. But recognition was very rapid, explicable only in feed-forward terms. The pinpointed response was in the STS, the monkey equivalent of human IT, and probably involved eight synapses through four visual areas from V1 to STS — such that 'the flow of information had to be entirely feed-forward' (1992: 81). In other words 'the first cell to transmit information to the next level becomes a "winner-takes-all" in purely feed-forward manner by using an inhibitory "veto" of late spike arrivals from competing inputs' (1992: 82). To us this has all the flavour of an argument for salience as the key to the recognition of canonical form — and for a part-for-whole understanding of salience.

Memory and the synapse

We think that the above provides, if not point-for-point support for our case for the perception of canonical form and its neural substrate, at least a framework within much or all of our argument might be positioned. However, the recognition of the canonical form of a kangaroo or giraffe etc. presupposes we have seen the animal in the past, in short, that we retain it in the memory. Visual memory may be envisaged as a process of abstraction, but we take it the dominant contemporary view is that it is perceptual. According to 'image-based' theories of visual memory, we store percepts, not concepts (Palmeri and Tarr 2008: 166, 171). Of course percepts are not 'pictures' in the brain, which would require the presence of a so-called homunculus to observe them and a second homunculus to observe the first and so on, regressively. Percepts are stored in the brain not as pictures but as coded information. There are in fact several types of visual memory, two of which require mention. Visual 'working' or 'short-term memory' (VSTM) ensures continuity of perception across saccades, blinking and short-term occlusion, and works by sustained firing of neurons. Visual 'long-term memory' (VLTM), on the other hand, effects *changes* in neurons, often permanently. We refer the reader to Hollingworth and Luck (2008: 5–7) for an account of the above. It seems the development from short- to long-term memory involves the hippocampal area (Rose 1987: 457) or, in greater detail, subregions of the medial temporal lobe (Chun and Turk-Browne 2008: 229). It must be the case that recognition is reliant on long-term memory, but what is the neural relation between them, i.e. between neural seeing and remembering?

We pointed out above that the contemporary view is that we see, imagine — and remember — in closely-related ways (Palmeri and Tarr 2008: 163, 194; Logothetis and Sheinberg 1996: 613). Specifically, the same neuronal ensembles fire for seen, imagined and remembered objects. This means that visual long-term memories of given objects are stored in those same neural areas in which the perception of those objects is processed. As we have seen, this will be the inferotemporal (Hollingworth and Luck 2008: 7, 151; Rose 1987: 458). But if short-term memory depends on sustained neuronal activity, while long-term depends on neuronal change, is this capacity for change a feature of some or all cells? Rose (1987: 460) raises the question and Deutsch (1987: 462) suggests it may be a capacity of most neurons. Where, then, in the neuron, is change to be located? The celebrated psychologist Donald Hebb was the first to nominate the synapse as the likely area, and this now appears generally accepted (Rose 1987: 458; Greenfield 1997: 177; Hollingworth and Luck 2008: 7). Greatly simplifying, we may say that the neuron, which consists of a cell body, a varying pattern of surrounding ramification (dendrites), and a large fibre branch (the axon), receives signals through its multiple minuscule dendrites, then transmits them via the axon to another neuron. If it is 'excited' it raises the frequency of its own firing and these impulses pass from neuron to neuron along a given path across the gaps between neurons, viz. synapses. The synapse is the neural zone in which electric signals from one neuron activate chemical changes that will affect the next neuron once more in the form of electrical signals. It is in the chemistry of the synaptic space that we locate neuronal change. A long-term visual memory is one that involves synaptic modifications along a particular neural path. According to Greenfield (1997), from whom we draw the above description of the neural signalling process, the key element in the chemistry of cell modification appears to be calcium.

Conclusion

The key evolutionary role of visual recognition and its dominant influence on the way in which humans have depicted (among other things) animals, not least, but not solely, in rock art, cannot be doubted. We have foregrounded some of the essential perceptual principles involved in recognition, both in life and art, and sought to relate these to possible neural substrates. So in summary and on the basis of partial but still considerable evidence from observation, experiments in psychology and findings in neurophysiology, we may hypothesise that we best recognise or most easily perceive a kangaroo or giraffe or llama — whether real or depicted — in a dominant view, usually profile, and via the part-for-whole mechanism of salience. We have referred this to the idea of seeing-as and termed it canonical form. What we quickly recognise is a particular canonical form, probably characterised at basic level. Canonical form underpins the possibility of iconic representation — not to be confused with culture-specific conventions of Realism. The way it works through perception of the whole in the part is better understood holistically than additively — not least in the context of our understanding of the role of saccadic eye movement in the process of recognition. What contribution can neurophysiology make to this discussion? To begin with, it must be emphasised that a neural explanation for perceptual (or any other) phenomena does not constitute, as it were, *the* explanation. Neurophysiology merely alerts us to the brain mechanisms engaged in seeing; it identifies the neural correlates of the perceptual experience — and in so doing adds to our knowledge of that experience without explaining it away. With this proviso, it seems highly relevant for us to find a degree of *matching* between the experience of looking at an animal and present knowledge as to how we process what we see. It appears that objects — especially faces, but also bodies, and, we conjecture on incomplete evidence, animal bodies — are terminally processed in the inferior temporal area, where long-term memory of such objects is also located. The processing apparently involves both holistic and non-holistic aspects, with the former predominating — assuming we may take viewer-centred neurons as operating holistically (doubtless in close association with object-centred neurons). Last but not least it seems evident that the neural visual system is extremely sensitive to the orientation of objects — presumably left or right profile for animals — and to features of an object that stand out, i.e. constitute salient features, whose part-for-whole operation finds its correlate in the fast-tracking mechanism of feed-forward. We think all this adds up to a plausible and to an extent testable matching of our experience of seeing a kangaroo with available knowledge of brain mechanisms minutely and invisibly processing the array of light which gives us the visual experience of the kangaroo.

Acknowledgments

We have been greatly encouraged in our present line of inquiry by B. Watson's recent investigation into the literature of neuroscience and its relevance to rock art studies. We thank the Kimbell Art Museum, Fort Worth, Texas, for permission to reproduce the painting by George Stubbs, and Iziko Museums of Cape Town Social History Collections (Archaeology) for Linton Panel B.

Dr Livio Dobrez
College of Arts and Social Sciences
A. D. Hope Building
Australian National University
Canberra, ACT 0200
Australia
dobrezl@grapevine.com.au

Dr Patricia Dobrez
9 Blair Street
Watson, ACT 2602
Australia
u4045668@anu.edu.au

Final MS received 15 January 2013.

REFERENCES

Arnheim, R. 1969. *Visual thinking*. University of California Press, Berkeley and Los Angeles, California.

Arnheim, R. 1974. *Art and visual perception, a psychology of the creative eye: the new version*. University of California Press, Berkeley and Los Angeles.

Attneave, F. 1954. Some informational aspects of visual perception. *Psychological Review* 61(3): 183–193.

Backer Cave, C. and S. M. Kosslyn 1993. The role of parts and spatial relations in object identification. *Perception* 22: 229–248.

Baylis, G. C. and J. Driver 1995. One-sided edge assignment in vision: 1. Figure-ground segmentation and attention to objects. *Current Directions in Psychological Science* 4(5):140–146.

Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychological Review* 94(2): 115–147.

Blanz, V., M. J. Tarr and H. H. Bülthoff 1999. What object attributes determine canonical views? *Perception* 28: 575–596.

Bülthoff, H. H., S. Y. Edelman and M. J. Tarr 1994. How are three-dimensional objects represented in the brain? A.I. Memo No. 1479, C.B.C.L. Paper No. 96, pp. 1–19. Retrieved from: *publications.ai.mit.edu* [pathname: ai-publications/1994/AIM-1479.ps.Z] Massachusetts Institute of Technology, Boston.

Buswell, G. T. 1935. *How people look at pictures: a study of the psychology of perception in art*. The University of Chicago Press, Chicago.

Celebrini, S., S. Thorpe, Y. Trotter and M. Imbert 1993. Dynamics of orientation coding in Area VI of the awake primate. *Visual Neuroscience* 10: 811–825.

Cheyne, J. A., L. Meschino and D. Smilek 2009. Caricature and contrast in the Upper Paleolithic: morphometric evidence from cave art. *Perception* 2009(38): 100–108.

Chun, M. N. and N. B. Turk-Browne 2008. Associative learning mechanisms in vision. In S. J. Luck and A. Hollingworth (eds), *Visual memory*, Oxford Series in Visual Cognition, pp. 209–245. Oxford University Press, Oxford.

Clegg, J. 1991. !Pictures and pictures of ... . In P. Bahn and A. Rosenfeld (eds), *Rock art and prehistory: papers presented to Symposium G of the AURA Congress, Darwin 1988*, pp. 109–111. Oxbow Monograph 10, Oxbow Books, Oxford.

Damasio, H., T. J. Grabowski, D. Tranel, R. D. Hichwa and A. R. Damasio 1996. A neural basis for lexical retrieval. *Nature* 380: 499–505.

Delorme, A., G. Richard and M. Fabre-Thorpe 2000. Ultra-rapid categorization of natural scenes does not rely on color cues: a study in monkeys and humans. *Vision Research* 40: 2187–2200.

Deręgowski, J. B. 1984. *Distortion in art: the eye and the mind.* Routledge and Kegan Paul, London.

Deręgowski, J. B. 1995. Perception — depiction — perception, and communication: a skeleton key to rock art and its significance. *Rock Art Research* 12: 3–10.

Deręgowski, J. B. and commentary 1989. Real space and represented space: cross-cultural perspectives. *Behavioural and Brain Sciences* 12: 51–119.

Desimone, R. 1991. Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience* 3(1): 1–8.

Deutsch, J. A. 1987. Memory: experimental approaches. In R. L. Gregory (ed.), *The Oxford companion to the mind*, pp. 460–463. Oxford University Press, Oxford.

Dobrez, L. 2012. American ikon: how to choose an ARARA logo. In E. W. Ritter, M. Greer, and P. Whitehead (eds), *American Indian Rock Art*, Volume 38, pp. 145–164. American Rock Art Research Association, Glendale, AZ.

Downing, P. E., Y. Jiang, M. Shuman and N. Kanwisher 2001. A cortical area selective for visual processing of the human body. *Science* 293: 2470–2473.

Edelman, S. and H. H. Bülthoff 1992. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research* 32(12): 2385–2400.

Gibson, J. J. 1950. *The perception of the visual world*. Houghton Mifflin Company, Boston.

Gibson, J. J. 1966. *The senses considered as perceptual systems*. Houghton Mifflin, Boston.

Gibson, J. J. 1971. The information available in pictures. *Leonardo* 4(1): 27–35.

Gibson, J. J. 1979. *The ecological approach to visual perception*. Houghton Mifflin Company, Boston.

Gombrich, E. H. 1960. *Art and illusion: a study in the psychology of pictorial representation.* Phaidon Press, London.

Greenfield, S. 1997. *The human brain: a guided tour*. Orion Books, London.

Gross, C. G. 1992. Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London B* 335(1273): 3–10.

Gross, C. G., C. E. Rocha-Miranda and D. B. Bender 1972. Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology* 35: 96–111.

Hagen, M. A. 1986. *Varieties of realism: geometries of representational art*. Cambridge University Press, Cambridge and Melbourne.

Halverson, J. 1992. Paleolithic art and cognition. *Journal of Psychology* 126(3): 221–236.

Halverson, J. 1995. Information and typicality. Reply to Deręgowski. *Rock Art Research* 12: 14–15.

Hochberg, J. 1972. The representation of things and people. In E. H. Gombrich, J. Hochberg and M. Black (eds), *Art, perception, and reality*, pp. 47–94. The Johns Hopkins University Press, Baltimore and London.

Hochberg, J. 1978. Art and perception. In E. C. Carterette and M. P. Friedman (eds), *Handbook of perception X*, pp. 225–258. Academic Press, New York.

Hochberg, J. 1987. Gestalt theory. In R. L. Gregory (ed.), *The Oxford companion to the mind*, pp. 288–291. Oxford University Press, Oxford.

Hodgson, D. 2003. The biological foundations of Upper Palaeolithic art: stimulus, percept and representational imperatives. *Rock Art Research* 20: 3–22.

Hollingworth, A. and S. J. Luck 2008. Visual memory systems. In S. J. Luck and A. Hollingworth (eds), *Visual memory*, pp. 3–8. Oxford Series in Visual Cognition, Oxford University Press, Oxford.

Hubel, D. H. and T. N. Wiesel 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160: 106–154.

Kihlstrom, J. F. 2006. A new reversible figure and an old one. Paper presented at the annual meeting of the Society of Experimental Psychologists, La Jolla, CA, March 15, 2006. Available at: *http://socrates.berkeley.edu/~kihlstrm/SEP06.htm*

Kosslyn, S. M. 1994. *Image and brain: the resolution of the imagery debate*. The Massachusetts Institute of Technology Press, Cambridge, MA.

Logothetis, N. K., J. Pauls and T. Poggio 1995. Shape representation in the inferior temporal cortex of monkeys. *Current Biology* 5(5): 552–563.

Logothetis, N. K. and D. L. Sheinberg 1996. Visual object recognition. *Annual Reviews Neuroscience* 19: 577–621.

Lowe, D. G. 1987. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence* 31(3): 355–395.

McCarthy, G., A. Puce, A. Belger and T. Allison 1999. Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex* 9: 431–444.

Marr, D. and H. K. Nishihara 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B* 200(1140): 269–294.

Navon, D. 1977. Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology* 9: 353–383.

Oram, M. W. and D. I. Perrett 1992. Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology* 68(1): 70–84.

Palmer, S. E. 1977. Hierarchical structure in perceptual representation. *Cognitive Psychology* 9: 441–447.

Palmeri, T. J. and M. J. Tarr 2008. Visual object perception and long-term memory. In S. J. Luck and A. Hollingworth (eds), *Visual memory*, pp. 163–207. Oxford Series in Visual Cognition, Oxford University Press, Oxford.

Peelen, M. V. and P. E. Downing 2005. Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology* 93: 603–608.

Perrett, D. I., E.T. Rolls and W. Caan 1982. Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47: 329–342.

Perrett, D. I., J. K. Hietanen, M. W. Oram and P. J. Benson 1992. Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London* B 335(1273): 23–30.

Perrett, D. I., A. J. Mistlin and A. J. Chitty 1987. Visual neurons responsive to faces. *Trends in Neuroscience* 10(9): 358–364.

Rolls, E. T. 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London B* 335(1273): 11–21.

Rosch, E. and C. B. Mervis 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605.

Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson and P. Boyes-Bream 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382–439.

Rose, S. 1987. Memory: biological basis. In R. L. Gregory (ed.), *The Oxford companion to the mind*, pp. 456–460. Oxford University Press, Oxford.

Schwarzlose, R. F., C. I. Baker and N. Kanwisher 2005. Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience* 25(47): 11055–11059.

Stanbury, P. and J. Clegg 1990. *A field guide to Aboriginal rock engravings with special reference to those around Sydney*. Sydney University Press, Sydney,

Tanaka, K. 1996. Inferotemporal cortex and object vision. *Annual Reviews Neuroscience* 19: 109–139.

Tarr, M. J. and S. Pinker 1989. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* 21: 233–282.

Thorpe, S. J., K. R. Gegenfurtner, M. Fabre-Thorpe and H. H. Bülthoff 1999. Categorisation of complex natural images in extreme peripheral vision. *Perception* 28 (Suppl.):61.

Thorpe, S. J., K. R. Gegenfurtner, M. Fabre-Thorpe and H. H. Bülthoff 2001. Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience* 14: 869–876.

Thorpe, S., D. Fize and C. Marlot 1996. Speed of processing in the human visual system. *Nature* 381: 520–522.

Tinbergen, N. 1960. *The herring gull's world: a study of the social behaviour of birds*. Revd edn. Basic Books, New York.

Tong, F., K. Nakayama, M. Moscovitch, O. Weinrib and N. Kanwisher 2000. Response properties of the human fusiform face area. *Cognitive Neuropsychology* 17(1/2/3): 257–280.

Tsao, D. Y., W. A. Freiwald, T. A. Knutsen, J. B. Mandeville and R. B. H. Tootell 2003. Faces and objects in macaque cerebral cortex. *Nature Neuroscience* 6(9): 989–995.

Tversky, B. and K. Hemenway 1984. Objects, parts, and categories. *Journal of Experimental Psychology: General* 113(2): 169–193.

Ungerleider, L. and M. Mishkin 1982. Two cortical systems. In D. J. Ingle, M. A. Goodale and R. J. W. Mansfield (eds), *Analysis of visual behaviour*, pp. 549–586. The Massachusetts Institute of Technology Press, Cambridge, MA.

VanRullen, R. and C. Koch 2003. Visual selective behaviour can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience* 15(2): 209–217.

VanRullen, R. and S. J. Thorpe 2001. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception* 30: 655–668.

Wachsmuth, E., M. W. Oram and D. I. Perrett 1994. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex* 5: 509–522.

Welch, D. M. 2012. Oh dear! No deer! *Rock Art Research* 29: 171–177.

Wittgenstein, L. 1968 [1953]. *Philosophical Investigations*, transl. by G. E. M. Anscombe. Reprint of English text with index. Basil Blackwell, Oxford.

Woodhouse, B. [n.d.]. *Bushman art of South Africa*. Art Publishers, Durban.

Wollheim, R. 1973. *On art and the mind: essays and lectures*. Allen Lane, London.

Wollheim, R. 1998. On pictorial representation. *The Journal of Aesthetics and Art Criticism* 56(3): 217–226.

RAR 30-1071