WHAT IS A HAND? WHAT IS A MIND?

Mohan MATTHEN

1. Movements in the philosophy of mind often turn on different approaches to definitional questions regarding the entities that are its subject-matter — the perceptual systems, the emotions, rationality, and the like. For example, Gilbert Ryle's philosophical behaviourism was motivated by the suggestion that the mind should be understood as a complex of behavioural dispositions, and the functionalism of Jerry Fodor and David Lewis by the idea that it should be modularly analysed in terms of the causal roles mental capacities occupy in the processing and linguistic representation of information. Ruth Millikan suggests another type of analysis, in terms of the historical contribution that "language, thought, and other biological categories" make (or more accurately "made", since hers is a historical analysis) towards the increased reproductive fitness of an organism. Finally, the idea that mental entities be defined in terms of intentional content or experienced character — these being treated as something unanalysable in terms of other approaches — has seen something of a revival.

In this article, I am concerned with a factor that is not much considered by philosophers of mind. It is generally conceded nowadays that, whatever else they might be, mental organs are biological in nature. Evolutionary biologists tend to characterize organs as homologous kinds. So it seems worth asking whether the notion of a homology ought somehow to enter into the ways in which we conceive of mental organs. My task will be to explore some of the ways in which it does. I shall be arguing that the notion of a homology essentially modifies the movements in philosophy of mind listed above.

2. The notion of a homology was first used in the context of the idea, which finds its place in the thought of the "transcendantalist" natu-

© Revue Internationale de Philosophie 4/2000 - n° 214 - pp. 653-672.

ralists — the German “Nature Philosophers” (notably Goethe and Oken) and some nineteenth century French naturalists (particularly Etienne Geoffroy Saint-Hilaire) — that all organisms are built according to an architectural plan out of interlocking structural elements. Within the confines of such a plan, such interlocking elements might be present in different animals in a distorted or topographically transformed form, but each had a definite position relative to others. When an element in one organism was identified as an instance of the same architectural item as that in another, they were called “homologues” (or sometimes “affinities”). In this section, I shall show that this notion is not as yet rich enough to be able to give us satisfactory answers to questions concerning the classification of *organs*. It cannot, for example, give us an answer to the question, “What is a hand?”

The English naturalist, Richard Owen, was perhaps the first to develop the idea of homology in a precise and theoretical manner. Owen was famous for his articulation of the “vertebrate archetype,” which, according to him, consisted of a linked series of “vertebrae,” each consisting of central bony hub attached to a pair of “arches” arrayed in the dorsal-ventral direction, and ribs projecting outwards, perpendicular both to the plane of the arches and to the longitudinal axis in which the vertebrae themselves line up. Owen’s thesis was that all vertebrate organisms were constructed according to this plan, the differences amongst them resulting from modifications of the vertebrae themselves. For instance, the tetrapods have skulls and arms, and more primitive vertebrates do not. This looks like a modification of the plan itself, but according to Owen it is not. The skulls consist, according to Owen, of modified vertebrae fused together, and the forelimbs are actually the modified arches and ribs that belong to these.

Owen labelled the correspondence of a skeletal item in a particular organism to one in an archetype the relation of “general homology”(1) — thus each vertebra in the human skull is a “general homologue” of, or “general-homologous” with, a particular vertebra in the archetype. We may define general homology as follows:

> *X* is *G*-homologous with *Y* if *Y* is a distinguished item in an abstract archetype, and *X* instantiates *Y* in some particular animal.

The transcendentalists regarded general homology as a relation like that of universal and particular. It is a source of posits concerning the underlying character (or “meaning” — *Bedeutung*, as they would have said) of various skeletal structures in particular animal types.

Owen was struck by the repetition of elements in archetypes: for example, the multiple iterations of the vertebral structure in the vertebrate archetype. Owen supposed that such repetitions were the result of a “polarising force” similar to that which he took to be responsible for crystallization in the mineral world, and for the profusion of leaves and branches etc in the vegetable kingdom. He called such repeated parts of the same animal “homotypes” — each of our vertebrae is a homotype of the others, and each vertebra in the archetype is a homotype of the others. Following Alec Panchen,(2) I shall call homotypes *iterated* homologues.

> *X* is *I*-homologous with *Z* if *X* and *Z* are
>
> a) repetitions in an archetype of the same ideal structure,
>
> or
>
> b) repetitions in a particular type of animal of the same item in the archetype, (i.e. if *Y* is a distinguished item in the archetype, and *X* and *Z* are *G*-homologues of *Y*. (3)

---

(1) On Richard Owen, see especially Nicolaas A. Rupke, “Richard Owen’s Vertebrate Archetype,” *Isis* **84** (1993): 231-51, and *Richard Owen: Victorian Naturalist* (New Haven: Yale UP, 1994), chapters 3-4, and Alec L. Panchen, “Richard Owen and the Concept of Homology,” in B.K. Hall (ed.) *Homology: the Hierarchical Basis of Comparative Biology* (San Diego: Academic Press, 1994): 21-62. The following works are also useful (*ad loc*): Martin J.S. Rudwick, *The Meaning of Fossils: Episodes in the History of Palaeontology* 2nd edition (New York: Science History Publications, 1976), Dov Ospovat, *The Development of Darwin’s Theory: Natural History, Natural Theology, and Natural Selection 1838-1859* (Cambridge: Cambridge UP, 1981), and Adrian Desmond, *Archetypes and Ancestors: Palaeontology in Victorian London, 1850-1875* (London: Blond and Briggs, 1982).

(2) Panchen, *Classification, Evolution and the Nature of Biology* (Cambridge: Cambridge UP, 1992): 70.

(3) Note that repetition can occur both in the archetype (which consists of an iterated series of vertebrae), and also in a particular type of organism. (It is the latter that accounts for the fact that a snake, for instance, might have many more vertebrae than a mammal.) It is a bit unclear why there is a need for both mechanisms. Given the existence of a “polarizing force”, would it not have made sense for Owen to posit only items like the vertebra as ideal entities, and allow the polarising force to account for the occurrence of linked iterated homologues in actual animals? Such a move would have eliminated the archetype (but retained ideal structural elements), and given Owen more flexibility in dealing with vertebrate skeletons with different numbers of elements.

The correspondence of elements in *different* animals constitutes a third (and for us the crucial) relation of homology. If two elements in different animals are *general* homologues of the same item in the vertebrate archetype, they are *special* homologues of each other. Thus:

> *X* is *S*-homologous with *Z* if *X* and *Z* are parts of distinct animals, and there is some *Y* such that *X* and *Z* are both *G*-homologous with *Y*.

It is this relationship that corresponds to the contemporary notion of homology. According to Owen, a particular bone in the human hand is a homologue of one in the wing of a bat, because each corresponds to the same bone in the vertebrate archetype. As Michael Ruse has said:

> … special homologies can be explained as two general homologies back to back, eliminating the middle term, the archetype.(4)

This is stated in a slightly misleading way. The “middle term” is never completely “eliminated,” since the relationship of special homology implies a common term to which *X* and *Z*, the special homologues, are compared. What Ruse means is that, in contrast to what holds of the first two homology relations, neither term of the special homology relation is an archetypal element.

Some interesting logical characteristics of Owen’s conception of special homology should be noted immediately. First, and most important, it is an absolute notion: given two items in distinct animals, either they are *S*-homologous or they are not. To appreciate the significance of this, let us compare Owen’s thesis with the somewhat later proposal of the biblical scholar Paul Anton de Lagarde that “all manuscripts of the Hebrew Bible go back to a single authoritative text from the early part of the second century A.D.” (5) Call that text **A**, the “archetype.” Now imagine that in a certain place in **A**, there occurs the word ‘PQR’. (See Figure 1.) Two copies **B** and **C** are made from **A**, and in **B**, the copyist mistakenly writes ‘POR’. (No such mistake is made in **C**.) Two copies are made from **B**: in **B’** the copyist writes ‘POR’, copying faithfully from his source, but in **B***, the copyist corrects the error, and writes ‘PQR’ again. Now, applying Owen’s

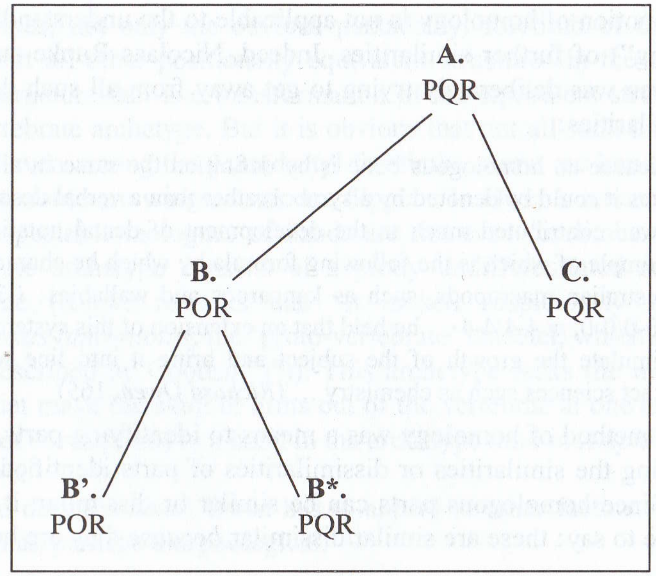

Figure 1

definitions as rendered above, *all* these words in the four copies are *S*-homologous with one another, because all are *G*-homologous with PQR in the archetype. There is no temporal layering in Owen’s scheme, as there is in the example of manuscripts; copies of copies must be treated simply as instances. Though Owen recognizes a progressive modification of animals further and further away from the archetype, he does not emphasize in this context that the later items in this ordering are modified from earlier items; each animal in the progression is an instance of the archetype. Perhaps this is a consequence of his “Platonism.” Each item in a concrete animal is an *instance* of an item in the archetype, and there is no such thing as an instance of an instance.

This leads to two consequences that are somewhat counter-intuitive from a contemporary perspective. The ‘O’ in **B’** and the ‘Q’ in **B*** are dissimilar. Nevertheless, they are (as we have just seen) *S*-homologues, that is, each is what corresponds to the middle item of ‘PQR’ in **A**. When Owen asks whether two things are *S*-homologous, he assumes that one is asking about their position as defined by some archetypical scheme. Do they occupy the same position or not?

---

(4) RUSE, *The Darwinian Revolution* (Chicago: U of Chicago Press, 1979): 119.

(5) See RUPKE, “Owen’s Vertebrate Archetype,” 231.

Owen's notion of homology is not applicable to the understanding (or "meaning") of further similarities. Indeed, Nicolaas Rupke suggests that Owen was deliberately trying to get away from all such descriptive similarities:

> Because an homologous bone is by definition the same in all verterates it could be denoted by a symbol rather than a verbal description. Owen contributed much to the development of dental notations, an example of which is the following formula by which he characterized Australian macropods, such as kangaroos and wallabies: *i* 3-3/1-1, *c* 0-0/0-0, *m* 4-4/4-4 … he held that an extension of this system would stimulate the growth of the subject and bring it into line with the exact sciences such as chemistry … (*Richard Owen*, 165)

Owen's method of homology was a means to identifying parts, not to explaining the similarities or dissimilarities of parts identified as the same. Since homologous parts can be similar or dissimilar, it makes no sense to say: these are similar/dissimilar *because* they are homologous.

This leads to the second consequence. The occurrences of PQR in **C** and **B*** are *S*-homologues: it makes no difference that the similarity between the two — the fact that they are both 'Q's — cannot be explained by the fact that both are copied from the archetypal item. For Owen, each species is an instance of the archetype; he apparently believed in some kind of transformation of animal kinds, but evolution is not, for him, a series of copies of copies in the way that Lagarde's biblical manuscripts are. Homology is correspondence with respect to the archetype, and that is all that it is. Consequently, Owen lacks the room to say something like: the occurrences of 'Q' in **C** and **B*** are homologous with respect to their position, because their position can be traced back to an archetypal item which has that position, but they are not homologous with respect to being 'Q' because there is no archetypal item to which this feature can be traced. These results indicate a limitation in the applicability of Owen's homological scheme to questions of the classification of organs. Once it is given that the item that occupies the place between 'P' and 'R' counts as a transformation of the 'Q' in the original archetype, Owen's conception of homology is exhausted.

Suppose then that one were to ask Owen: "What is a hand?" He cannot say: "organs *S*-homologous with the human hand." The *S*-homology relation would pull in every equivalent structure among the vertebrates, not only the obvious pentadactyl forelimb of the tetrapoda, but all other positionally equivalent structures in the class — every structure that is a transformation of the equivalent structure in the vertebrate archetype. But it is obvious that not all such things are *hands*: even among the pentadactyl forelimbs, some are hands, some are feet, some are wings, some are flippers. In Owen's scheme, some of the special-homologues of hands are not even forelimbs. Owen's vertebrate archetype consists of a pretty undifferentiated series of vertebrae (Rupke remarks that "it looked suspiciously like the *Amphioxus lanceolotus*, the 'proto-vertebrate' lancelet, which had just been described by Goodsir." (6)). This archetype lacks the modifications that make the skull or arms out of the vertebrae at one end. The item that corresponds to a hand in the archetype (and in *Amphioxus*) is just a rib.

So to define a *hand*, Owen has to appeal to other factors, perhaps functional, perhaps morphological.

3. The logic of the *evolutionary* theory of homologies more closely parallels the historical orientation of de Lagarde's thesis concerning the manuscript origin of the Hebrew Bible. It is by now a truism that the theory of evolution historicized the pre-existent idea of a structural homology, substituting ancestors for archetypes. Two items are understood as homologous in evolutionary theory if they are evolutionary transformations of the same item in a common ancestor.

We can transform Owen's definitions of homology by putting an actual concrete ancestor in place of his abstract archetype:

> *X* is *G*-homologous (in the evolutionary sense) with *Y* if *Y* is a distinguished item in some animal *A*, and *X* is an evolutionary transformation of *Y* in an evolutionary descendant of *A*.
>
> *X* and *Y* are *S*-homologous (in the evolutionary sense) if *X* and *Y* are parts of distinct types of animal, and there is a *Z* such that *X* and *Y* are both *G*-homologous (in the evolutionary sense) with *Z*.

In these definitions, the ancestor animal takes the place of the archetype, with the relation of evolutionary transformation replacing that of

---

(6) *Richard Owen*, 190.

general homology. The idea of parts of animals with common ancestry takes the place of Owen's "special homology," and I will simply use the term "homology" to denote this relation, assuming now an evolutionary interpretation.(7)

The ancestor relationship has many layers, since an ancestor of an ancestor is an ancestor. Two things that lack a common ancestor a certain number of generations back may still have a common ancestor a larger number of generations back. To go back to the example of manuscripts copied from **A**, there is now no absolute answer to whether the occurrences of 'Q' in **B*** and **C** are homologous or not. Relative to the original manusript **A**, they are homologous. Relative to **B** they are not. This relativity allows us to apply the notion of homology to the explanation of features other than position. The 'Q's in **B*** and **C** constitute a homologous-similarity with respect to their position, but not with respect to being 'Q's. There is no common ancestor of the text to which the 'Q's can be traced, but there is a common ancestor to which their position can be traced. Consequently, a similarity between *X* and *Y* can be explained homologically if there is a common ancestor to which their shared feature can be traced. The 'Q's in **B*** and **C** cannot be explained homologically, but the 'P's can, as can the position of the 'Q's between 'P' and 'R'.

The historical or evolutionary version of Owen's theory is able, then, to consider the provenance of similarities *within* a homology, i.e. on whether a particular similarity between two homologous items is traceable to a common ancestor which shares the feature in question. Return then to the question of what a hand is. We might try to build a functional component into the notion. For example:

> *M* A hand is a tetrapod pentadactyl forelimb capable of grasping.

In Owen's conception, this definition does not delineate a homology: as we have seen, it is a sub-class of a homology. However, the class defined by *M* — call it {*M*} — might still be homological in the evolutionary sense. In other words, it might well be that all the members of {*M*} have been replicated, in the apposite evolutionary sense, from a corresponding item in an ancestral organism, i.e. from an ancestor that is a pentadactyl forelimb capable of grasping. Thus {*M*} might be "a homology within a homology." (8) Relative to the common function of grasping, the hand is a small homology consisting of some pentadactyl forelimb that possessed this function and all of its descendants. Relative to being a pentadactyl forelimb, the hand is part of a larger homology that ranges across the tetrapods. Relative to being simply the end of a vertebra in Owen's sense, it is a member of an even more inclusive homology. If something like this is the case, then we might say that the grasping function of the hand is explainable in terms of a homology.

---

(7) I am here following PANCHEN, *Classification, Evolution and the Nature of Biology*, 72- 73.

Let us now define a *homology-class* as follows.

> A *homology-class* is a class *C* such that:
> a) *C* contains an original item, *A*,
> b) if *x* is a member of *C* then all of its evolutionary transformations are members of *C*, and
> c) nothing else is a member of *C*.

I have argued, in effect, that {*M*} might well be a homology-class. In other words, it might well be the case that there is an original structure *H*, which is a hand (i.e. a pentadactyl forelimb capable of grasping), such that all and only its descendants are hands. But clearly {*M*} *need not* be such a class. It is possible that grasping evolved more than once within pentadactyl forelimbs. And it is also possible an evolutionary descendant of a hand come to have lost the function of grasping. In other words, the extension of a mixed definition such as *M* might not constitute a homology class for two different reasons:

a) because there might be no common ancestor from which all pentadactyl forelimbs capable of grasping all descended, and
b) because, even if all such forelimbs *are* descended from a single ancestor, there might be descendants of that ancestor which are not capable of grasping.

There is thus no guarantee that the similarity to which the mixed definition *M* appeals is explainable homologically.

---

(8) The phrase is Karen Neander's: she was extremely helpful in getting me to see more clearly the logical and structural differences between a conception of homology tied solely to archetypal position and one based on descent.

4. *M* makes reference to the function of hands, but it is not a *pure* functional specification; it defines *hand* as a sub-class of a more inclusive homology-class, that of the tetrapod pentadactyl forelimb. What is the significance of this? Contrast *M* with the following *purely* functional, or *homoplastic*, definition which does not depend at any point on homologies:

H An *f-hand* is an organ capable of grasping.

An arthropod claw or reptile tail that is so capable would qualify as an f-hand, and thus the definition admits of instances that do not conform to the same body plan. The mixed definition *M* above does not do this. But it is surely not true that all f-hands are *hands*. The tail (or entire body) of a python might be capable of grasping its prey; nevertheless, it is not a hand. This is not merely a terminological point. It is not a matter just of what we call a hand. The more important consideration is that if were to use a homoplastic definition, then we would be specifying an organ by means of characteristics that are derivative from an explanatory point of view.

Does a hand have the morphological characteristics it does because it has a certain function, or does it have that function because it has those morphological characteristics? Proponents of evolutionary function like Ruth Millikan tend to emphasize the former sort of dependency. Natural selection has so shaped the morphology of our hands that they are now fit for grasping. This is because it is an advantage for us to have forelimbs that are capable of this. This argument seems to imply that the morphology follows function. But clearly this is not the case. At the very least, we have to distinguish between morphological characteristics that precede function and those that are subsequent to it. For though it may be true that certain fine morphological characteristics of our hands make them fit for grasping, and are there because they do, it is not the case that the pentadactyl structure of the hand evolved for this reason. Many such limbs are adapted to other functions. This structure pre-existed the grasping function of the hand, and only later became adapted to this function. The latter process of adaptation takes the broad structural characteristics of the pentadactyl limb as given and uses function to explain why it was modified in a given way.

So the situation seems to be this:

A. The pentadactyl forelimb arose for reasons unconnected with the grasping function.

B. The original gross structure of this organ, *G*, was such that it could be modified in order to allow it to grasp.

C. Once modified, the organ came to acquire fine morphological structure, *F*, which specifically adapts *G* to the grasping function.

The gross morphological structure *G* is not explained by the grasping function. However, the fine morphological detail *F* is so explained. It might seem that the advantage of the mixed definition is that it permits us to appeal to the pre-existent characteristics of the pentadactyl forelimb, *G*, in order to explain why it is that the grasping function dictates a particular form *F*. Thus the mixed definition is able to explain *F* as an adaptation of *G* to the function of grasping. The homoplastic definition is unable to do this because it is concerned with grasping as such, not with pentadactyl-forelimb-grasping. The morphological characteristics that would result from the tail vertebrae becoming adapted to grasping (as in the case of the python) are very different from those that result from adapting the pentadactyl forelimb to this purpose. This is where the attempted homoplastic definition of the hand falls down. However, the mixed definition will do better than the homoplastic definition only if the grasping function in pentadactyl forelimbs can be homologically explained. If this is not the case, i.e. if the grasping function evolved more than once, then there will be two adaptations of *G* to the grasping function, two different morphological fine structures that are adapted to grasping.(9)

Ron Amundson and George Lauder have argued recently that a certain kind of functionalism fails because it relies on homoplastic definitions. The "selected effect functionalism" of Ruth Millikan and Karen Neander fails, they say, in cases where morphological criteria are used to identify organs.(10)

---

(9) The points being made above bear some similarity to points I heard made by Larry Shapiro in a paper delivered to the conference on "Intentionality, Naturalism and Evolution" held at the University of Western Ontario in April 1998.

(10) "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology," *Biology and Philosophy* **9** (1993-94): 443-69. (This article is reprinted in David L. Hull and Michael Ruse (eds), *The Philosophy of Biology* [Oxford: Oxford Readings in Philosophy, Oxford University Press, 1998]: 227-57. My page references are to the latter.)

> [I]sn't 'kidney' a function category? Well, kidneys do all perform common functions (in vertebrates). But they are also homologous. This means that we could identify all members of the category 'kidney' by morphological criteria alone (morphological connectedness and developmental origin). So, at least in that sense, 'kidney' is not a function category, or at least not *essentially and necessarily* a function category. (242)

This is an odd argument. "Anatomists use morphological criteria to identify kidneys. This specification succeeds. Therefore, *kidney* is not essentially a function criterion." Can't one simply turn this point on its head? Amundson and Lauder concede that "kidneys do all perform common functions (in vertebrates)." So the vertebrate kidney can be identified by means of a functional criterion. Does it follow that *kidney* is not essentially and necessarily a morphological category? There seems to be nothing here that "selected effect functionalists" need to fear.

True, but a different and more telling point against functionalism can be extracted from Amundson and Lauder's critique. They claim that morphological identification succeeds only when a specification of organ captures a homology class:

> The histological structure of mammalian cardiac muscle could not be mistaken for any other tissue. Thus, it is incorrect to suggest that the hearts that characterize natural evolutionary clades cannot be characterized by anatomical criteria. *This situation will obtain just when all of the members of the functional category are homologous within the taxon.* Since all *vertebrate* hearts are homologous they can be identified by anatomical criteria notwithstanding the name they share with their molluscan analogues. (ibid, my emphasis.)

The extra-functional criteria — i.e. morphological criteria — that serve to define the class of vertebrate kidneys and the class of hearts succeed, Amundson and Lauder maintain, precisely because kidneys are homologous. In other words, function predicts morphology only when "all of the members of the functional category are homologous."

This claim has a startling consequence when yoked to the dictum, adopted by many philosophers, that explanatorily derivative characteristics should not be used as definitions. Suppose that F is the function of an organ X. The dictum demands that X should be defined in terms of F only if F is predictive with respect to the morphological and other properties of X. If this is the case, the morphological properties of X will be strongly constrained by its function. This implies that a morphological characterization of X will be possible on the basis of its function. But Amundson and Lauder have just informed us that morphological identifications are possible only when "all of the members of the functional category are homologous." So, functional characterizations are explanatory of morphological characters only when any two items so characterized are homologous, i.e. only when the functional characterization specifies a sub-class of a homology-class.

Of course, it is hard to know how much credence one should give to the very strong claim that Amundson and Lauder make, namely that morphology is predicted by function only when the functional similarity is explained homologically, i.e. where all the items that possess the function in question can be traced to a single ancestor that also possessed that function. After all it may well be that even if hands evolved twice, all hands might share the fine morphological characteristic F referred to above. If the gross morphological characteristic G of the pentadactyl forelimb permits only very few ways of being adapted to the grasping function, it might seem that function might well predict morphology. But then, if that is so, G, the ancestral morphology, is really doing a lot of the work. We need not go further with this line of thought: the situation is clearly quite imponderable in the abstract. Still the point is clear: a functional characterization of an organ is not going to be very predictive with respect to morphology unless it is closely tied to a homology class.

Amundson and Lauder argue that because "selected effect functionalists" like Ruth Millikan and Karen Neander ignore homologies, they fall afoul of practices well established in the evolutionary theory of classification. This point has been taken up, with independent argumentation, by Paul E. GRIFFITHS in *What Emotions Really Are* (Chicago: University of Chicago Press, 1997) and by myself in MATTHEN, "Biological Universals and the Nature of Fear," *Journal of Philosophy* **95** (1998): 105-32. It is a rather strange feature of Amundson and Lauder's polemic that they attempt to vindicate a form of functionalism which they attribute to Robert Cummins, "causal role" functionalism. But causal role functionalism is certainly no better off in terms of respecting homological boundaries than selected effect functionalism; indeed one would think it was worse off, since it is utterly ahistorical in its tendency. Again, I am grateful to Neander — one of the selected effect functionalists targetted by Amundson and Lauder — for discussion of this point, and for making me recognize the possibility of what in the text I call imixed definitions." She plans her own response to Amundson and Lauder.

5. The points made in the last section transfer to functionalism in the philosophy of mind. It is supposed to be an attractive feature of functionalism that it allows multi-realizability. Consider one classic articulation of this doctrine. David Lewis suggests that mental kinds should be defined by the causal role they play in humans.(11) Pain, for example, might be defined by its occasions and by its functions, i.e. by the situations which make us feel pain, and the ecological significance of the behaviour that we tend to perform while in pain. This is a homoplastic definition. As we saw in section 4, it may fail for two reasons to define a homology class. The first is that it might include things that are not homologous with human pain with respect to the characteristics it invokes to characterize pain. The second is that the evolutionary descendants of a state that has the defining characteristic of pain might fail to have those characteristics.

Lewis's definition falls prey to both these dangers. He asks what we should do in the case of a "Martian" who is subject to a state functionally very like human pain, but phenomenologically very different from it. The identification of this state as pain goes against certain deep seated intuitions, because we tend to think that pain is non-accidentally linked to the experiential character of the pain that we feel. Lewis attempts to solve this difficulty by means of the idea that the same causal role property can be instantiated by different states in different "populations." The state that instantiates pain in us *might* be different from the one that instantiates pain in a Martian, Lewis suggests. This ought to remind us of the point that was made in the last section: the grasping organ in pythons is morphologically different from the grasping organ in humans. Lewis is proposing that pain be defined in such a way as to include instances that are not only not homologous with one another, but are phylogenetically unrelated at every level — Martian life must have originated independently of life on Earth, one imagines. The conclusions of the last section throw some doubt on the propriety of such a definition. We can expect Martian pain to be utterly different from a morphological point of view from human pain. Thus, the functional definition will have no predictive value with respect to the morphological characteristics of pain *in general*. What then is the point of defining pain so broadly?

---

(11) Lewis, "Mad Pain and Martian Pain," in his *Philosophical Papers*, volume 1 (New York: Oxford, 1983): 122-32.

A further difficulty is that homoplastic definitions will exclude the evolutionary descendants of pain which have lost their function. To prepare us for the appreciation of this point, imagine the following case. Suppose that a neurological patient loses some vital part of his pain circuit with the result that he feels pain in his right arm in a totally unfunctional way — he feels pain when there is no proper pain occasion, and fails to feel pain when there is. A neurosurgeon decides to rectify this situation. She rigs up a prosthetic device by which the pain circuit is rerouted: when there is a proper pain occasion in the right arm, the patient now experiences a feeling of warmth in his right index finger. (Unfortunately, he continues to experience pain in all sorts of random situations.) The patient is then trained to react to such feelings of warmth in his right index finger in much the same way as he used to react when he felt pain in his right arm, and also to disregard *pain* in the right arm. Should we then say that for him, as for the "Martian," the feeling of warmth in his right index finger *is* pain? Surely not. We would simply say that some other sensation is functioning in this patient in just the way that pain in the right arm functions in our case. And similarly we should want to say that the non-functioning pain sensations in the right arm are still pain, even though they have lost their function.

Now, Lewis responds to this kind of case as follows. He says that the identification of pain is tied to "populations," not individuals. The state which fills the causal role in the human population at large is the one which we identify as pain. Since the patient is a member of this population, it is not the warm feeling in the right index finger that counts as pain, but the state that counts as pain in the population as a whole. And this is a very reasonable response. Because the patient's sensations of pain in his right arm are phylogenetically and morphologically connected with pain as it exists in the human "population" it is too much like the latter to be excluded — it is still confoundedly unpleasant, and the patient continues to have all sorts of unconditioned responses to the sensation.

But change the example a little bit. Suppose that a change similar to the one that we have just described occurs during the course of evolution. Then we will have a case in which a descendant population comes to possess a state that is homologous with pain but fails to function in the way that human pain does, and another, non-homologous, state that comes to fill that role. Lewis is now committed to saying that the state that functions like pain — i.e. the feeling of warmth — *is* pain. We will assume that the functional associations of pain are now attached to warmth in the right index finger, that is the sensation that is unpleasant and brings about avoidance behaviour, it is the state that has to be used to bring about conditioned avoidance, and so on. Even so, the morphological characteristics of both the warmth sensation and the non-functional pain sensation will be traceable more to phylogeny than to function, and will be more readily explained by ancestry. The functional definition falls prey to the difficulty that it is no longer predictive with respect to the morphology of pain. Even though the new state is now endowed with the unconditioned responses of pain not the old one — it is now warmth in the right index finger that is confoundedly unpleasant, not the random occurrences of pain in the right arm — we really ought to say, as we said in the case of the neurological patient, that a new state has taken the place of pain in the right arm, and that pain is now non-functional. Only such a psychological typology will explain the "accidental" characteristics of the non-functioning state as it occurs in the right arm: that it is a "firing of C-fibres," that it is subjectively similar to pain in the left arm (even though not as unpleasant), etc.

6. Philosophers seem in fact to be aware of such difficulties in "functional", i.e. homoplastic definitions, and tend, without necessarily realizing it, to offer mixed rather than homoplastic definitions when they think that they are in a functionalist mode. Consider the following scenario.

> We acquire a mental habit that takes the place of fear. That is, we coolly and rationally identify certain situations as dangerous in greater or lesser degree, and try to avoid them to the extent that they threaten our safety. (Think of Spock in *Startrek.*) Is this mental habit the *same* as fear? The mental habit is not accompanied by the physiological symptoms or phenomenological experiences that are the hallmarks of fear. It is a "cool" and calculated response that is quite unlike the experience of fear, and plausibly the way in which it causes avoidance behaviour is very different.

Scenarios like this one lead many philosophers to build it into their definitions of fear that it must be a state of "involuntary arousal" that

has certain effects.(12) Since there is, in humans, only one kind of state of involuntary arousal that helps us identify certain situations as dangerous and to avoid them, this seems to get us out of the difficulty.

That this is the motivation for introducing the stipulation that fear must be a state of involuntary arousal illustrates the point that it is not just the functional role that matters in our explanations of mental events, but the identity of the item that plays that role — the added condition plays much the same role as the specification that hands must be pentadactyl forelimbs. The additional condition has some bite because the identity of the item that plays the role is taken to depend on accompaniments that are not essential to the role — i.e. on characteristics the absence of which do not effect the successful performance of the role. In general it is felt that the accompaniments that matter go to the physiological and phenomenological character of the states in question.

This complements the conclusion we have arrived at in connection with the definition of *hand*. We argued that if we were interested in the explanatory priority of the structures that are adapted to the functional roles as delineated above, we will need to introduce the kinds of broad morphological considerations that define a homology-class within which the requisite function occurs. Just as hands are defined by a certain function within a broader homology class, so also pain and fear must be. Though there might be things outside that homology-class that share the function, those things will not generally share in the very characteristics that allow these things to perform their functions.

7. Mixed definitions allow us to restrict the extension of a functional definition to a sub-class of a homology-class, and thus they gain the ability to predict the morphological characteristics of the thing they define. This is a virtue. But mixed definitions are subject to an epistemological and methodological limitation. They are not, as we have seen, *guaranteed* to capture a homology sub-class. And this means that the functional component of mixed definitions might well be mistaken: that is, they may simply cross phylogenetic boundaries without realizing it, and hence be worthless. To avoid this problem, mixed

(12) See, for example, Wayne DAVIS, "A Causal Theory of Experiential Fear," *Canadian Journal of Philosophy* **18** 1988: 459-84.

definitions should, I believe, be backed up by a kind of definition that is not subject to this difficulty. (13) I shall illustrate the point by means of a case where there is disagreement concerning the true essence of a mental entity. My contention is that mixed definitions are inevitably partisan in the midst of such disputes. Consequently, further investigation of the dispute requires another form of definition, an ostensive definition that appeals to the notion of homology.

Consider the following disagreement concerning the essential nature of fear. Jenefer Robinson(14) believes that emotions are bodily states that render some feature of the environment salient and mediate an appropriate behavioural or communicative response to that situation. Consequently, she includes states like hunger, pain, and sexual deprivation among the emotions. These are bodily states that mediate a behavioural response to an environmental situation. Now consider the startle response. Robinson believes that it should be classified along with fear because it mediates the same sorts of response to roughly the same kinds of environmental situations as fear does. Robinson's classifications are tied to a particular conceptualization of what the emotions are. How would her classifications fare in a different sort of conceptualization of emotion? Suppose that one thought of the emotions as precisely those states that *in humans* are triggered by beliefs. Then one would have to exclude hunger, pain, and sexual deprivation (as distinct from lust or love) from the emotions, on the grounds that they are more closely tied to hormonal than to cognitive states. One is not typically hungry because one *believes* that a certain situation obtains, i.e. that one ought to be eating. Nor is one in pain because one *believes* that one has undergone some sort of physical trauma. On the contrary, the occurrence of hunger and pain are reasons for arriving at such beliefs; hunger is a hormonal state that indicates low nourishment, pain a neurophysiological state that indicates trauma. The startle reflex falls into the latter category. While fear is (as Wayne Davis says, see note 11 above) normally the result of "the occurrent propositional belief" that harm is imminent, the startle reflex is much more automatic and is not mediated by any such belief.

---

(13) The discussion that follows complements and to some extent parallels that in section III of my article, "Biological Universals and the Nature of Fear," cited in note 10 above.

(14) *Robinson*, "Startle," *Journal of Philosophy* **92** (1995): 53-74.

It should not be linguistic "intuition" that decides cases like this. For me the key question is this: is the startle reflex homologous with fear? That is to say: is the evolutionary ancestor of fear something close to the startle reflex, or is the latter something which originated independently and continues to be independent. One can formulate this issue in terms of an evolutionary version of Owen's conception of homotype, or *iterated homology*, as we have called it.

> *X* and *Z* are *I*-homologous (in the evolutionary sense) if *X* and *Z* are repetitions in a particular type of animal of the same item in an ancestor (i.e. if *Y* is a distinguished item in the ancestor, and *X* and *Z* are both *G*-homologues (in the evolutionary sense) of *Y*.

What we want to know is this: is fear *I*-homologous (in the evolutionary sense) with the startle reflex as these states occur in humans. Is fear a different version of the same ancestral thing? Is it morphologically similar in ways that would make a biological systematist think that it was the same thing? And I should want to ask the same sorts of questions concerning hunger, pain, and sexual deprivation. Are each of these things, as they occur in humans, *I*-homologous with human emotional states. For Robinson the question is not settled like this; it is a matter of an inductive definition. Having surveyed the emotions as she conceives of them, she arrives at a functional definition that includes the startle reflex. In my view, this methodology is mistaken, regardless of whether she is right or not concerning the inclusion of the startle reflex in the same category as fear. The basis for inclusion should not be functional, because as we have seen, sameness of function is compatible with difference in kind. The basis for including two biological items in the same class ought to be homology. And the evidence that two biological items belong together ought to be the kind of evidence that puts them in the same homology-class. Since functional specifications are not, in general, evidence for homologies, it is not appropriate to define biological kinds functionally.

Earlier we saw that though the mixed definition of *hand* might well capture a homology class, it is possible for it not to do so. There could be, in the future, evolutionary transformations of contemporary hands that are not adapted to the grasping function, but since they are members of the same homology-class, they would be instances of the same kind of thing, adapted to a different purpose. Further, one could be wrong. There could be other pentadactyl forelimbs adapted to grasping, but this point of similarity with our hands might not be explainable by common origin. These points cross over to the mental kinds. Things with one function in us might well evolve in such a way that they have a different function in our evolutionary descendants. And we might be wrong about what constitutes their defining functions. These problems should not be allowed to get in the way of successful definition. Mental kinds can, like anatomical kinds, be defined by reference only to their origins.

University of British Columbia.