Cognitive Modules, Synaesthesia and the Constitution of Psychological Natural Kinds

Richard Gray

For Final Version See *Philosophical Psychology*, 14, no. 1 (2001) 65-82.

ABSTRACT *Fodor claims that cognitive modules can be thought of as constituting a psychological natural kind in virtue of their possession of most or all of nine specified properties. The challenge to this considered here comes from synaesthesia. Synaesthesia is a type of cross-modal association: input to one sensory modality reliably generates an additional sensory output that is usually generated by the input to a distinct sensory modality. The most common form of synaesthesia manifests Fodor's nine specified properties of modularity, and hence, according to Segal, it should be understood as involving an extra module. Many psychologists believe that synaesthesia involves a breakdown in modularity. After outlining how both theories can explain the manifestation of the nine alleged properties of modularity in synaesthesia, I discuss the two concepts of function which initially motivate the respective theories. I argue that only a teleological concept of function is properly able to adjudicate between the two theories. The upshot is a further application of so-called externalist considerations to mental phenomena.*

1. Introduction

Exactly how we should think of natural kinds in science is contentious. Jerry Fodor remarks that a natural kind might be initially thought of as: 'a class of phenomena that have many scientifically interesting properties in common over and above whatever properties define the class' (Fodor, 1983: 46). He suggests that all cognitive modules are members of a psychological natural kind in this sense since they have interesting properties in common over and above the functional similarity of the input systems which originally define the class. If Fodor's suggestion is right then cognitive modules might serve to distinguish psychology from other levels of enquiry.

Fodor claims that the mind is not fully explicable in terms of cognitive modules; central systems would be required to cut across and thus relate the outputs of modular systems. He claims that cognition possesses a (more or less) tripartite functional arrangement: transducers (sensory organs) are analogue systems which turn proximal signals into co-varying neural signals, input systems are computational systems which perform complex inferential transformations on the inputs they receive from transducers, and central systems operate on the representations of distal properties computed by the input systems (Fodor, 1983: 38-46). More recently, evolutionary psychologists have argued that central processes are also subserved by modules and therefore, that the mind should be considered massively modular (e.g. Tooby & Cosmides, 1995, and for a challenge to this view Samuels, 1998).

Input systems are, according to Fodor, cognitive modules, and, in turn, members of a psychological natural kind, in virtue of their possession of most, or all, of nine specified properties:

(i) Input systems are domain specific.
(ii) The operation of input systems is mandatory.
(iii) There is only limited central access to the mental representations that input systems compute.
(iv) Input systems are fast.
(v) Input systems are informationally encapsulated.
(vi) Input systems have shallow outputs.
(vii) Input systems are associated with fixed neural architecture.
(viii) Input systems exhibit characteristic and specific breakdown patterns.
(ix) The ontogeny of input systems exhibits a characteristic pace and sequencing.

Fodor notes that these need not be considered individually as necessary conditions of modularity, but it is essential that a number of them be realised for the attribution of modularity and (i) - (ix) are anyway such that if several of the properties are realized then most of them are likely to be realized. What is significant for present purposes is that, taken together, they offer a jointly sufficient condition for the attribution of modularity; whatever possesses most or all of these properties should be regarded as a cognitive module (Fodor, 1983: 47 is explicit on this point: 'if there are other psychological systems which possess most or all of these properties then, of course, they are modular too'.)

The challenge to Fodor's view of modularity considered here comes from synaesthesia. Synaesthesia is a distinct type of cross-modal association: input to one sensory modality of one type of physical stimulus reliably generates an additional phenomenal character of experience that is usually generated by the input to a distinct sensory modality of another type of physical stimulus. For instance, a particular auditory stimulus, such as a linguistic utterance, will produce an experience of a particular colour as well as an auditory experience. Whilst synaesthetic-like manifestations may be acquired in the presence of specific brain lesions, synaesthetic manifestations can also be the result of endogenous factors. It is this latter form of synaesthesia which will be of concern here.[1] The challenge arises because one well-studied form of synaesthesia manifests most, if not all, of Fodor's nine specified properties of modularity. Hence synaesthesia, according to Fodor's understanding, should be considered as involving the emergence of a new type of module. I shall call this proposal the *Extra Module* thesis (the EM thesis for short). The EM thesis has been outlined by Gabriel Segal (Segal, 1997).[2] However, it is claimed by some psychologists that synaesthesia involves a breakdown of modularity. Baron-Cohen *et al.*, for instance, have suggested that the above-mentioned form of synaesthesia indicates a breakdown of barriers between the speech and colour processing modules (Baron-Cohen *et al.*, 1993). I will call this proposal the *Modularity Breakdown* thesis (the MB thesis for short).[3]

The modularity hypothesis is by no means undermined by this disagreement, indeed, the fact that both accounts presuppose the theoretical usefulness of postulating cognitive modules endorses the hypothesis. Nevertheless, this theoretical difference presents a clear challenge to the notion of modularity as characterized by Fodor: either it is correct as it stands and synaesthesia should be regarded as involving an extra cognitive module, which is in turn a member of a psychological natural kind, or synaesthesia should not be so regarded and thus the properties Fodor postulates for the individuation of cognitive modules are insufficient for individuating modularity. Many philosophers take it for granted that a further story has to be told about the provenance of cognitive modules; one aim here is to focus that issue.[4]

The paper is arranged as follows. The next section outlines the plausibility of the EM thesis by outlining the way the nine properties of modularity seem to be manifested by colour-graphemic synaesthesia. At the same time the discussion shows how these features can be explained in a different way by the MB thesis. Section 3 examines how the EM thesis originally arises from a computational, non-teleological view of function. Although this view of function is consistent with Fodor's view of function, it is arguably in some tension with his attitude to natural kinds. Instances of natural kinds, according to Fodor, are those instances whose terms are the bound variables in proper laws (Fodor, 1974). Section 4 discusses the MB thesis as a consequence of a teleological view of function, which would view cognitive modules in a lawlike context. The claim to be endorsed here is that what ultimately distinguishes cognitive modules is that they can figure positively in equations of evolutionary fitness.

The argument is another application of so-called externalist considerations, in this case to the constitution of a psychological natural kind. If synaesthesia is a breakdown in modularity then one needs to introduce externalist considerations to explain why. If externalist considerations are needed to support the MB thesis then these should apply to the individuation of all cognitive modules. To decide between the EM thesis and the MB thesis a tenth property needs to be added to the list of properties of modularity: input systems are teleofunctional kinds.

2. Two Theories of Synaesthesia

If the traditional division of the sensory modalities is correct then there could be a wide variety of types of synaesthetic connections: experiencing colours when perceiving sounds, experiencing tastes when perceiving colours, experiencing sounds when perceiving smells. Cases of several different types of synaesthesia have already been described in the literature (Baron-Cohen & Harrison, 1997). Coloured hearing is the most commonly occurring. What Baron-Cohen *et al.* have termed chromatic-graphemic (hereafter CG) synaesthesia appears to be one of the most common forms of coloured hearing synaesthesia. In coloured hearing synaesthesia experiences of colours and Euclidean shapes are triggered by experiences of sounds in general. In CG synaesthesia experiences of colours are triggered by the sounds of words via their spellings.[5] From their study of CG synaesthesia Baron-Cohen *et al.* concluded that a breakdown of barriers between modules might be responsible for this form of synaesthesia. More recently, Segal has claimed that their results might actually indicate the emergence of an extra module and thus further confirm Fodor's account of modularity.[6]

The immediate task is to examine CG synaesthesia in the context of Fodor's nine specified properties. This will demonstrate the plausibility of the EM thesis. At the same time it will show how the MB thesis can also explain the manifestation of these nine features in synaesthesia. As should quickly become clear, the constitution of modules via their nine defining properties underdetermines the nature of modularity. The exercise is of additional interest in that it sheds some further light on the features of the nine alleged properties of modularity.[7]

(i) *Domain Specificity*. Modules could have been domain-specific simply in virtue of the determinate causal relations holding between each of them, sensory organs and specific properties of objects. If for instance the mechanism for the analysis of shape is perceptually related only to specific distal stimuli then 'it follows trivially that their computational domain qua mechanisms of visual perception is specific to the class of possible retinal outputs' (Fodor, 1983: 48). This is not how Fodor conceives of domain-specificity. Nothing interesting about cognitive processing would follow from such specificity for it is consistent with domain-specific processing so defined that each mechanism uses the same types of computations. According to Fodor, modules are specialized in virtue of the specific computational processes they each utilize to generate the range of representations they produce. The specific computational mechanisms that are used to represent colours for instance cannot be employed to represent shapes. Evidence for domain-specificity comes largely from experiments on language processing in which only a specific class of stimuli are 'capable of throwing the switch' for the perceptual systems that effect the phonetic analysis of speech. The restriction of actual linguistic systems to only a small subclass of the logically possible linguistic systems is according to Fodor evidence of a speech input system which generates representations of distal linguistic utterances by means of idiosyncratic computational processes upon proximal acoustic signals.

Can the EM thesis accept Fodor's notion of domain-specificity as stated? Segal admits that it is not clear to what extent the module which is supposed to realise CG synaesthesia is domain-specific. In CG synaesthesia there does appear to be a specific processing of information in that representations of sounds are mapped onto representations of letters which, in turn, are mapped onto representations of colour. For instance when a subject hears the word 'phonology' the underlying mechanism always produces the particular colour experience associated with the letter 'p'. It is generally believed that there is a computational system that maps the sounds of words onto their spellings which are stored in a mental lexicon. The EM thesis claims that there is an additional capacity to process representations of one type - in this case colours - from representations of another type - in this case written letters - via a domain-specific mechanism. But it is not clear that we have here evidence for the existence of idiosyncratic computational mechanisms, as Fodor conceives them. For one thing non-linguistic sounds also produce synaesthetic representations of colour. Part of the difficulty is that it is not clear what computational task the alleged module would be undertaking. This problem is particularly sharp in that different subjects have different associations.

The MB thesis claims that, in the case of CG synaesthesia, there is a breakdown of barriers between the modules which typically process speech inputs and inputs from colour sources respectively. This seems to raise a dilemma for the MB thesis: either there are no domain-specific computations or there is no barrier to break down between speech processing and colour processing modules. How, if computational domain-specificity is partially constitutive of the barrier between modules, is it possible for the information which has been processed by the speech-processing module to be further processed by the visual processing module? Although breakdowns within modules can, and should be acknowledged, breakdowns between modules should be impossible because one module, by hypothesis, would provide a representation to which any other processing module should not be able to respond, except at the proper interfaces. In the present case, the speech-processing module would provide a representation to which the visual processing module should, by hypothesis, be unable to respond. Perhaps an answer is easily available: the speech processing module already seems to generate visual representations in the form of graphemic representations (remember that cognitive modules are not to be identified with perceptual modalities), so there will be no special difficulty generated by the domain-specific mechanisms. But, if this is the case, then the dilemma ensues: there seems to be no clear barrier between modules to break down.

One response might be that the speech-processing module may normally produce graphemic representations of speech which typically fail to receive the attention of the central processing system. They do receive the attention of the central processing system only when there is a breakdown between the speech processing module and the colour-processing module. In this case the breakdown of barriers would involve not domain-specificity but other properties of modularity. This may be a satisfactory response for CG synaesthesia, but there seem to be other forms of coloured-hearing synaesthesia, such as colour-phonemic synaesthesia, which do not require the mediation of graphemic representations. This form of synaesthesia (which would coincide more closely with the breakdown of traditional notions of sensory modalities) suggests either that there are no computational domain-specific mechanisms preventing cross-modal associations or that there are extra domain-specific mechanisms as the EM thesis claims. In order to maintain the MB thesis, it seems a revision of Fodor's notion of domain-specificity is required. But this need not be controversial, for there seems to be no clear consensus amongst psychologists concerning the specificity of computational processes underlying particular psychological abilities (see Shapiro & Epstein, 1998: 174 for discussion). It may even turn out that synaesthesia is further evidence that the domain-specificity of modules is constituted by hardware rather than software.

(ii) *Mandatory operation.* The mandatoriness of operations captures the idea that once an input of the relevant kind, such as a token utterance or a written token of a language we understand, is received by the sensory transducers we cannot prevent ourselves from processing it in a way such that we are aware of it as a familiar linguistic token. The experience of synaesthetes suggests that synaesthesia too has this property of modularity. Salient amongst coloured hearing synaesthetes' remarks is that the colours evoked are automatic and unsuppressible.[8] Synaesthetes sometimes remark that they can prevent their synaesthetic experiences but, consistent with the mandatory operation of modules, it seems as though they do this indirectly, by attending to other stimuli, as we ordinarily do if we are trying to ignore a stimulus. But if this supports the EM thesis it also supports the MB thesis. Synaesthesia could be the result of a breakdown of barriers between modules such that the signal from one module cannot be prevented from reaching another module.

(iii) *Information is inaccessible to central processes.* Part of the reason for our ignorance about synaesthesia lies with the third of the specified properties of modularity. According to the modularity hypothesis the internal operations of input systems are inaccessible to inspection by the subject. The details of the processes which generate synaesthetic experiences are certainly unavailable to the subject. If synaesthetes are able to deduce the association between graphemes and colours, such a deduction would be by processes one thinks of as typically central processes. But the consequence of this is that neither synaesthetes themselves nor psychologists can tell directly whether they have an extra module or suffer a breakdown in modularity.

(iv) *Processes are rapid.* Evidence for the rapidity of synaesthetic processes is largely anecdotal, although some early attempts were made to measure them (Clavière, 1898). Even if the relative rapidity of synaesthesia could be measured, it is not obvious that this would provide evidence in favour of either the MB thesis or the EM thesis.

(v) *Information is encapsulated.* One feature of perceptual illusions is that, regardless of our beliefs about the world, we cannot directly change our perceptual experiences. This illustrates the property of informational encapsulation, arguably the central defining property of Fodorian input systems. It is the property of a cognitive system whereby access to it by information from elsewhere in the individual's cognitive system, especially feedback mechanisms from central processing, is denied. Synaesthetic experiences, in this respect, are like illusions which we cannot but be subject to. Consider the following observation reported to Cytowic: 'It is not an hallucination but it is hard for me to describe. As I look at a page, I see the colors there even though I see the colour of the real ink that's before me. I know it isn't there for real, but I still can't help seeing it. There is still a sensation that the colour is there.' (Cytowic, 1989: 43).

Again both the MB thesis and the EM thesis can account for the apparent encapsulation of information. The EM theorist will argue that synaesthetes are unable to modify their experiences because the module which underlies synaesthesia is encapsulated from the belief that letters of the alphabet are not intrinsically coloured. The MB theorist will argue that synaesthetes are unable to correct their experiences because the reason for them lies in a pathology of the neural system.

It is worth briefly considering here, in the context of synaesthesia, the issue of the scope of modules. Are input modules divided into levels, or do putative levels of processing form component modules of input systems? One might think of the various representations in Marr's model of visual input processing - the grey level description, primal sketch, 2.5D sketch and object description - as corresponding to processing interlevels or to the outputs of individual modules (Marr, 1982). If Fodor is right about the scope of the speech input module and the visual input module, CG synaesthesia could constitute either a breakdown between modules at an interlevel stage, more precisely, at the stage at which a representation of the graphemes of a word have just been generated, or an extra module at the interlevel stage. If this is the generation of an extra module then it would appear more plausible to postulate a number of smaller modules rather than one large Fodorian module as constitutive of input systems. Of course, if a number of small modules is the preferred option, CG synaesthesia could still be caused by the breakdown of barriers between them.

(vi) *Outputs are shallow*. We commonly draw a distinction between appearance and reality. Most things have a depth that is not immediately available to our inspection. In philosophy of science the issue is 'where to draw the line between observation and inference', in psychology the issue is 'where to draw the line between perception and cognition' (Fodor, 1983: 86). Fodor takes it that what are to be classed as appearances ('observations' in philosophy of science and 'perception' in psychology) are more than what are often taken to be the appearances of things by philosophers - the perceptible properties, such as the visible properties colour and shape. According to Fodor since the visual input system is a module which generates representations ready for use by the central cognitive system these representations have to be representations of basic categories of objects e.g. cats and dogs.

In CG synaesthesia the output representations are of shape and colour. This might be interpreted in a number of ways. Perhaps there is an extra module which maps representations from the language input system onto the visual processing module, the output of which is shallower than is suggested by Fodor. Alternatively there might be a breakdown in the barriers between modules, either at an inter-level stage, as suggested by the Fodorian model, or between the barriers of smaller modules.[9]

(vii) *Fixed neural architecture*. Fixed neural architecture would contribute to the constitution of the barriers between modules. (The extent of the constitution would depend upon the upshot of the earlier considerations of domain-specificity.) Positron Emission Tomography (PET) scans have been used to determine the particular areas of the brain active in CG synaesthetic experience. The details can be found in Paulesu *et al.*, 1995. In synaesthetes, a number of additional visual associative areas, including the posterior inferior temporal cortex and the parieto-occipital junction were activated. The synaesthetes also showed activations in the right prefrontal cortex, insula and superior temporal gyrus. The posterior inferior temporal cortex has been implicated in the integration of colour with shape and in verbal tasks which require attention to the visual features of objects to which words refer. It is believed that some of these brain areas are those which underlie the transition from viewer-centred to object-centred representations (Goodale, 1995, Wager, 1999).

The evidence therefore indicates a dedicated link between language areas and visual association areas. The equation is often made between localisation and modularity. As Fodor emphasizes, modules are to be individuated functionally rather than physiologically (Fodor, 1983: 98). Input modules may be distributed about brain tissue. We know neurons make connections across areas of the brain, therefore localization does not appear to be a necessary condition for modularity. But does it nevertheless support the notion of modularity?

Segal points out that the evidence for localization not only supports Fodor's claim that modules are realized in dedicated neural architecture but this particular evidence for localization supports the claim that there is an extra module at work. But once again the evidence is far from conclusive. The additional areas of activation are those which are believed to be already employed for the integration of colour with shape and in verbal tasks which require attention to the visual features of objects to which words refer. The inference is that the extra module would be realised in an independent area mediating the functional areas in use by both synaesthetes and controls (the language areas) and the functional areas in use by synaesthetes alone (the visual areas). But the evidence we have is silent on this. It is suggestive of modularity in so far as it is suggestive of dedicated or fixed neural architecture, but it does not tell us whether this additional dedicated neural architecture is the activation of an extra module or of a breakdown between two other modules.

How can additional neural architecture be construed as other than additional cognitive architecture? Perhaps synaesthetes develop extra neural connections as infants. There is an alternative possibility. It has been suggested that synaesthesia is a stage of development all neonates undergo. Modularity develops when neural connections are lost (Maurer, 1993). The manifestations of synaesthesia would in fact be better understood not as the breakdown of barriers between modules but as the failure to develop the barriers between modules in the first place. If this is true the wider implications which form the topic of this paper remain the same. (The MB thesis might be reconstrued as the breakdown of what builds up the barriers between modules.) So the presence of additional neural connections does not necessitate additional cognitive processes, in particular, it does not necessitate additional modular processes. The issue is precisely one of how we are to characterize psychologically extra neural connections.[10]

(viii) *Characteristic pattern of breakdown*. Modules have traditionally been inferred from characteristic patterns of cognitive breakdown. For example, Marr's theory of visual processing stages has received some support from clinical cases in which some subjects are unable to process shapes and other subjects are unable to recognise objects (see Ellis & Young, 1988 and Shallice, 1988). In considering whether CG synaesthesia is a cognitive breakdown or whether it can itself be subject to cognitive breakdown we come to the heart of the matter.

Each thesis can elicit support by means of criticisms of its rival. Consider the MB thesis from the point of view of the EM thesis. How can there be a breakdown of function involving the spread of information to other modules if the information is also processed correctly by the module which is supposed to be undergoing that breakdown? This is not like cases of neuropathology arising from injury or stroke in which there is an absence of normal output. In those cases breakdowns occur because an interlevel of the modular input systems or a component module breaks down. Current evidence suggests that cognitive function is not negatively affected by synaesthesia. On that evidence CG synaesthesia should not be regarded as a cognitive breakdown.

Now consider the EM thesis from the point of view of the MB thesis. It can firstly be argued that synaesthesia is not a breakdown of a particular module but a breakdown of those properties of modularity, which constitute the barriers between two modules. In that case the burden of proof can be returned to the EM theorist who needs to show what the characteristic pattern of breakdown would be. And it is not clear just what would count as evidence of any characteristic pattern of evidence of breakdown in the CG module, since it is not clear what the function of the CG synaesthesia mechanism is supposed to be (as Segal admits 1997: 220). One response the advocate of the EM thesis can make (unconvincing as it may be to the advocate of the MB thesis) is that it is not necessary that all the properties of modularity be instantiated. But actually a functional module underlying synaesthesia probably can be construed, as the next section shows, and thus an analysis of its elements indicated.

(ix) *Fixed Pattern of Development*. Few would want to suggest that acquired synaesthesia (the occurrence of synaesthetic-like symptoms following neural damage or drug usage) indicates an extra module. Nevertheless, one cannot automatically infer from this that all synaesthetic symptoms are the result of a breakdown of modularity. There are a number of important differences between acquired synaesthesia and endogenous forms of synaesthesia, in particular the occurrence of common patterns of development in the latter. Modules, according to Fodor, are to be characterised by their fixed pattern of development. Fixed patterns of development are controlled at the genetic level. If synaesthesia is characterised by genetic differences then it seems plausible that it too should have a distinctive fixed pattern of development. Evidence indicates that synaesthesia runs in families and thus does have such a genetic component (Baron-Cohen *et al.*, 1995).

Again, that synaesthesia has a genetic component does not decide between the EM thesis and the MB thesis. If synaesthesia arises in adults because neural connections are preserved this could be accounted for by the genetic differences between synaesthetes and non-synaesthetes. That synaesthesia has a genetic basis and thus a fixed pattern of development does not confirm the existence of an extra module. Even if the gene or genes for the CG mechanism are discovered, the proteins these genes code for are determined and the development processes they mediate are clarified, we may be no closer to determining whether the EM thesis or the MB thesis is correct.

As should be clear by now the presence of an alternative shows that each thesis is underdetermined by the evidence of synaesthesia. Even if the nine properties which are distinctive of modularity are necessary for the individuation of cognitive modules they do not appear to be sufficient. Something further is required to adjudicate between the two rival theses.

3. Function and the Extra Module Thesis

The concept of function operative in cognitive psychology is sometimes left unspecified. In some cases - such as damage to the brain through stroke or infection - there is little controversy over the attribution of dysfunction. The nature of function and dysfunction becomes more significant when considering a case such as CG synaesthesia, where the question of neuropathology is more controversial.

The EM thesis is a consequence of a computational, non-teleological concept of function and functional organization. According to Fodor the functions of psychological systems can be understood by comparison with the organization of idealised computing machines (Fodor, 1983: 38-46). Idealized computing machines are closed symbol-manipulating devices. Their functional architecture amounts to a small number of interacting subsystems (tape, scanner, printer and executive) and a small number of primitive machine operations (stop, start, move the tape, read the tape, change state and print). The system functions in the way it does because of the physical dispositions of its components. If the central cognitive system is no more than such a symbol manipulator then it can be fully explained in local causal terms. In order to act as a relevant model of human cognitive processes such a computational machine has to be embedded within input systems which can allow the exchange of information between the machine and its environment. These input systems would model the modules of present interest. The way that input systems transform information from the environment into symbolic representations is not, in general, any different from the way the central system then operates on them; in particular, input systems are also *solipsistic* (in the sense that the internal mechanisms are all that count). Input systems thus function in a particular way because of the pattern of dispositional properties and local causal relations which constitute them. Additional processes which have the features Fodor notes should consequently be regarded as extra modules.

Segal grants that we often talk of computers as having goals, and that the having of goals is related to the functions certain computations have. But he claims that this is only a way of talking which could be fully understood in terms of local causal properties. For instance, when modular processing gives rise to visual illusions, according to Segal, the visual input system is not doing anything it should not be doing. It was not designed in such a way that it is now failing to fulfil the parameters of any design. The same can be said with respect to synaesthesia; when the CG synaesthesia module gives rise to illusions it is not doing anything it should not be doing for it was not designed in such a way that it is now failing to fulfil the parameters of its design. In short, Segal claims that cognitive systems in general and cognitive modules in particular, are what they are independently of their origins (Segal, 1997: 215).

Computational functionalism can be related to a more general approach to functional explanation. Cummins has observed that Fodor made the connection between the analytical strategy in psychological theorising and functional characterisation without however extending it to a general account of functional explanation (Cummins, 1975: ft. 20). Cummins has himself offered a general account. According to Cummins, functional ascription and functional explanation arise from an analysis of a complex system into its component parts. He argues that the function of any part should be understood without reference to ends: the function of a part of a complex system is just the causal contribution it makes to a specified activity of the system. What the specific properties are of a component part, x, of a system, s, or, alternatively speaking what x functions as, ϕ, depends upon the analytic account, A, the theorist has of the activity, ψ, of the system of which x is a component part. The idea is that if one can make sense of a system by ascribing it goals this is enough to justify the ascription, which does not mean that the ascription cannot then be fully understood in causal terms. (In this respect Cummins offers a more interpretative account of function than Segal, who claims that the function of a part of a system is just what it is disposed to do.) Cummins defines function in the following way:

> x functions as a ϕ in s (or: the function of x in s is to ϕ) relative to an analytic account A of s's capacity to ψ just in case x is capable of ϕ-ing in s and A appropriately and adequately accounts for s's capacity to ψ by, in part, appealing to the capacity of x to ϕ in s.

Consider cognitive modules. Against the background of explaining (A) the processing of representations (ψ) of the visual system (s) we can analyse the causal contribution or function of the neural system (x) as a scene surface analyser (ϕ) because the capacity of the visual system (s) to process representations (ψ) requires the capacity of the neural system (x) to represent the surfaces of objects (ϕ). So Cummins' analysis can be instantiated in the following way:

> neural system (x) functions as a scene surface analyser (ϕ) in the visual system (s) (or: the function of the neural system (x) in the visual system (s) is to analyse scene surfaces (ϕ)) relative to an analytic account (A) of the visual system's (s's) capacity to process representations (ψ) just in case the neural system (x) is capable of analysing scene surfaces (ϕ-ing) in the visual system (s) and A appropriately and adequately accounts for the visual system's (s's) capacity to process representations (ψ) by, in part, appealing to the capacity of the neural system (x) to analyse scene surfaces (ϕ) in the visual system (s).

It is clear why damage to the brain through stroke or infection would be dysfunctional on this account: if we assume that intact cognitive performance presupposes a system having a number of components all of which are required to operate in a well-defined way for the system as a whole to operate in a specified way then neuropathologies arise when components do not function appropriately within the system so specified. Of course, viewing neuropathologies as such depends upon viewing the systems to which they are related as cognitive systems which perform a determinate set of cognitive tasks.

With respect to CG synaesthesia, all that is required for an account of function is that one be able to characterize a system whereby inputs to the speech processing module reliably cause representations of colour. Synaesthetic experiences, in which the sound of particular words are experienced as having distinctive colours, can then be assumed to be subserved by modules which compute the functions which are determined by the dispositional properties of their components. Against the background of explaining (A) the processing of representations (ψ) of the synaesthetic system (s) we can analyse the causal contribution or function of the neural system (x) as a colour analyser (ϕ) because the capacity of the synaesthetic system (s) to process representations (ψ) requires the capacity of the neural system (x) to represent its inputs as coloured (ϕ). So Cummins' analysis can be instantiated for synaesthesia in the following way:

> neural system (x) functions as a colour analyser (ϕ) in the synaesthetic system (s) (or: the function of the neural system (x) in the synaesthetic system (s) is to analyse the inputs as colours (ϕ)) relative to an analytic account (A) of the synaesthetic system's (s's) capacity to process representations (ψ) just in case the neural system (x) is capable of analysing inputs as colours (ϕ-ing) in the synaesthetic system (s) and A appropriately and adequately accounts for the synaesthetic system's (s's) capacity to process representations (ψ) by, in part, appealing to the capacity of the neural system (x) to analyse inputs as colours (ϕ) in the synaesthetic system (s).

An initial uneasiness might be felt about this approach. Cummins proposes that the effectiveness of such an analysis is proportional to the extent to which the capacities of the *analysans* are less sophisticated than and different in type from the capacities of the *analysandum*; there is no point in sophisticated hypotheses about simple capacities. CG synaesthesia would certainly not involve as complex a system as either the visual input system or the language input system. Since there is no clear cut-off point, it is open to dispute whether CG synaesthesia would be sufficiently complex. And it would be correspondingly less clear why the relevant brain systems damaged through stroke or infection would be dysfunctional on this account.

A deeper difficulty with this view is that there is another possible explanation for synaesthesia: it is a breakdown of modularity. It is difficult to say what analytical strategy (A) the computational account could adopt here. And if there were a strategy it would be difficult to give a reason why we should adopt that one rather than the above strategy without adverting to an alternative theory of functional explanation altogether. Such a theory of functional explanation would not fit in with the current explanatory strategy because it denies that every causal property needs to have a function in a wider system. This theoretical explanation turns out to be better fitted to adjudicate between the EM and MB theses.

4. Function and the Modularity Breakdown Thesis

One reason for adopting the MB thesis is that it arguably fits better with Fodor's own suggestions about natural kinds. In an earlier paper discussing the unity of science and the relation between the laws of the basic sciences and the special sciences, Fodor suggests that we should think of natural kinds in science with respect to laws: 'the kind predicates of a science are the ones whose terms are the bound variables in its proper laws' (Fodor, 1974: 87). The upshot for Fodor is that the laws and natural kinds of the special sciences are irreducible to those of the basic sciences. To resist the view that CG synaesthesia involves an extra module a view of function which *does* reflect wider law-like processes needs to be outlined. Only this will eventually provide us with the individuation conditions which will allow us to distinguish genuine cognitive modules. It is not that computational considerations are not interesting, nor that they are not explanatorily useful, it is just that in the long run they are not sufficiently so.

This discussion can now be seen as belonging to a larger discussion about the relative merits of the computational and biological frameworks in psychology. Sober has argued with reference to psychological functionalism that 'function' is ambiguous and the doctrine was developed with the wrong meaning in mind (Sober, 1985: 165). He favours a teleofunctional as opposed to a computational view of function. Exactly how one should think of functional explanation teleologically is a matter of current debate. Common ground in the debate is the view that functional explanation should be a distinctive form of explanation because the function of an item is determined by the contribution it has made or does make to individual fitness.

Millikan argues for the claim that functions are explained by reference to evolutionary history. For Millikan, to describe the biological function of an item is not to describe its dispositional capacities, it is to describe the role that the ancestors of that item played in a historical process, including birth, development and reproduction over numerous generations. If individuals possessing a trait have been favoured by natural selection because their token traits have performed in a certain way, then that is the function of the trait. Thus function ascription involves saying what something is for by saying why it is there. Consideration of the biological context allows the selection of the relevant properties of a trait for proper function ascription among others of its properties, thus providing the opportunity for a distinction to be drawn between function and accident, and function and malfunction. If a trait has a function because it has increased the fitness of individuals then any property of the trait is accidental if it has not increased individual fitness. A property of a trait is a malfunction if it does not increase fitness and is different from a related property of the trait which has historically increased fitness (Millikan, 1984).

One issue that divides the parties to the debate is how one is to relate teleological function and causal explanation. It should be recalled that Segal, following Fodor, argues that psychological processes, as causal processes, can be studied without reference to their history. Millikan argues that teleological explanation is quite different from causal explanation. But a consideration of history alone cannot introduce teleofunctional considerations and such a stark contrast between teleological function and causal explanation need not be preserved, as aetiologists maintain, in order to maintain the distinctiveness of teleological explanation.

Walsh and Ariew have recently argued that the aetiological theory championed by Millikan and others is incomplete. Its incompleteness is particularly evident when the present utility of a trait is different from its past utility. They argue that functional explanation in biology must be analysed with respect to relevant regimes of selection. The relational theory they advocate claims that: 'the way a trait contributes to fitness may vary wildly according to the environment [...] one must specify the contribution to fitness with respect to a selective regime' (Walsh & Ariew, 1996: 498). The relational theory explains both the persistence of traits and why we should expect a trait to persist into the future. In this way Walsh and Ariew are able to present the further claim that functional explanation in biology can be viewed as a specific sub-category of functional explanation as characterized by Cummins - that category which is distinguished by the context of the contribution to average individual fitness: 'the evolutionary function of a trait token (with respect to a regime) is that Cummins function which constitutes the (positive) contribution to average fitness for tokens of the trait's type (with respect to that regime)' (Walsh & Ariew, 1996: 509). By the right reference to the external environment, teleofunctional explanation can both be viewed as a distinctive form of explanation and be compatible with causal explanation.

This would go some way to satisfying Fodor's suggestion that the kind predicates of a science are the ones whose terms are the bound variables in its laws. The biological law-like explanation for the evolution of specific cognitive modules, simply put, would be that the ancestors of cognitive modules, as a means of processing information, would have caused a differential in average fitness between the individual organisms which possessed them and those which did not, and thus an increase in the descendants of those ancestors of cognitive modules in the population. The evolutionary psychologists Tooby & Cosmides argue that: 'modules are kinds invented by natural selection during the species' evolutionary history to produce adaptive ends in the species' natural environment' (Tooby & Cosmides, 1995: xiii). It is not clear what the full evolutionary story of cognitive modules generally would be. Fodor notes that modules might have evolved from central processes (the most efficient means of processing information might have required freedom from background beliefs) or central processes might have evolved from modular processes (the most efficient means of processing information might have required freeing certain sorts of problem-solving systems from the constraints under which input analysers operated). But, as a consideration of synaesthesia shows, the difficulty of articulating the full evolutionary story is not a reason for eschewing evolutionary considerations as determinants of the properties of cognitive modules, as Fodor seems to suggest (Fodor, 1983: 43).

Whether synaesthesia is realized by a breakdown in modularity or by an extra module depends on how this alleged module would contribute to the fitness of individuals. Baron-Cohen has argued that if the MB thesis is correct there has to be a clear cost to fitness produced by CG synaesthesia. The problem for the MB thesis according to Baron-Cohen is that CG synaesthesia does not appear to be maladaptive. Baron-Cohen tries to save the MB thesis by discussing a case in which a synaesthetic subject not only experiences colours when she hears sounds, but experiences sounds when she sees colours. The dysfunctional nature of synaesthesia, according to Baron-Cohen, only becomes apparent when the condition is bi-directional (Baron-Cohen, 1996). But the MB thesis is credible without this line of reasoning for it is not the MB thesis which has to show a cost to individual fitness but the EM thesis which has to show a benefit to individual fitness.

Some evidence that synaesthesia does not increase fitness derives from its relative scarcity in the population.[11] It may be that the novelty of synaesthesia is such that selection has not yet had time to act on it. In that case, what could the utility of synaesthesia possibly be? Synaesthesia does not allow individuals to perceive more, or the same more quickly, since the experiences had are non-veridical, e.g. a synaesthete may experience a word to be coloured when it is not. It is perhaps hard to see how future selective regimes might allow synaesthesia to bestow a positive contribution to average fitness of an individual synaesthete. Still it might be the case that synaesthesia does have a function, for teleofunctions are not always easy to discern. It is an open empirical question.[12] If synaesthesia does confer an advantage then the EM thesis would thereby be vindicated. If the EM thesis is wrong then synaesthesia need not be considered malfunctional; it need only be considered accidental. Synaesthesia can be regarded as accidental rather than malfunctional because it does not involve a breakdown of modules, but a breakdown of certain properties of modularity.[13]

5. Conclusion

Since we are dealing with two concepts of function, we are dealing with two concepts of modularity. An examination of synaesthesia shows that the characterisation of modularity by computational considerations is insufficiently discriminating. It is the concept of teleofunction or proper function that allows us to pick out membership of a psychological natural kind. [14] In sum, Fodor suggests that cognitive modules are members of a psychological natural kind because they possess nine distinctive properties; the preceding argument claims that we should consider cognitive modules as possessing a tenth property which they share with other biological characteristics. Input systems should be considered as properly teleofunctional kinds.

Acknowledgements

I am grateful to Denis Walsh and Timothy Williamson for detailed comments on earlier drafts of this paper and to members of the audience at the 1999 British Society for the Philosophy of Science annual summer conference for their responses to some of the ideas discussed in the paper.

Notes

[1] Synaesthetic-like manifestations acquired through brain lesions might be the manifestation of characteristic patterns of cognitive breakdown but such manifestations would not themselves exhibit characteristic and specific breakdown patterns as is required by property (viii) above.

[2] The present paper focuses on one form of synaesthesia. Other forms of synaesthesia may constitute related challenges to Fodor's modularity hypothesis.

[3] For reasons given in the discussion of the fixed neural architecture of input systems the explanation for synaesthesia might lie not in the breakdown of certain properties of modularity but in the failure to develop certain properties of modularity. If this is so a slightly different thesis could be substituted for the MB thesis without significant implications for the overall argument. Call it the *Modularization Breakdown* thesis.

[4] Opponents of the modularity hypothesis typically argue that the nine purported properties of modularity can be explained just as well by other models of cognitive architecture such as production systems or connectionist systems. See Stillings, 1989. The present paper brackets this issue, although it might be that its line of argument can be construed as undermining support for the modularity hypothesis in so far as it emphasizes the underdetermination of the modularity hypothesis by the present data.

[5] The experimental procedure adopted by Baron-Cohen *et al.*, 1993 was designed to distinguish between a number of possible forms of coloured-hearing synaesthesia that appear to be associated with the auditory presentation of language. Chromatic-lexical synaesthesia occurs when different spoken words produce experiences of colour in virtue of the particular words they are, i.e. there seems to be no clear determination of the colour either by the sound of the word, or by the alphabetical representation of the word, or by the meaning of the word. Chromatic-phonemic synaesthesia occurs when different spoken words produce experiences of colour in virtue of their sounds. Chromatic-graphemic synaesthesia occurs when different spoken words produce experiences of colour in virtue of their spelling.

[6] There are other theories of synaesthesia. Grossenbacher, 1997, for instance, argues that synaesthesia might involve a feedforward/feedback mechanism. This might still be construed as a breakdown of modularity, albeit of a different form from the one intended by Baron-Cohen *et al.*. It is not clear that there is a theory of synaesthesia which is more plausible than the theories considered here and which rejects the modularity hypothesis

completely. So I take it that the most plausible theories of synaesthesia either advocate some kind of breakdown of modularity or the emergence of an extra module.

[7] The literature on modularity is fast expanding. Specific challenges have been made to the details of Fodor's modularity hypothesis from both philosophers and psychologists. On the issue of domain specificity see papers in Garfield, 1989, especially those by Arbib and by Stillings. On the issue of the mandatoriness of central processes see Marlsen-Wilson & Tyler, 1989. Putnam, 1984, and Churchland, 1989 have challenged the detail of the encapsulation of modules and the nature of their supposedly shallow outputs. See also Marlsen-Wilson & Tyler, 1989 (with respect to the encapsulation of language processing modules) and Arbib, 1989 (with respect to the encapsulation of visual processing modules). Marshall, 1984, has challenged the impenetrability of the internal processing of modules, whilst Karmiloff-Smith, 1994, has challenged the innateness of modules.

[8] See the questionnaire response in Baron-Cohen *et al.*, 1993: 423. 'They all recalled the surprise of discovering that this was not the case for everyone. All subjects also reported that the colours evoked were automatic and unsuppressible.'

[9] There is also the question of the constitution of the content of this output. For a discussion see Wager, 1999.

[10] It is doubtful whether more careful study of could help. Maybe further experiments, which allowed the subtraction of the functional areas in use only by synaesthetes, could address this. Perhaps specific additional neurotransmitters could be found. But even if the evidence supported additional neural architecture or chemistry of a specific type this would not be conclusive evidence for a new module.

[11] It might here be objected that although synaesthesia does not benefit the individual the gene responsible for it may be linked to genes which do confer benefit upon their possessors. Even if this were the case, synaesthesia would not be the focus of positive selection. Linkage might explain synaesthesia's frequency in the population, but it would not explain its own adaptive advantages, and therefore its function.

[12] Grossenbacher, 1997: 156 suggests that additional colour labelling might serve as a natural cognitive resource when stimulus conditions preclude colour sensations, e.g. under poor lighting conditions perceived shape might automatically evoke colour imagery in order to facilitate memory retrieval for object recognition.

[13] It may be that some forms of synaesthesia, as the breakdown of the barriers of modules, are to be considered accidental because they do not prevent proper functioning, whilst other forms of synaesthesia are to be considered malfunctional because the breakdown of the barriers of modules prevents the module from doing what it was originally designed to do.

[14] If this view is right then it may be used to defend the notion of psychological natural kinds from challenges provided by McGinn, 1978, and Kim, 1993, amongst others. Exactly how would be the topic of another paper.

References

ARBIB, M. A. (1989). Modularity and interaction of brain regions underlying visuomotor coordination. In J. GARFIELD (ed.) *The modularity of mind*. Cambridge, Mass.: MIT Press.

BARON-COHEN, S. & HARRISON, J. (eds.) (1997). *Synaesthesia: classic and contemporary readings*. Oxford: Basil Blackwell.

BARON-COHEN, S., HARRISON, J., GOLDSTEIN, L. & WYKE, M. A. (1993). Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception*, 22, 419-426.

BARON-COHEN, S., BURT, L., SMITH-LAITTAN, F., HARRISON, J. & BOLTON, P. (1996). Synaesthesia: prevalence and familiarity. *Perception*, 25, 1073-1079.

BARON-COHEN, S. (1996). Is there a normal phase of synaesthesia in development? *Psyche*, 2.

CHURCHLAND, P. M. (1988). Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philosophy of Science*, 55, 167-187.

CLAVIÈRE, J. (1898). L'audition colorée. *L'Année Psychologique*, 5, 161-178.

CUMMINS, R. (1975). Functional analysis. *Journal of Philosophy*, 72, 741-765.

FODOR, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 28, 97-115.

FODOR, J. (1983). *The modularity of mind*. Cambridge, Mass.: MIT Press.

GARFIELD, J. (ed.) (1989) *The modularity of mind*. Cambridge Mass.: MIT Press.

GROSSENBACHER, P. G. (1997). Perception and sensory information in synaesthetic experience. In S. BARON-COHEN & J. HARRISON. (eds.) *Synaesthesia: classic and contemporary readings*. Oxford: Basil Blackwell.

KARMILOFF-SMITH. A. (1994). *Beyond modularity: a developmental perspective on cognitive science*. Cambridge, Mass.: MIT Press.

KIM, J. (1992) Multiple-realizability and the metaphysics of reduction. *Philosophy and Phenomenological Research*, 52, 309-35.

MARLSEN-WILSON, W. D. & TYLER, L. K. (1989). Against modularity. In J. GARFIELD (ed.) *The modularity of mind*. Cambridge, Mass.: MIT Press.

MARR, D. (1982). *Vision*. San Francisco: Freeman.

MARSHALL, J. C. (1984). Multiple perspectives on modularity. *Cognition*, 17, 209-242.

MAURER, D. (1993). Neonatal synaesthesia: implications for the processing of speech and faces. In B. DE BOYSSON-BARDIES, S. E. SCHONEN, P. JUSCZYK, P. MCNEILAGE & J. MORTON (Eds.) *Developmental neurocognition: speech and face processing in the first year of life*. Dordrecht: Kluwer.

McGINN, C. (1978). Mental states, natural kinds and psychophysical laws. *Proceedings of the Aristotelian Society*, 78, 195-221.

MILLIKAN, R. G. (1984). *Language, thought and other biological functions*. Cambridge, Mass.: MIT Press.

PAULESU, E., HARRISON, J., BARON-COHEN, S., WATSON, J., GOLDSTEIN, L., HEATHER, J., FRACKOWISK, R. & FRITH, C. (1995). The physiology of coloured hearing: a PET activation study of colour-word synaesthesia. *Brain*, 118, 661-676.PUTNAM, H. (1984). Models and modules. *Cognition*, 17, 253-264.

SAMUELS, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *The British Journal for the Philosophy of Science*, 49, 575-602.
SEGAL, G. (1997). Synaesthesia: implications for the modularity of mind. In S.BARON-COHEN & J. HARRISON. (eds.) *Synaesthesia: classic and contemporary readings*. Oxford: Basil Blackwell.
SHAPIRO, L. & EPSTEIN, W. (1998). Evolutionary theory meets cognitive psychology: a more selective perspective. *Mind and Language*, 13, 171-194.
SOBER, E. (1985). Panglossian functionalism and the philosophy of mind. *Synthese*, 64, 165-193.
STILLINGS, N. (1989). Modularity and naturalism in theory of vision. In J. GARFIELD (ed.) *The modularity of mind*. Cambridge, Mass.: MIT Press.
TOOBY, J. & COSMIDES, L. (1995). Forward in S. BARON-COHEN, *Mindblindness: an essay on autism and theory of mind*, Cambridge, Mass.: MIT Press.
WAGER, A. (1999).: Synesthesia and representationalism: the extra qualia problem. *Philosophical Psychology*, 12, 263-281.
WALSH, D. M. & ARIEW, A. (1996). A taxonomy of functions. *Canadian Journal of Philosophy*, 26, 493-514.