A dilemma for modular architectures of the mind

Dustin Stokes (University of Toronto)
Vincent Bergeron (University of Ottawa)

Abstract

Modular architectures of the mind can vary both with respect to the strength of the notion of modularity and the scope of the modularity of mind. We propose a dilemma for modular architectures, no matter how they vary along these two dimensions. First, if a modular theory commits to the *informational encapsulation* of modules, then modules are on this account impenetrable. However, there are plausible cases of the cognitive penetrability of perception. And so any strongly modular theory of perception is threatened. Second, many recent massive modularity theories weaken the strength of the notion of module, while broadening the scope of modularity. These theories do not require any robust informational encapsulation, and thus avoid the incompatibility with cognitive penetrability. However, the weakened commitment to informational encapsulation significantly weakens the explanatory force of the theory and, ultimately, is conceptually incompatible with the core of modularity.

Amongst philosophers and cognitive scientists, modularity remains a popular choice for an architecture of the human mind. Jerry Fodor (1983), who was very influential in establishing the concept of a module, writes: "One day . . . Merrill Garrett made what seems to me to be the deepest remark that I have yet heard about the psychological mechanisms that mediate the perception of speech. 'What you have to remember about parsing is that basically it's a reflex.'" (Dedication). Reflexes are quick, inflexible, and involuntary responses to stimuli, and Fodorian modules are reflexes. In its most recent and general form, the modularity hypothesis consists in viewing the human mind, or at least part of it, as a configuration of quick specialized mental mechanisms, or subsystems, that are functionally independent of one another, that are localized in definite regions of the brain, and that operate over a distinct domain of information.

There are a number of compelling theoretical and empirical motivations for this approach. Theoretically, modularity provides a clear and tractable explanandum for cognitive science, it provides materials for a simple explanation of cognitive and perceptual dissociations, and nicely accommodates adaptationist and other evolutionary explanations of mental phenomena. Modularity also well-accommodates important empirical data. For instance, it provides a simple explanation for the speed and robustness of processing enjoyed by the human mind.

Although it is sometimes misrepresented as doing precisely this, Fodor's pioneering discussion of the concept did not involve a definition of 'module'. (Fodor 1983; see also Coltheart 1999). Fodor did, however, provide a list of properties symptomatic of modules. Fodorian modules are *typically* domain specific, hardwired, computationally autonomous, informationally encapsulated, fast, and their operation is likely mandatory. It is noteworthy how much of this characterization follows the *reflex* metaphor. Domain-specificity parallels the singularity of the stimulus that sets off a reflex; autonomy, mandatoriness, hardwiring, and encapsulation mirror the standard reflex-arc model. Fodor maintains that "The notion of modularity ought to admit of degrees" (Fodor 1983: 37), and that "if a psychological system has most of the modularity properties, then it is very likely to have all of them" (Fodor 1983: 137). Importantly, Fodor claimed only that *input systems* are modular. His primary subject matter was perceptual systems, but he also made the case for systems devoted to low-level linguistic decoding. Higher-level conceptual or cognitive systems, then, are not modular on Fodor's general architecture.

Commitments with respect to Fodor's original analysis of modularity vary. Several modularity theorists take domain-specificity to be definitive of modularity (e.g. Coltheart 1999). Fewer require innate specificity, even if related explanations and arguments often invoke evolutionary considerations. Others maintain that modules are informationally encapsulated and computationally autonomous (e.g. Farah 1994, Sperber 1996; 2001.)[1] Recent theorists have extended the modularity thesis beyond Fodor's input systems. It is common among evolutionary psychologists to endorse some version of what Dan Sperber (1994) has called the *massive modularity thesis*. The general hypothesis states that all, or nearly all, of the mind is modular, and modules have been postulated to account for cognitive capacities as diverse as theory of mind, face recognition, cheating detection, reading, and a variety of social understanding abilities.

As it will be understood here, if a module *m* is informationally encapsulated then *m* cannot, during the course of its processing, access or compute over information found in other components of the overall system. As such, an encapsulated module *m* is *impenetrable* with respect to the other components of the system, since the processing of *m* is insensitive to the information and processing occurring elsewhere in the system (Pylyshyn 1980). In the modularity literature, such a system is thereby

[1] In at least one place, Fodor himself explicitly states that "informational encapsulation is an essential property of modular systems" (Fodor 1985: 3). Though elsewhere he is less clear on his commitment regarding this very point.

considered computationally autonomous. This basic analysis of modularity is important to any substantive modular account of the mind because it constitutes the foundation of modularity in so far as modules are, in a sense to be explained later, *functionally independent systems*.

In this respect, the modularity theorist faces a dilemma that hinges on the commitment to informational encapsulation. On the one hand, a commitment to informational encapsulation is inconsistent with the cognitive penetration of perceptual experience. And, we will attempt to show, there are plausible cases of cognitive penetration of perceptual experience. On the other hand, as recent modularity theorists have done, one might weaken the notion of 'module' so as not to require any robust informational encapsulation. The result, however, is an account that is inconsistent with one of the central motivations for modular architectures and, more fundamentally, with the conceptual core of the very notion of modularity. The first horn challenges *Fodorian modularity*, or a strong notion of modularity, which includes a commitment to informational encapsulation. The second horn challenges *massive modularity* which broadens the scope of modularity but weakens the notion of modularity so as not to require informational encapsulation. Either way, modular architectures appear to be significantly challenged.

I. Informationally encapsulated modules: Cognitive penetrability and the challenge for Fodorian modularity

Both Fodorian and massive modularity theorists take perceptual systems to be modular. If perceptual modules are informationally encapsulated, then at the very least, they are not penetrable by the information or processing of higher-level cognitive or conceptual systems. Most theorists seem to take the concepts 'informational encapsulation' and 'cognitive impenetrability' to be co-extensive, if not equivalent—Fodor in fact originally argued for the encapsulation of modules by arguing against claims about the cognitive penetrability of input systems (Fodor 1983: 73-86). The following discussion requires only the assumption that informational encapsulation *of perceptual modules* entails cognitive impenetrability.

On this account, then, perceptual processing is not influenced by cognitive states like beliefs, desires, or concepts. Evidence of cognitive penetration of perception will thus threaten any modular theory that includes a commitment to informational encapsulation of perceptual modules.

A cognitive impenetrability thesis, regarding perceptual experience, says that for any two perceivers, if one holds fixed the object or event of perception, the perceptual conditions, the spatial attention of the subject, and the conditions of the sensory organ(s), then the perceptual experiences of those perceivers will be identical (see Macpherson, forthcoming). If the experiences of the two perceivers are distinct in these circumstances, and as a result of distinct cognitive states of the perceivers, then experience is, instead, cognitively *penetrable*.[2]

[2] Susanna Siegel defines cognitive penetration as follows. "*Cognitive Penetrability (second pass):* If visual experience is cognitively penetrable, then it is nomologically possible for two subjects (or for one subject in different counterfactual circumstances, or at different times) to have visual experiences with different contents while seeing and attending to the same distal stimuli under the same external conditions, as a result of differences in other cognitive (including affective) states" (Siegel, forthcoming). For a definition of one

It will be useful here to offer some clarifications. First, distinguish perceptual experience from higher-level cognitive and affective states and processes like belief, judgement, desire, emotion, and so on. The former is, whatever else one says about it, characterized by phenomenal character or content and depends non-trivially on one or more sensory organ. Debates abound regarding how to draw the line between perception and cognition. The only point that need be granted here is that there are clear cases of perceptual states and clear cases of cognitive states. So there *are* visual experiences, auditory experiences, olfactory experiences, and so on; and these can be distinguished from states like belief and processes like decision making.

Second, distinguish the cognitive penetration of perceptual experience from the cognitive penetration of perceptual processing. The former concerns some difference in the phenomenal content or character of a perceptual experience, where this difference depends non-trivially upon some cognitive state or processing in the system. The latter only concerns some cognitive effect on perception at the level of processing. The fact that processing at *some* stage is cognitively penetrated does not, by itself, entail the cognitive penetration of experience. Experience may depend on a wider class of processing and, in principle, the cognitive influences on perceptual processing (at some particular stage or other) may not ultimately influence conscious experience. Moreover, perceptual processing may not give rise to a conscious experience but rather, for example, the sub-personal guidance of motor performance (Goodale and Milner 1992).

particular type of cognitive penetration—desire-influenced perception—and how it engages with the relevant debates concerning cognitive penetrability and theory-ladenness, see [AUTHOR MASK].

However, cognitive penetration of perceptual experience does, by definition, entail cognitive penetration of perceptual processing *at some level*. There is much to be said here. The only assumption we need regarding the relation between perceptual experience and perceptual processing is this. Whether one takes experience to be identified with, constituted by, supervenient upon, or the output of perceptual processes, a difference in perceptual experience implies a difference in perceptual process. This will be generally true—albeit for different reasons—no matter how one's metaphysics of mind varies according to these alternatives. So if experience is penetrated, then information relevant to cognitive systems, or the processing of cognitive systems, directly influences the processing of perceptual systems. It is this entailment relation that is important for the criticism we offer below: if perception is cognitively penetrated, then the modularity theorist cannot maintain a mental architecture whereby perception is generally encapsulated (or even whereby distinct sensory modalities are encapsulated).

One final point: Although the cognitive penetration of experience entails the penetration of processing at some stage, the cognitive penetration of experience is compatible with the cognitive impenetrability of processing at some stage or even most (but not all) stages. This point is instructive: one cannot argue from the purported fact that some particular perceptual module is impenetrable to the claim that perception broadly or perceptual experience (in that modality) is impenetrable. Zenon Pylyshyn, for instance, argues that "early vision" is not penetrable by cognition (Pylyshyn 1999). Even granting Pylyshyn's empirical claim about early visual processing, this claim does not imply that experience is cognitively impenetrable. His claims about the impenetrability of a particular stage of processing, early vision, even if true are insufficient to support the thesis that perception is cognitively impenetrable. Indeed, Pylyshyn admits that the output of this component of the visual system, as he and most theorists understand it, does not (alone) determine perceptual experience. His defence of cognitive impenetrability is thus consistent with the cognitive penetrability of perception; one might accept that the computations performed by the early visual system are impenetrable by cognitive states but maintain that perceptual processing is penetrated elsewhere such that the resulting perceptual experience is causally dependent upon cognition[3]

In the middle 20th century, the New Look movement in psychology, in a slogan, theorized perception as an active construction of representations of the environment. Importantly, perceptual representations are constructed in a way informed by the perceiver's "mental set"—her expectations, needs, values, desires, and other higher-level states. The New Look psychologists performed a number of studies that ostensibly provided evidence for the cognitive penetration of experience. Although initially very influential, the New Look model of perception, and the purported evidence for the model, has been largely dismissed. And out with the theory went plausible candidate experimental cases for the cognitive penetrability of experience. This dismissal was unwarranted. Some of the classic research done by New Look

[3] A number of critics have questioned Pylyshyn's conclusions in this general way (Bermudez 1999; Macpherson, forthcoming; Moore 1999; Noë and Thompson 1999). It is also worth noting that Pylyshyn's empirical claim can be challenged (see Boynton 2005; Kamitani and Tong 2005).

psychologists evades the standard strategies employed to reject it, and in turn remains some of the best existing evidence for cognitive penetrability of experience.[4]

Most of the relevant experimental cases adduced by the New Look psychologists aim to show a difference in the perceptual experience had by experimental subjects (e.g. seeing a stimulus as bigger than it is), where this difference is supposed to be a direct causal result of cognitive states of the perceiver (Bruner and Goodman 1947 is discussed in detail below. See also Postman, Bruner, and McGinnies 1948;. Bruner, Postman, and Rodrigues 1951; Bruner and Minturn 1955). Thus, it appears, cognition penetrates perception. Critics of the New Look movement have most commonly deflected these putative cases of cognitive penetration by the following three strategies.

First, critics claim that what is affected by the subject's cognitive states is the subject's memory rather than her perceptual experience. Subjects recall the stimulus to be some way (i.e. a way different than if the relevant cognitive states are controlled for) as a result of some other cognitive state, and report a memory of the stimulus rather than a perceptual experience. This evidences cognitive penetration of cognition. And this is uncontroversial: memories can be faulty, and in ways influenced by what we believe, desire, or otherwise think. Call this the *memory interpretation*.[5]

---

[4] See Balcetis and Dunning 2006 for a brief historical discussion of the rise and fall of the New Look movement, as well as a new set of studies in the New Look spirit.

[5] For one example, see McCurdy 1956. Note also that if one takes memory to be veridical, then this interpretation is conceptually problematic. We will set to one side this defence against the criticism, since even granting the conceptual coherence of the memory interpretation, it fails, as argued below, to deflect important cases.

A second strategy is the *attention-shift interpretation*. This interpretation maintains that in the cases in question, cognitive states of the experimental subjects cause a shift in attention, generally involving some overt action, which then results in the change in perceptual experience. Thus the link between cognition and perception is mediated by an external movement or action. For example, Pylyshyn rules out attention-shift cases as non-genuine cases of cognitive penetrability by appeal to the fact that in such cases there is no internal, logical connection between the belief, goal, or other cognitive state and the computations performed by the perceptual system (Pylyshyn 1999: 343). Lacking this internal connection, there is nothing to properly call 'penetration'.[6]

Finally, critics have suggested that the experimental subjects are not reporting a cognitively affected perceptual experience, but instead a judgement of the perceived stimulus. So the perceptual experience of the stimulus remains unaffected. At most, the subject judges or evaluates the stimulus in a way she would not if she lacked some background cognitive state/s. This difference manifests in the different reports of the experimental subjects versus the control subjects in the New Look studies. Call this the *judgement interpretation*.[7]

Grant that if any putative case of cognitive penetration can be interpreted in one of these alternative ways, then the critics are correct: it is *not* a genuine case of cognitive penetration of experience. We can then define cognitive penetration so as to rule out these interpretations, and ask if any case plausibly meets the definition. If the answer

[6] Fodor also appeals to this general response in his debate with Paul Churchland on the theory-ladenness of perception/observation (see Fodor 1988; Churchland 1988; see also Fodor 1983).
[7] For extended discussion of these and other strategies for the cognitive impenetrability theorist, see Macpherson, forthcoming; AUTHOR MASK XXXX.

is ‘yes’, then the critics must secure some alternative interpretation to deflect the case/s. Here is such a definition:

> *(CP) A perceptual experience E is cognitively penetrated if and only if (1) E is causally dependent upon some cognitive state C and (2) the causal link between E and C is internal.*

The definition requires a few qualifications. First, assume an orthodox understanding of ‘cognitive state.’ Cognitive states are representational and possess some kind of linguistic or propositional content, and they play a role in reasoning and decision making. On this minimal understanding of ‘cognitive state’, standard cognitive states include beliefs, desires, concepts, values and, perhaps, emotions.

Second, clause (2) says that if an unscreened internal cause involves a cognitive state—that is, the causal chain runs from experience back to a belief, desire, or some other cognitive state without deviating from internal mental processes—then the perception depends (internally) upon a cognitive state. Counterfactually, had C not been present in the process, E would not be had by the subject. C is thus a necessary causal condition for E. Understood probabilistically, C is not a strictly necessary causal element, but one that is highly relevant to the probability of that perceptual experience; E is more likely to be had when C is present, and less when not present.

The preferred notion of causation is of little matter so long as the internal causal dependence is maintained.[8]

Finally, (CP) excludes non-genuine cases of cognitive penetration. For example, a desire to see the show, coupled with a true belief about the location of the show, may result in a perceptual experience of the show (several experiences in fact). But this should not count as an instance of cognitive penetration of experience, lest the concept 'cognitive penetration' becomes trivial. (CP) delivers the appropriate result: such a case fails to satisfy our definition, since it fails to satisfy clause (2). In cases like this, a cognitive state (or some cognitive states) motivates an action (or set of actions) which eventually results in the relevant experience. The perceptual experience thus causally depends upon the relevant cognitive state/s. Clause (2) insures, however, that such cases are not instances of cognitive penetration, since in each case the cognitive states are screened from being *internally*, causally efficacious: the cognitive states cause an (external) action which eventually results in the experience.[9]

Now for the importance of (CP): A perception that satisfies this definition cannot be interpreted in any of the three ways described above. Clause (1) of (CP) rules out both the memory and judgement interpretation, since it requires a cognitive influence on perception, rather than just an influence on some other cognitive state in the system. Clause (2) of (CP) rules out the attention-shift interpretation, since it requires a non-externally mediated causal link between the cognitive state and the perceptual experience. The question now becomes: are there any experimental cases that satisfy

---

[8] One should also note that C is a non-sufficient cause of E. There are other relevant causal factors. Again, counterfactually, cognitive states are causally relevant, such that if state C had not been present in the perceptual-cognitive system of the agent, then that agent would not have had perceptual experience E.
[9] In this way, CP is consistent with other recently proposed definitions of cognitive penetrability, discussed above (Macpherson, forthcoming; Siegel, forthcoming).

(CP)? The answer is 'yes'. In fact, one of the most compelling cases comes from one of the earliest of New Look studies.

In a now famous experiment, Jerome Bruner and C.C. Goodman tested perceptual experiences of objects of social value (Bruner and Goodman 1947). Three groups (10 persons per group) of 10 year old children, two experimental and one control, were put before a wooden box with a glass screen on its face. In the centre of the screen was a small patch of light, nearly circular in shape, the circumference of which could be adjusted by a small knob located on the bottom right corner of the box. The two experimental groups of children were presented with ordinary coins of varying values. As they looked at the coins, placed flat in the palm of the left hand, positioned at the same height and six inches to the left of the adjustable patch of light, they were asked to adjust the patch to match the size of the presented coin. The subjects could take as much time as they liked to complete the task. The control group was instead presented with cardboard discs of sizes identical to the relevant coins, and asked to perform the same task. In the experimental group, perceptual experience of the coins was "accentuated." The experimental subjects systematically overestimated the size of the coin, and sometimes by a difference as high as 30% as compared with control subjects.[10]

The second, more famous, experimental variation divided experimental groups into subgroups comprising "rich" and "poor" children. The task was the same, except only real coins were used. Here, rich children, as the previous results would suggest, still overestimate the size of the coins, but at percentages significantly lower than the

[10] For example, experimental subjects overestimated the size of a dime by an average of 29%; controls underestimated the size of the cardboard analogue of a dime by -1%.

poor children. Indeed, poor children systematically overestimate the size of coins, by as much as 50%, and by differences as high as 30% as compared to rich children.[11]

Bruner and Goodman's explanation was that the social value or desire for money was somehow affecting the perceptual experiences of the children. Children desire money and this results in their seeing the coins as bigger than they are. And for the poor children, who had a greater desire or need for money, this effect was greater. This case, *prima facie*, satisfies (CP). The experimental subjects have a perceptual experience, the character or content of which causally depends on a cognitive state, in this case, a desire or value. And the causal link between the experience and cognitive state is internal.

Nonetheless, critics have traditionally rejected this as a genuine case of cognitive penetration, and by appeal to one of the strategies outlined above. However, this is a mistake premised on a failure to carefully consider Bruner and Goodman's experimental procedures. In all the variations described above, subjects took as much time as they needed to adjust the light patch to match the size of the coins. The coins were presented at the same time as, at the same horizontal level as, and six inches to the left of, the adjustable light patch. Subjects did *not* visually inspect the coin and then shift to a distinct visual field, adjusting the light patch by memory. Instead, their task was to adjust the patch of light to match what they were seeing, *while* they were

[11] Although there were a number of theorists critical of particular details and the broad scope of the New Look approach (Klein, Schlesigner, and Meister 1951; Carter and Schooler 1949; Lysak and Gilchrist 1955), the Bruner and Goodman 1947 results have been replicated by a number of similar studies (in addition to those listed above, see Bruner and Postman 1948; Dukes and Bevan 1952; Bruner and Rodrigues 1953; Blum 1957; Holzkamp and Perlwitz 1966), at least insofar as these studies all evidence some higher level effect on perceptual experience. Only this last result is needed for present purposes; no commitment to the full scope and detail of the New Look theory is required.

seeing it. In no relevant sense were they forced to base their report just on memory. The memory-interpretation thus fails.

For the same reason, the attention-shift interpretation fails: subjects did *not* attend to one stimulus (the disc or coin) and then shift attention to a distinct visual field where the second stimulus (the adjustable light patch) was located. The subjects would have shifted their gaze from disc/coin to light patch, but this slight shift would not have differed across control and experimental subjects (and would thus fail to explain the relevant differences between the two groups).

The judgement interpretation is most commonly used to deflect the Bruner and Goodman case. The only version of this interpretation that is inconsistent with cognitive penetrability is one that claims that the perceptual experiences of the subjects are accurate across control and experimental subjects alike, while the experimental subjects make a misjudgement of the size of the coins. This interpretation is less plausible than the interpretation it opposes, since it requires attributing a judgement or belief to the child which does *not* correspond to the perceptual experience that she has simultaneously with that judgement or belief. It requires that the subject, while inspecting the visual stimulus—which is, again, at a constant location six inches to the immediate left of the adjustable light patch—consistently makes erroneous judgements about what she is seeing.

Here, the cognitive impenetrability theorist is likely to respond by invoking instances where perception and judgment do come apart in just this way. So, for example, although one sees the Müller-Lyer lines as being of different lengths, one believes (if one knows the illusion) that the lines are of the same length. And indeed one cannot manage to see them accurately in spite of this background knowledge (Fodor 1983, 1985, 1988; Pylyshyn 1999). So, the critic would argue, a consistent mismatch between simultaneous experience and judgment is not so uncommon. However, the subjects in the Bruner and Goodman experiments are importantly different from standard perceivers of the Müller-Lyer and other such illusions. When one judges and reports that the Müller-Lyer lines are of the same length, one bases this report *not* on perceptual experience, but on knowledge of the illusion. Bruner and Goodman's subjects are different in this regard: they intend for their report to be one of what they see. If asked, the subjects would certainly report that they are matching the light patch to what they are seeing. To treat these subjects like perceivers of the Müller-Lyer illusion requires that they are systematically mistaken about this: the subjects are not correctly reporting what they see.

The judgement interpretation, then, must maintain that these subjects are continually ignoring, remaining unconscious of, or somehow otherwise failing to accurately report their perceptual experience. This, given the experimental circumstances, is far less plausible than the interpretation it opposes—namely, that experience itself is penetrated and then reported by the subjects. The judgement interpretation thus fails.[12]

If the above discussion is successful, then the standard strategies for dismissing putative cases of cognitive penetration fail to deflect the Bruner and Goodman cases, and the latter remain genuine cases of cognitive penetration of perception. The

[12] There are other experimental cases, old and new, that seem to evade the same strategies. Fiona Macpherson argues that Delk and Filenbaum's 1965 colour perception results cannot be deflected by the standard strategies (see Macpherson, forthcoming). More recently, see Balcetis and Dunning 2006; Balcetis and Dunning, 2010; van Ulzen 2008.

experiences of the experimental subjects are penetrated by some higher-level (non-perceptual) state or process. We should note that there is flexibility regarding how the case is further interpreted, in the following respect. We favour interpreting the results so the cognitively influencing state is a desire or value. Alternatively, one might think that it is a concept or belief regarding money that is causally operative. Or one might favour a less direct mechanism, where a belief or concept influences a visual image, which in turn influences the character of experience.[13] In any case, Fodorian modularity will be challenged by this case, since whether it is a desire, value, concept, belief, or some other higher-level mental state, there is *some* non-perceptual state (internally) influencing experience. And therefore, as we will now argue, perception is not operating in an informationally encapsulated way.

Recall that perceptual systems are paradigms for modular systems. And recall further that if one is a Fodorian modularity theorist, then one commits to the informational encapsulation of modules. The informational encapsulation of perceptual modules entails cognitive impenetrability. Finally, the cognitive penetration of perceptual experience entails, at some level, the cognitive penetration of perceptual processing. Thus any legitimate case of the cognitive penetration of experience undermines the purported informational encapsulation of the relevant perceptual systems, and in turn challenges any modularity theorist that commits to this feature as necessary for modules.

Here, finally, is the first horn of the dilemma for modular architectures of the mind. There are legitimate cases of the cognitive penetration of experience (Bruner and

[13] See Macpherson (forthcoming) for such an account.

Goodman 1947 is the example we have defended). And so perceptual systems—in this case *vision*—are not informationally encapsulated. Any modularity theory that commits to informational encapsulation (and by implication: cognitive impenetrability) as necessary for modularity is therefore threatened.

To clarify our critique, it will be useful to briefly consider a hypothetical defence for the Fodorian modularist in response to this first horn of the dilemma. Our suggestion is not that perception (or vision, more specifically) is, as it were, unencapsulated through-and-through. As discussed above, the entailment relations between the cognitive penetration of perceptual experience and the cognitive penetration of perceptual processing would not support this last inference. So, the modularity theorist might retort, the penetration of experience is compatible with the impenetrability (and thus encapsulation) of *some* (but not all) components or systems in perceptual processing. Thus one can reasonably maintain that components of perceptual systems are encapsulated and, to this degree, strongly modular.[14]

All of this may be correct, but it does nothing to save a modular *architecture* of perception. A theorist might reasonably maintain that sub-components in vision, say, are encapsulated and thus modular. For example, feature detecting components like groups of simple and complex cells in the primary visual cortex are likely encapsulated, as are many other neural circuits and low-level components in the overall visual system. In fact, theorists may be able to maintain that certain sub-systems in vision—for example, Pylyshyn's early vision—are encapsulated in spite of the penetration of visual experience. This would be to maintain the commitment to

[14] Thanks to XXXX for pressing us to consider this response.

informational encapsulation and thus a *strong* Fodorian notion of 'module'. But note that the *scope* of modularity on such a view is significantly weakened: such a modularity theorist can only claim that *some* of the visual system is modular and, importantly, cannot claim that vision is, generally, modular. This last claim *is* inconsistent with genuine cases of cognitive penetration.

Where does this leave the view? The claim that individual low-level circuits are encapsulated and thus strongly modular is largely uncontroversial among cognitive scientists. And the claim that *some* sub-systems—even Pylyshyn's early vision—in perception are strongly modular is insufficient to support the claim that the general structure of perception (or, more specifically, vision) is strongly modular. In short, one cannot save a Fodorian modularity by appeal to encapsulated components or sub-systems of perception. To do so would be to opt for strength of modules over scope, in turn undermining the theory as an *architecture* of the mind.

II. Informationally unencapsulated modules: A challenge for the massive modularity hypothesis

A number of recent theorists have accordingly weakened the notion of modularity with respect to Fodor's original characterization and, in particular, with respect to informational encapsulation. This change in the notion of modularity has tended to accompany a broadening of the scope of modular theories. Thus, massive modularity theorists take much if not the whole of the human mind to be modular, including higher level conceptual and cognitive systems. If, as we have argued in the previous section, cognitive impenetrability (and thus encapsulation) seems too strict a requirement on the modularity of perceptual systems, then it makes sense not to require it of higher-level conceptual and cognitive systems. Weakening modularity in this way, however, comes with significant costs to any modular account of cognition. First, it undermines one of the primary theoretical advantages of the modular approach. Second, it threatens the internal coherence of modular theories.

Peter Carruthers, a massive modularity theorist, argues that

> *if a thesis of massive mental modularity is to be even remotely plausible*, then by 'module' we cannot mean 'Fodor-module'. In particular, the properties of having proprietary transducers, shallow outputs, fast processing, significant innateness or innate channelling, and *encapsulation* will very likely have to be struck out. (Carruthers 2006: 12; emphasis added.)

According to Carruthers, massive modularists should expect most (if not all) central cognitive modules to be *un*encapsulated. He writes:

> …even where a system has been designed to focus on and process a particular domain of inputs, one might expect that in the course of its normal processing it might need to query a range of other systems for information of other sorts. (Carruthers 2006: 10).

In other words, an unencapsulated module, in order to perform its task, will often need to compute over information that is made available by other systems. For example, the mind-reading system "may need to query a whole range of other systems for information relevant to solving the task in hand" (Carruthers 2006: 11).

Evolutionary psychologists, many of whom subscribe to the massive modularity hypothesis, also tend to argue for (or at least assume) the compatibility of modularity with unencapsulation. Hagen (2005) explicitly states what is often implicitly assumed in this field:

> Why, except when processing speed or perhaps robustness is exceptionally important, should modules not have access to data in other modules? Most modules should communicate readily with numerous (though by no means all) other modules when performing their functions, including querying the databases of selected modules. (163).

Any such modularity theorist thus claims that systems, like the mind-reading system, can be modular *in spite of* being informationally unencapsulated. As Carruthers suggests, this might be a necessary adjustment of a general modular architecture for the simple reason that anything stronger is implausible.

One main theoretical advantage of, and indeed motivation for, postulating modular architectures is that they explain dissociations between perceptual or cognitive functions. A cognitive task *A* is said to be dissociated from cognitive task *B* when at least some individuals are observed who show a significant deficit with respect to *A* in the absence of a corresponding deficit in *B*. *A* and *B* are said to be *doubly* dissociated when, in addition, we observe individuals in whom *B* is significantly impaired without a corresponding deficit with regard to *A*. Cognitive scientists generally hold that dissociations are signs of functional separateness, and will often postulate the existence of cognitive modules on the basis of such behavioral patterns. If *A* is observed to fail when *B* does not, then one may infer that *A* is performed by a process, or module, *M* that *B* does not recruit. When this process *M* is obstructed, it is argued, *A* fails and *B* does not. In the case of a double dissociation, the inference is stronger, namely that *A* and *B* are each performed by a process, or module, that the other does not recruit.

A classic, although by no means uncontroversial, case of the stronger version of the inference concerns the face recognition module hypothesis. Many theorists postulate the existence of a face recognition module on the basis of a double dissociation between face recognition and other cognitive and perceptual functions such as visual object recognition. On the one hand, there are well-documented cases in the neuropsychological literature of patients who have a face recognition deficit (prosopagnosia) but who appear to retain all other cognitive and perceptual capacities, including the ability to respond to and identify an array of physical objects (i.e. non-faces) (e.g. Rossion et al. 2003). On the other hand, there are reports of patients with impaired visual object recognition who appear to retain all other cognitive and perceptual capacities, including the ability to respond to and identify familiar faces (Rumiati & Humphreys 1997). This is, ostensibly, a double dissociation between face and object recognition and it has motivated some theorists to posit two corresponding, separate modules: a face recognition module and a visual object recognition module (Coltheart 1999: 119).

This reasoning from dissociation data to modularity—call it the *functional modularity inference*—has been central to the development of modern neuropsychology. In the last thirty years, philosophers and cognitive scientists have refined concepts of dissociation and narrowed the scope of the inference (see Shallice 1988 for a detailed discussion of this methodology). Moreover various authors have questioned the legitimacy of the inference, on both theoretical and empirical grounds, by arguing that the existence of a double dissociation between subjects' performances on two different cognitive tasks does not necessarily constitute strong evidence for the existence of separate cognitive functions or modules (Dunn & Kirsner 2003; Juola, & Plunkett 2000; Plaut 1995; Van Orden, Pennington, & Stone 2001). In the light of these developments, a better way to think about the functional modularity inference is to see it as an inference to the best explanation. In this case, one infers that a cognitive system is modular on the grounds that this hypothesis best explains a set of dissociation data, given the fact that modular systems are known to produce dissociations when damaged in different ways (Coltheart 2001).

But suppose, as the weakened modularity theory we're considering does, that modules are *not* encapsulated. Suppose, for example, that the face and object recognition modules are not encapsulated, that in order to perform their task they each need to compute over information that is made available by other systems (and perhaps each other). It is difficult to see, in this case, how this assumed modular architecture could explain (let alone *best* explain) the observed double dissociation between face and object recognition, since it is the assumed functional separateness of cognitive modules, in the functional modularity inference, that is supposed to explain the existence of dissociation data.[15] Since both the face and object recognition modules are by hypothesis unencapsulated, dissociation data concerning these tasks could no longer be taken to be signs of functional separateness, as either module (or both) could fail for reasons that have nothing to do with a failure of their mechanisms. Note that the point here is not that unencapsulated face and object recognition modules could not, when lesioned in certain ways, give rise to a double dissociation between the two tasks—a dissociation between face and object recognition could occur, for example, if there is a type of information outside the face and object recognition modules that face recognition more significantly draws on than object recognition does, and the opposite dissociation could occur for similar reasons (see Shallice 1988 for discussion). It is rather that a double dissociation between face and object recognition could not, in this case, be *explained* by an appeal to (unencapsulated) modularity.

By contrast, a double dissociation between face and object recognition is just what we would expect from Fodorian-strength modularity. Importantly, because the face recognition module would in this case be informationally encapsulated, its normal functioning would thereby not depend on information made available by other modules, including the object recognition module. And moreover, the normal functioning of these other modules would not be impaired by the malfunction of the

[15] The other property most commonly associated with modules is neurological distinctness. For example, Tim Shallice defines modularity in terms of *functional* and *neurological distinctness*. While neurological distinctness is what explains the (contingent) fact that functional dissociations often arise in patients showing fairly localized brain lesions, it is functional distinctness (defined in terms of computational autonomy) that explains the very presence of *functional* dissociations.

face recognition module. This kind of modularity can thus easily account for the impairment in the ability to recognize familiar faces while other object recognition abilities remain perfectly intact. And the complementary explanation would be given for a patient showing the opposite dissociation.

Functional dissociations are not the only kind of evidence in favour of modular architectures. One might appeal, as Fodor (1983) does, to the reflex-like nature of input systems. Generally, perceptual processes are fast, mandatory, inflexible, and involuntary. Two of Fodor's favourite types of examples are speech recognition processes and visual illusions. "You can't help hearing an utterance of a sentence (in a language you know) as an utterance of a sentence" (52-3), and "[t]he very same subject who can tell you that the Müller-Lyer arrows are identical in length . . . still finds one looking longer than the other" (66). As these cases show, various other properties of cognitive systems (e.g. speed of processing, inflexibility) can be used to motivate a modularity thesis. However, as these are themselves symptoms of informational encapsulation and cognitive impenetrability (the properties which Fodor considers to be the core of modularity), they hardly constitute an independent line of evidence in favour of (unencapsulated) modular architectures.

Carruthers, nonetheless, wants to maintain that in the weakest sense of the term, a module is a "dissociable functional component" (Carruthers 2006: 2). Carruthers' commitment here is once again representative: modularists of all strengths maintain that modules are *dissociable*, and thus are *functionally independent systems* (for example, theorists as diverse as Tim Shallice, Jerry Fodor, and Dan Sperber understand modules in this way). This seems to be, if there is one, the core commitment of a modular architecture.

The challenge here for the modularity theorist is that a cognitive system cannot be both informationally unencapsulated *and* functionally dissociable.

On the one hand, a cognitive system is considered functionally independent if it can be dissociated from other cognitive systems. This means that the system is minimally affected by what happens to other systems, and that it can be modified (or damaged) without affecting the normal functioning of the other systems. Functional dissociation is thus a sufficient condition for functional independence.

On the other hand, when a system is unencapsulated—when it needs to compute over information made available by other systems—that system depends for its normal operation on substantial interaction with other systems. This system is thus functionally dependent on other systems, and is therefore *not* dissociable from them. Therefore, if functional independence is understood in terms of dissociation, then a module cannot be both an unencapsulated and functionally dissociable system.

An illustration may help. As Carruthers suggests, modules must be, at the very least, dissociable functional components. To illustrate this "everyday" sense of module, he offers the following metaphor.

> The hi-fi is modular if one can purchase the speakers independently of the tape-deck, say, or substitute one set of speakers for another for use with the same tape-deck. Moreover, it counts towards the modularity of the system if one doesn't have to buy a tape-deck at all—just purchasing a CD player along with the rest—or if the tape-deck can be broken *while the remainder of the system*

continues to operate normally. (Carruthers 2006: 2; emphasis added).

Carruthers goes on to suggest that although operationally distinct in the above ways, the components of the hi-fi, once conjoined as a system, do depend upon one another in other ways: the CD player requires the amplifier to distribute sound, the speakers require input from the amplifier to make sounds, etc. Indeed, some of these dependence relations will be asymmetric: the CD player needs the amplifier to distribute sound, but not vice versa. The important point to note for present purposes is that in spite of these dependence relations, the hi-fi components are (relevantly) computationally or operationally autonomous: the CD player may require the amp to deliver its output, but it does not use or access information in the amp, in reading data off of a CD. In other words, in performing its function, it is encapsulated from the amp, the speakers, and so on. Likewise for other components in the system: the tape-deck reads data, the tuner acquires a radio signal, the speakers deliver a range of sounds, and so on, all independently.

Thus the hi-fi metaphor is a useful one, at least for modularity as traditionally understood. The trouble is that unencapsulated cognitive modules are relevantly disanalogous to hi-fi components. Like the hi-fi modules, the cognitive modules envisaged by massive modularity work as a system, exchanging input and output, and often asymmetrically. But unlike the hi-fi components, the unencapsulated modules, as per the massive modularity theorist, depend on other systems to perform their computational tasks. This is crucial, since it implies that cognitive modules—*un*like the CD player, the amplifier, and so on—depend for their computational operations on other components in the system. Thus if cognitive module *m* breaks down, the remainder of the system will not function normally. And, importantly, if *m* cannot recruit other components in the system, or other components in the system break down, then *m* will fail to perform its function (or, at least, fail to perform its function optimally). Unencapsulated cognitive modules, unlike hi-fi modules, are not dissociable functional components.

But, although dissociation implies functional independence, perhaps functional independence does not imply dissociation. In this case, a cognitive system (e.g. a module) could both be functionally independent and interact substantially with other systems and even recruit them as needed. This means that operations in such a system could both affect and be affected by other systems' operations.

Perhaps we can make this proposal work by stating that a cognitive system is functionally independent if its operations can be specified independently of cognitive operations occurring in other systems. Here, functional independence is defined not in terms of the amount or type of interaction occurring between a system and other systems, but in terms of how distinctly specifiable its contribution is to a given cognitive capacity. For example, a growing body of evidence shows that Broca's area is involved, together with other brain regions, in a relatively large number of linguistic and non-linguistic operations—for example, syntactic operations in both music and natural languages, object manipulation, action perception (Patel 2003; Schubotz & Fiebach 2006). In an effort to explain these findings, Fiebach and Schubotz have proposed a model that could account for the basic contribution of Broca's area to a "wide range of attentional, cognitive, and motor processes" (2006:

499). According to their model, Broca's area would be a hypersequential processor, i.e. a functional component that performs the "detection, extraction, and/or representation of regular, rule-based patterns in temporally extended events" (2006: 501). Broca's area would thus be functionally independent not because it is informationally encapsulated or because its operations can be dissociated from other systems' operations, but because it is an area of the brain that performs a set of operations that can be specified independently of the larger cognitive operations in which it participates.

This account of functional independence is perfectly compatible with cognitive systems being informationally unencapsulated. It is also compatible with the thesis that all, or nearly all, of the mind (or brain) is composed of such functionally independent systems. And if modules were minimally understood as functionally independent systems in this sense of the term, then perhaps this thesis would be a version of the massive modularity hypothesis. But if this is all the hypothesis amounts to, then it is not something about which cognitive scientists disagree. It is equivalent in substance to Norman Geschwind's account of the general architecture of the brain as "more or less specialized groups of cells connected by relatively discrete pathways" (Geschwind 1965). Cognitive scientists consider this to be the most basic architectural principle of the brain. Modular and non-modular theorists alike see specialization within the brain as an undisputed fact. What is at issue is how best to characterize this specialization, and the massive modularity thesis thus weakened is no different from what all cognitive scientists take for granted. Therefore, if functional independence is understood in this way, then (massive) modularity is not a controversial thesis.[16]

This, to conclude, is the second horn of the dilemma for modular architectures of the mind. To weaken the notion of modularity so as not to require informational encapsulation is to weaken the commitment to functional independence of modules. To do so is to opt for scope of modularity over strength of modules. Weakening modularity to this degree undermines one of the main empirical motivations for the theory: explaining behavioural dissociations. Moreover, the weakened notion either is conceptually incompatible with the core of modularity—namely, viewing modules as dissociable functional components—or leads to an uncontroversial thesis about the general architecture of the mind.

Bibliography

Anderson, M. (2010, in press), 'Neural re-use as a fundamental organizational principle of the brain,' *Behavioral and Brain Sciences*.

Balcetis, E. and Dunning, D. (2006) 'See what you want to see: Motivational influences on visual perception,' *Journal of Personality and Social Psychology* 91(4): 612-25.

______(2010) 'Wishful Seeing: Desired objects are seen as closer' *Psychological Science*: 21: 147-52.

Bermudez, J. (1999) 'Cognitive impenetrability, phenomenology, and nonconceptual content,' *Behavioral and Brain Sciences* 22(3): 367-8.

[16] Cowie 2008 challenges Carruthers' modularity theory with a dilemma, one horn of which charges that his notion of modularity is too thin to be conceptually distinct from various other (non-modular) architectures of the mind. For other recent (but distinct) criticism of modularity and massive modularity architectures, see Anderson 2010; Machery 2008; Wilson 2008.

Blum, A. (1957) 'The value factor in children's size perception, *Child Development* 28: 14–18.

Boynton, G. M. (2005) 'Imagining orientation selectivity: Decoding conscious perception in V1,' *Nature Neuroscience*, 8: 541–542.

Brewer, W. F., & Lambert, B. L. (2001) 'The Theory-Ladenness of Observation and the Theory-Ladenness of the Rest of the Scientific Process,' *Philosophy of Science* 68: 176-86.

Bruner, J.S., & Goodman C.C. (1947) 'Value and need as organizing factors in perception,' *Journal of Abnormal and Social Psychology* 42: 33-44.

Bruner, J. S., & Minturn, A. L. (1955) 'Perceptual identification and perceptual organization,' *Journal of General Psychology*, 21-28.

Bruner J. S., & Postman, L. (1948) 'Symbolic value as an organizing factor in perception,' *Journal of Social Psychology* 27: 203-208.

Bruner, J.S., Postman, L., and Rodrigues, J. (1951) 'Expectation and the Perception of Color' *American Journal of Psychology* 64: 216-227.

Bruner, J.S., & Rodrigues, J. S. (1953) 'Some determinants of apparent size,' *Journal of Abnormal and Social Psychology* 48: 17-24.

Carruthers, P. (2006) *The Architecture of the Mind*, Oxford: Oxford University Press.

Carter, L. F., & Schooler, K. (1949) 'Value need and other factors in perception,' *Psychological Review* 56,: 200-207.

Churchland, P. M. (1988) 'Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor,' *Philosophy of Science* 55: 167-187.

Coltheart, M. (1999) 'Modularity and Cognition,' *Trends in Cognitive Science* 3: 115-20.

Cowie, F. (2008) 'Us, Them and It: Modules, Genes, Environments and Evolution,' *Mind and Language* 23: 284-92.

Delk, J. L. and Fillenbaum, S. (1965) 'Differences in Perceived Colour as a Function of Characteristic Color,' *The American Journal of Psychology,* 78: 290-93.

Dukes, W. F., & Bevan, W. (1952) 'Size estimation and monetary value: a correlation,' *Journal of Psychology* 34: 43-53.

Dunn, J. C. & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex*, *39*(1), 1-7.

Farah, M. (1994) 'Neuropsychological inference with an interactive brain: A critique of the locality assumption,' *Behavioral and Brain Sciences* 17: 43-104.

Fodor, J. (1983) *Modularity of Mind*, Cambridge, MA: MIT Press.

______(1985) 'Precis of *The Modularity of Mind*', *The Behavioral and Brain Sciences* 8: 1-5

______(1988) 'A reply to Churchland's "Perceptual plasticity and theoretical neutrality,"' *Philosophy of Science* 55: 188-19

Gentaz, E. and Rossetti, Y. (1999) 'Is haptic perception continuous with cognition?' *Behavioral and Brain Sciences* 22(3): 378-9.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, *88*, 237-294, 585-644.

Goodale, M. A. and Milner, D. (1992) Separate visual pathways for perception and action. *Trends in Neurosciences* 15(1): 20-25.

Hagen, Edward H. (2005). Controversial issues in evolutionary psychology. In D. M.

Buss (Ed.), *The handbook of evolutionary psychology*. (pp. 5-67). Hoboken: John Wiley & Sons.

Holzkamp, K. and Perlwitz, E. (1966) 'Absolute oder relative Größenakzentuierung? Eine experimentelle Studie zur sozialen Wahrnehmung,' *Zeitschrift für experimentelle und angewandte Psychologie* 13: 390-405.

Juola, P. & Plunkett, K. (2000). Why double dissociations don't mean much. In G. Cohen, R. A. Johnston, & K. Plunkett (Eds.), *Exploring cognition: damaged brains and neural networks*. (pp. 319-327). Psychology Press.

Kamitani, Y., & Tong, F. (2005) 'Decoding the visual and subjective contents of the human brain,' *Nature Neuroscience* 8: 679–685.

Klein, G. S., Schlesinger, H. J., & Meister, D. E. (1951) 'The effect of personal values on perception—an experimental critique,' *Psychological Review* 58, 96-112.

Lysak, W., & Gilchrist, J. C. (1955) 'Value, equivocality, and goal availability as determinants of size judgments,' *Journal of Personality* 23: 500-501.

Machery, E. (2008), 'Massive Modularity and the Flexibility of Human Cognition,' *Mind and Language* 23: 263-72.

Macpherson, F. (forthcoming) 'Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism,' *Philosophy and Phenomenological Research*.

McCurdy, H. G. (1956) 'Coin perception studies and the concept of schemata,' *Psychological Review* 63: 160–168.

Moore, C. (1999) 'Cognitive impenetrability of early vision does not imply cognitive impenetrability of perception,' *Behavioral and Brain Sciences* 22(3): 385-6.

Noë, A. and Thompson, E. (1999) 'Seeing beyond the modules toward the subject of perception,' *Behavioral and Brain Sciences* 22(3): 386-7.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat Neurosci*, *6*(7), 674-681.

Plaut, D. C. (1995). Double dissociation without modularity: evidence from connectionist neuropsychology. *Journal of Clinical & Experimental Neuropsychology*, *17*(2), 291-321.

Postman, L., Bruner, J. S., & McGinnies, E. (1948) 'Personal values as selective factors in perception,' *Journal of Abnormal and Social Psychology* 43: 142–154.

Pylyshyn, Z. (1980) 'Computation and cognition: issues in the foundations of cognitive science,' *Behavioral and Brain Sciences* 3:111-132.

______(1999) 'Is vision continuous with cognition? The case for cognitive impenetrability of visual perception,' *Behavioral and Brain Sciences* 22 (3):341-365.

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-facesensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126* (Pt 11), 2381-2395.

Rumiati, R. I. & Humphreys, G. W. (1997). Visual object agnosia without alexia or prosopagnosia. *Visual Cognition*, *4*, 207-217.

Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge: Cambridge University Press.

Schubotz, R. I. & Fiebach, C. J. (2006). Integrative models of Broca's area and the ventral premotor cortex. *Cortex*, *42*, 461-463.

Siegel, S. (forthcoming) 'Cognitive Penetrability and Perceptual Justification,' *Nous*.

Sperber, D. (1994) 'The Modularity of Thought and the Epidemiology of Representations,' in L.A. Hirschfeld and S.A. Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture*, New York: Cambridge University Press.

______(1996) *Explaining Culture: A Naturalistic Approach*, Oxford: Blackwell.

______(2001) 'Defending massive modularity,' In E. Dupoux (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, MA: MIT Press

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociation prove? *Cognitive Science*, *25*, 111-172.

van Ulzen, N.R., Semin, G.R., Oudejans, R., Beek, P. (2008) 'Affective stimulus properties influence size perception and the Ebbinghaus illusion', *Psychological Research* 72: 304–310.

Wilson, R. (2008), 'The Drink You Have When You're Not Having a Drink,' *Mind and Language* 23: 273-83.

Young, A. W. (1996). Face recognition. In J. G. Beaumont, P. M. Kenealy, & M. J. C. Rogers (Eds.), *The Blackwell dictionary of neuropsychology*. Blackwell.