Modular architectures and informational encapsulation: A dilemma

Abstract

Amongst philosophers and cognitive scientists, modularity remains a popular choice for an architecture of the human mind, primarily because of the supposed explanatory value of this approach. Modular architectures 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 these architectures vary along these two dimensions. First, if a modular architecture commits to the *informational encapsulation* of modules, as it is the case for modularity theories of perception, then modules are on this account impenetrable. However, we argue that there are genuine cases of the cognitive penetrability of perception and that these cases challenge any strong, encapsulated modular architecture of perception. 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 at odds with the core of modularity. We then propose a non-modular notion of *functionally independent system* that, we argue, achieves the explanatory force sought by modularity theorists.

Amongst philosophers and cognitive scientists, modularity remains a popular choice for an architecture of the human mind. Jerry Fodor (1983), who was 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, involuntary responses to stimuli, and Fodorian modules are like 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, and that typically operate over a distinct domain of information.

There are compelling theoretical and empirical motivations for this approach. Theoretically, modularity nicely accommodates adaptationist and other evolutionary explanations of mental phenomena. It also provides materials for a simple explanation of important empirical data, including a wide range of behavioural dissociations, as well as the speed and robustness of processing enjoyed by the human mind. Most broadly, modularity provides an intuitive framework for characterizing the relations between brain structures and particular perceptual and cognitive functions.

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 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 (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 (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.

Our suggestion is that *informational encapsulation* is essential to a distinctive, non-trivial modularity theory. 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 (and so does not compute over) the information available elsewhere in the system (Pylyshyn 1980). 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, as made by modularity theories of perception, is inconsistent with the cognitive penetration of perceptual experience. And, we argue, there are genuine cases of the cognitive penetrability of perception. On the other hand,

---

[1] In at least two places, Fodor himself explicitly states that "informational encapsulation is an essential property of modular systems" (Fodor 1985: 3; see also 1983: 71). Elsewhere, however, he is less clear on his commitment regarding the same claim.

as recent modularity theorists have done, one might weaken the notion of module so as not to require 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 strong, *encapsulated modularity*: any modularity theory that 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, the modular approach to the study of cognitive architecture is significantly challenged.

In the concluding section, we argue that a non-modular notion of functionally independent systems can achieve much of the explanatory force originally sought by modularity theorists. This notion characterizes the specialization of the architectural units of the mind in terms of their cognitive *workings* rather than the cognitive or psychological *uses* to which they are put. This approach avoids the challenges to the modular approach, while allowing for robust functional decomposition and modeling of cognitive systems.

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

Both encapsulated and unencapsulated 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 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 modular input systems by arguing against claims about the cognitive penetrability of those 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 belief or desire. Evidence of this influence—that is, of the cognitive penetration of perception—thus threatens any modularity theory that includes a commitment to informational encapsulation of perceptual modules.

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 latter possess some kind of linguistic or propositional content, and play an immediate role in practical reasoning. 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. Philosophers debate 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 perceptual 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, some aspects of 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).

However, cognitive penetration of perceptual experience does 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 offered below.

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 alleged 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). Pylyshyn's empirical claim about early visual processing, even if true, is 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[2]

Although there is ample evidence for "top-down" processing in the brain—where information is exchanged between various areas of the cortex, including those areas believed to process higher-level or conceptual information—current neuroscience lacks an uncontroversial mapping from conceptual mental states (like belief) onto brain structures. And some such mapping would be necessary for neuroscience to provide a verdict on the actuality of cognitive penetration.[3] Consequently, empirical evidence for cognitive penetration must be obtained at the behavioural or psychological level, rather than merely the neurological level. Predictably, there are a number of possible alternative interpretations of this data, and so the inference structure is abductive. Critics of cognitive penetrability appeal to these alternative interpretations as better explaining alleged cases of cognitive penetration. We identify four such general skeptical strategies. With these strategies in hand, a working definition of cognitive penetration can be devised, in hopes of isolating the target phenomenon in a way agreeable to both sides of the debate.

First, for some experimental and/or anecdotal cases, 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 as a result of some other cognitive state, and report a memory of the stimulus rather than a perceptual experience. For instance, subjects might recall and report a stimulus to be a particular

---

[2] A number of critics have questioned Pylyshyn's conclusions in this general way (Bermudez 1999; Macpherson 2012; 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).

[3] Some have argued that evidence for reentrant neural pathways is evidence for cognitive penetration (Churchland 1988; 1989). Others have argued against this line of reasoning (Fodor 1988; Gilman 1991; Raftopoulos 2001). For purposes of this discussion, we simply assume that the neurological evidence is presently insufficient to count in either direction.

colour, where this memory causally depends upon beliefs about the stimulus type. 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*.[4]

A second strategy is the *attention-shift interpretation*. This interpretation maintains that in some of 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 action. This is no different in kind, critics urge, from an ordinary perceptual scenario where one, for example, has some belief about one's environment (e.g. that there is an irritating noise somewhere around here) and this belief causes some action (e.g. going to look for the thing making the irritating noise), which in turn results in a changed perceptual experience (e.g. seeing the squeaky faucet). This familiar cognitive-behavioural dynamic is important to everyday life, but unless cognitive penetration is trivially rampant, this is not cognitive penetration. Cases involving shifts of attention, this interpretation suggests, are to be treated similarly: these scenarios lack an appropriate internal connection, and so there is nothing in this causal chain to properly call 'penetration' (see Pylyshyn 1999: 343).[5]

Third, critics have suggested that 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

---

[4] For one example, see McCurdy 1956. Note also that if memory is *factive* as some have argued (Williamson 2002), then the memory-interpretation amounts to something like a *quasi*-memory interpretation.

[5] 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).

experimental subjects versus the control subjects in particular studies. Call this the *judgement interpretation*.[6]

Finally, an *intra-perceptual interpretation* claims that some of the evidenced effects are not cognitive ones but instead occur as adjustments or adaptations within the perceptual system. There are many ways to develop this alternative. One might claim that certain types of stimuli (for example, naturally occurring stimuli like ripe fruit) are such that the human perceptual system is appropriately "tuned" so as to represent these objects more quickly or in some enhanced way. These effects would not be learned, but would instead be artefacts of the evolution of human sensory systems. Or one might claim that a subject acquires a non-cognitive association with the stimulus type, and this association affects how perceptual information regarding tokens of this type are processed and/or how one acts in response to the stimulus. Or one may argue that sensory systems are sufficiently plastic to "learn" new ways to process sensory information, perhaps for some adaptive advantage.[7] Differences to one side, the thread common to these interpretations is that some cases may be better explained by changes in the sensory system that do not depend upon background cognitive states.

Grant that if any alleged 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 is 'yes', then the critics must secure some alternative interpretation to deflect the case/s. Here, following Stokes (2013) is such a definition:

---

[6] For additional discussion of these and other strategies for the cognitive impenetrability theorist, see Macpherson 2012; Stokes 2012, 2013.

[7] Against Churchland's appeal to subjects' adjustment to inverting lens as evidence for diachronic cognitive penetration, Fodor appeals to an intra-perceptual interpretation (see Fodor 1988: 193). For a more recent use of this kind of interpretation, see Deroy (2013), who analyzes some of the research also discussed below.

> *(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 and mental.*

The definition requires a few qualifications. First, 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.[8]

Second, (CP) excludes obviously non-genuine cases of cognitive penetration. For example, a desire to see the symphony, coupled with a true belief about the location of the symphony, may result in a perceptual experience of the symphony (several experiences in fact). But this should not count as an instance of cognitive penetration of experience, else the concept 'cognitive penetration' becomes trivial. (CP) delivers the appropriate result. 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) ensures, however, that this is not an instance of cognitive penetration, since the cognitive state (or states) is screened from being *internally*,

---

[8] The preferred notion of causation is of little matter so long as the internal causal dependence is maintained. One should also note that C is a non-sufficient cause of E. There are other relevant causal factors.

causally efficacious: the cognitive state causes an (external) action which eventually results in the experience.[9]

Importantly, a perceptual experience that satisfies this definition cannot be interpreted in any of the four ways described above. Clause (1) of (CP) rules out the memory, judgement, and intra-perceptual interpretation, since it requires a *cognitive* influence *on perception*, rather than just an influence on some other cognitive state in the system or an intra-sensory adjustment. 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 (CP)? We now consider three sets of studies that strongly suggest that the answer is 'yes'.

The first study was performed by Jerome Bruner and C.C. Goodman (1947). It is a pioneering study in at least two ways, it employed a methodology—involving *online perceptual tasks*, as will be explained below—which best isolates cognitive effects on perception. Second, it initiated a movement in psychology, the *New Look*. New look psychology was important both for its explicit opposition to its behaviourist predecessors—contra behaviourism, the New Look argued that the proper explananda of psychology include internal mental states and processes—and for the *way* it made that opposition—the New Look theorized perceptual experience as an active construction of representations of the environment. The thesis was essentially a universally quantified cognitive penetrability claim: (human) perceptual representations are always (or nearly always) constructed in a way informed by the perceiver's "mental set"—her expectations, needs, values, desires, and other higher-level mental states. Criticism, and eventual dismissal, of New Look targeted the strength of this claim: critics constructed a range of experiments where perception appeared unaffected by

---

[9] In this way, CP is consistent with other recently proposed definitions of cognitive penetrability: Macpherson 2012; Siegel 2011; Wu, forthcoming.

cognitive states. But a complete dismissal of New Look, like the universal claim upon which it was predicated, is overstated. Many of the New Look studies provide good evidence for an existentially quantified cognitive penetrability claim. And the claim of mere existential strength is the subject of today's debate.[10]

Bruner and Goodman's 1947 studies ran as follows. 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 American 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, who report the size of the cardboard analogues with near perfect accuracy.[11]

The second 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 poor children.

---

[10] See Balcetis and Dunning 2006, Stokes 2012, 2013, and van Ulzen 2008 for brief historical discussions of the rise and fall of the New Look movement, as well as (discussion of) new studies in the New Look spirit.

[11] 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%.

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.[12]

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 and mental.

Nonetheless, critics may, and have, resisted this as a genuine case of cognitive penetration, and by appeal to one of the strategies outlined above. But this rejection is a mistake premised on a failure to carefully consider Bruner and Goodman's experimental procedures. Importantly, the procedure for each experiment involved an online matching task. 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 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

---

[12] A number of theorists were 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). These critics were right to challenge the universally quantified New Look claim, and by simply acquiring evidence for cases where cognition apparently fails to affect perception. But, as suggested above, this evidence fails to undermine the more modest implication that cognition sometimes influences perception in the relevant ways. Moreover, the Bruner and Goodman 1947 results have been broadly replicated by a number of similar studies at least insofar as these studies all evidence some higher-level effect on perceptual experience. See Bruner and Postman 1948; Postman, Bruner, and McGinnies 1948; Bruner, Postman, and Rodrigues 1951; Dukes and Bevan 1952; Bruner and Rodrigues 1953; Bruner and Minturn 1955; Blum 1957; Holzkamp and Perlwitz 1966.

differed across control and experimental subjects (and so would fail to explain the relevant differences between the two groups).

The intra-perceptual interpretation, no matter how it is developed, simply does not apply. The experimental stimuli in question are artefactual and so are not stimuli for which we have an evolved perceptual sensitivity, nor a kind with which we are likely to form non-cognitive associations.

The judgement interpretation is most commonly used to deflect cases like 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 subject that 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 location six inches to the immediate left of the adjustable light patch—consistently makes erroneous judgements about what she is seeing. The reporting method here is key: since subjects report by adjusting the light patch, the judgement interpretation requires that subjects see the coin and light patch accurately (e.g. they see the light patch as significantly bigger than the dime at the time of concluding with a report, since this is the resulting data) but then judge, with both stimuli present, that the target and light patch are the same size.

Here, the cognitive impenetrability theorist might 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, and perhaps Bruner and Goodman's subjects can be explained similarly. 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 current 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 presently see (Bruner and Goodman take careful measures to adequately instruct the subjects). Indeed, if asked, the subjects would certainly confirm that their report—that the light patch matches the coin—is based on what they see. To treat these subjects like perceivers of the Müller-Lyer illusion requires that they are systematically mistaken about this: the subjects are not reporting on the basis of 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. Imagining the phenomenology of such a situation further reveals its implausibility. In the first experiment, the judgement interpretation requires that a subject have a veridical experience of, for example, a dime. And, simultaneously and upon inspection, she adjusts the matching patch of light to 129% of the (veridically) perceived coin (and, moreover, this light patch is also perceived veridically). And in the second experiment: a quarter is perceived (by "poor" subjects) veridically but, simultaneously and upon inspection, the light patch is over-adjusted to 150% of the (veridically) perceived coin. These are significant differences in size and so, given the online nature of the task, maintaining the judgement interpretation requires attributing significant cognitive error to the subjects. Moreover, what explains this error? The judgement interpretation must provide some answer here. The cognitive penetration interpretation, by contrast, only requires attribution of perceptual error (the effects of background cognitive states) that, in turn, explains the subjects' reports. Given the experimental circumstances, the latter interpretation is far more plausible; the judgement interpretation fails.

Crucial to disarming the alternative skeptical interpretations is the online nature of Bruner and Goodman's experimental procedure. Many recent studies in psychology are suggestive of cognitive penetration[13], but the most convincing studies follow some of Bruner and Goodman's basic methods. We now briefly present two such studies—the first on colour perception of natural and artificial objects, the second on the influence of racial categories on visual experience. Both studies involve relevantly controlled online perceptual tasks.

In a recent study, Thorsten Hansen and colleagues tested colour perception of objects with high "colour diagnositicity", objects the concepts of which are partly constituted by a distinctive colour concept: YELLOW for bananas, RED for strawberries, ORANGE for carrots, and so on (Hansen et al 2006). The procedure involved the presentation, on a computer monitor, of digital photographs of natural fruits/vegetables, presented in their typical colour, set against a uniformly grey background. The subject's task was to adjust the fruit image to what she judged to be a neutral (achromatic) grey. What in fact happens is that subjects adjust the image past achromatic grey and into the opponent colour range (e.g. adjusting a banana image past grey into the bluish hue). The researchers describe this as the *memory colour effect*. The researchers quantify this effect with a *memory colour index* (*MCI*), which in simplest terms provides a measure of the achromatic adjustment, towards the colour typical of the stimulus object (negative index) or away from it into the opponent hue range (positive index), relative to the

---

[13] For example, each of the following studies present data that may be plausibly explained in terms of cognitive penetration: Balcetis and Dunning 2006, 2010; Payne 2001, 2005; Stefanuci and Proffitt 2008, 2009; Witt and Dorsch 2009. However, the experimental controls in these studies are such that the results could also be plausibly explained in terms of one (or more) of the mentioned alternative interpretations.

typical colourfulness of the object. So for example, for a banana, the MCI is the ratio of the distance of shift past the perceiver's grey point into the bluish hue range *to* the distance of the shift from the typical yellow of the banana to the grey point, (with both of these distances measured along the same axis of typical adjustment for subjects).[14] For all of the experimental conditions, the MCI ranges from +4 to +13%, with a mean effect of +8.23%. As the researchers clarify, this quantification corresponds to an effect that is approximately three to five times above the threshold of discrimination. As a control, subjects perform the same task with uniformly coloured discs, and there is no memory colour effect: subjects adjust the discs to achromatic grey with perfect accuracy. We should emphasize that in this study (and those discussed immediately below), the task was clearly perceptual and online. Subjects took as much time as they felt necessary to make the adjustments, thus making adjustments to the perceptual stimuli in real time.

This case plausibly meets definition (CP). As the researchers hypothesize, a fruit/vegetable image, say a banana, still appears yellow to the subject at the point of achromatic grey. This hypothesis explains the fact that the subject adjusts the image into the bluish range, to compensate for the residual yellow, and then reports the fruit to be grey (when in fact it is slightly blue). This colour experience seems to depend, in a direct way, upon beliefs or conceptual associations with the relevant fruit/vegetable objects. Because the testing procedure involves online adjustment of the target stimuli itself, the memory interpretation is not appropriate. For similar reasons, the attention-shift interpretation fails: there is no plausible explanation whereby subjects, in

---

[14] More specifically, the researchers clarify the calculation of the MCI as it is used in all three of the colour perception studies discussed here, as follows. "For the MCI the achromatic adjustments are projected on the axis of the typical adjustments that leads through the subjective grey point. The distance of this projection from the subjective grey point measures how strong the shift along this axis was. For the MCI this measure is divided by the length, i.e. the saturation, of the typical adjustment. In this way, the MCI represents the ratio of achromatic shift relative to the colourfulness of the typical colour. The sign (+/–) of the MCI reflects the direction in which the adjustment is shifted away from the subjective grey point. A positive MCI indicates an achromatic adjustment opposite to the typical adjustment. A negative MCI implies, contrary to the memory colour effect, that there is a shift of the achromatic adjustments towards the same direction as the typical adjustments. The MCI has been calculated separately for each participant using their subjective grey point" (Witzel et al 2011: 37).

experimental but *not* control conditions, execute overt (or covert) attention to get the relevant effects.[15] And the judgement interpretation would require that, as the subject visually inspects and adjusts the target stimulus, she veridically perceives the stimulus (e.g. a banana image as slightly blue) but then reports a judgement that it is perfectly grey. So she sees the stimulus accurately but reports it erroneously. And this error has to be explained in a way that current (veridical) perception is bypassed or ignored as informing the subject's report, in spite of the task being an explicitly perceptual one. This looks much less plausible than an explanation where a non-veridical experience, itself causally dependent on background cognitive states, causes a judgement and report that the target is perfect grey (when it, the banana image for example, is in fact objectively, slightly blue). Here the report is erroneous—as the data make clear—but the error in report is explained by perceptual error, and the perceptual error is explained by cognitive penetration.

However, what about the intra-perceptual interpretation? One might worry, that since the target stimuli are all natural objects, the memory colour effect is symptomatic of hard-wired sensitivities of the human perceptual system. An enhanced perceptual sensitivity to ripe fruit and vegetables would plausibly be an evolutionary advantage for humans. And so granting that subjects still see the banana image as slightly yellow even when it is objectively grey, one might argue that this is best explained by facts about human perceptual processing and how it has evolved, without any needed appeal to cognition.

This interpretation may appear even more plausible in the light of a second study performed by some of the same researchers, where the memory colour effect was most pronounced for realistic images of fruits/vegetables (e.g. those depicting texture) and

---

[15] This is by contrast, for example, with Fodor's favoured explanation of the way one can shift, by attentional changes, one's experience of the Necker cube or the duck-rabbit. See Fodor 1988: 190.

mostly absent for mere fruit/vegetable outline shapes (Olkonnen et al 2008).[16] However, this interpretation is easily dispelled by a very recent study (Witzel et al 2011). These studies involve artificial, human-made objects as stimuli. In a preliminary study, the researchers identify artificial objects with maximal *colour diagnosticity*, the blue Smurf, the Pink Panther, the red Coca-Cola logo, a green ping pong table, and so on. Images of these objects are then included in an experiment where the task is the same as the above two studies, plus a few additional controls. Target objects are initially presented in a random colour (e.g. a fire extinguisher might appear as blue rather than its typical red colour) against a uniformly grey background, and subjects then adjust the object to what they perceive to be achromatic grey. Additionally, control objects that typically vary in colour (e.g. a sock) and control objects that are typically achromatic (e.g. a golf ball) are presented in a random colour where the task is the same. Under these conditions, there is no evident effect for control objects, and a significant effect for colour diagnostic objects. The mean MCI for fourteen colour diagnostic stimuli was +3.31%, with a high of +10.3% (for the blue Nivea tin). Just as in the earlier studies, the results provide strong evidence for a cognitive effect on perceptual experience. And importantly, the intra-perceptual interpretation is not applicable to this most recent study: there is no story to be told about the evolution or plasticity of perceptual systems for the perception of cartoon icons or soda logos.[17]

---

[16] See Deroy 2103 for an analysis that partly focuses on the Olkonnen et al 2008 study.

[17] It is worth adding that this effect is apparently culture-sensitive. In their initial study to identify colour diagnosticity for artificial objects, which was performed in Germany, Witzel et al (2011) found that some stereotypically German images were highly colour diagnostic (as measured by reaction time and accuracy of typical colour identification)—for example the orange Die Maus (a German television character), the yellow German mailbox, the yellow (German-made) UHU glue tube. But some non-German objects were not sufficiently colour diagnostic (relative to German subjects)—for example, the yellow Ferrari symbol and the red Soviet flag. These researchers did not run the study using these non-colour diagnostic (relative to German subjects) objects, but presumably if they had, any memory colour effect would have been insignificant at best. This factor lends additional plausibility to the effect being a cognitive one: subjects learn (and sometimes differently in different places) what colours are typical of artificial objects, and these beliefs or conceptual associations affect colour experience of (images of) those objects.

Finally, consider a recent study on racial stereotypes and face perception (Levin and Banaji 2006). In Experiment 1, subjects were presented, on a computer monitor, with realistic greyscale images of male faces with features stereotypical of either black or white persons (with hair removed). The task was to match the luminance of an adjustable greyscale face to the target face (in some conditions the adjustable face was of the same racial prototype as the target, and in others of the opposite racial prototype). Although the luminance of the two target prototypes was (objectively) identical, subjects consistently adjusted to a lighter grey for the stereotypical white faces and to a darker grey for the stereotypical black faces (in both mixed-race and same-race conditions).[18] Now, one might worry that in this experiment, the results are mere effects of optics. For example, in spite of both prototypes being identically greyscale, the contours of typical black facial features versus typical white facial features vary such that, one might conjecture, the first type of face naturally looks darker than the second type of face. Or one might attempt to invoke some variation of the intra-perceptual interpretation to deny the cognitive influence of racial stereotypes. These interpretations, however, fail to apply to a second variation of the study, one that in fact yields even more striking results.

In Experiment 2, Levin and Banaji first, in a preliminary study, created a racially ambiguous face by morphing a range of prototypical black and white-face features, and then confirmed the ambiguity of the face by appeal to racial classification results across 15 subjects. On an instruction screen, the ambiguous face (call this 'BW') was then paired with either an unambiguously white face (call this 'W') or an unambiguously black face (call this 'B'). And in each condition, both faces were labelled, either 'Black' or 'White'. So for example, when paired with an unambiguously white face, the

---

[18] More specifically, for example, with a black-face prototype as target, subjects adjusted a white-face prototype to 4.65 levels darker (out of 256 possible greyscale levels for the computer monitor) than a white-face prototype target (where, again, both targets are of identical luminance levels).

ambiguous face (BW) was labelled 'Black' and the unambiguous white face (W), labelled 'White'. Taking this condition as our example—Levin and and Banaji call this the "BW/W condition"—the task phase proceeded as follows. Subjects were presented with a series of trials, where each trial involved either the ambiguous face (i.e. the one labelled 'Black' in the instruction phase of the BW/W condition) or the unambiguous white face (labelled 'White' in the instruction phase), both of identical luminance, coupled with an adjustable rectangular region of uniform grey. The task in each trial was to adjust the grey report patch to match the face simultaneously perceived. Result: the racially ambiguous face is reported in a way that strongly correlates with the semantic labelling prime. So, in the BW/W condition, the lightness report for the ambiguous ('black'-labelled BW) face was .465 levels darker (than the objective luminance of the target) and 17.85 levels lighter for the unambiguous ('white'-labelled W) white face. And here is perhaps the most striking result: when the same ambiguous face BW is labelled 'White' (in the opposite "B/BW condition") the report for BW is 15.95 levels lighter. So: present a face identical both with respect to luminance and facial features, but change the label from 'Black' to 'White', and the reported match goes from .465 levels darker to 15.95 levels lighter (than the objective luminance of the ambiguous face)(2006: 505-6).[19] This is not an effect explained just by optics or (intra-) perceptual features; the linguistic label is clearly playing an operative role in the subject's perceptual experience.

To conclude discussion of this final set of studies, consider the remaining alternative interpretations. The Levin and Banaji results are not well-explained by memory since this study involves online perceptual matching tasks. Nor is there any reason to think that overt shifts in attention are explanatory of the effects.[20] Thus both the memory and

---

[19] See footnote 18 for clarification regarding the grey measures.

[20] In fact, the researchers devise a third experiment explicitly devoted to discounting an explanation where attention is drawn to facial contours (e.g. of the stereotypical black face) in a way that explains the perceptual differences that appear in the results. They construct greyscale line drawings—with either white lines or black

attention-shift interpretations fail to apply. What about the judgement interpretation? Considering Experiment 2, this interpretation would require that a subject, upon confirming her report and moving to another trial, has veridical experiences of the target face and the report region of grey—for example, where the grey report region would appear as (on average) 15.95 levels lighter than the simultaneously perceived ('white'-labelled) ambiguous face. But then somehow the subject, in spite of perceiving this difference, judges and reports the face and report region as matching. This is far less plausible than the opposing cognitive penetrability thesis. The best explanation, here and above, is that the subject is having a non-veridical experience. She sees the prototypical white and prototypical black face as lighter and darker, respectively, and in Experiment 2, this effect is exaggerated by a linguistic labelling prime. The non-veridical experience is a result of penetrating cognitive states, in this case, racial stereotypes or beliefs.[21]

Summarizing, there are crucial methodological features common to the three sets of studies discussed above. First, in all of these studies, subjects must perform online perceptual tasks. Thus the target stimulus—a coin, a Smurf image, a greyscale face—is present and perceivable while the subject makes her report. Use of this methodology disarms the memory and attention-shift interpretations. Second, the method for reporting involves some direct kind of manipulation, either of the target stimulus itself or of some match disc or region. It is this methodology that, coupled with the online methodology, disarms the judgment interpretation. (Compare: Many other experiments use verbal reports of some kind. And a task involving a verbal report—for example, providing a numerical estimate of the distance of a perceived object—opens space for judgement about perception and, in turn, encourages the judgement interpretation.)

---

lines providing the facial outlines, but with no other shading of facial features—of the white and black prototypes. The results are relevantly the same and statistically significant: subjects choose darker samples for the black prototype faces and lighter samples for the white prototype faces. See Levin and Banaji 2006: 506-8.

[21] Macpherson 2012 briefly discusses both Hansen et al 2006 and Levin and Banaji 2006. She also provides a detailed analysis of an earlier study on colour perception, Delk and Filenbaum 1965.

Finally, the stimuli used in these studies are all ones about which we learn and form beliefs, desires, and other cognitive states. It is this methodology that, at least partly, disarms the intra-perceptual interpretation. As a point of methodology, we prescribe that any experimental attempts to test for cognitive penetration should employ, at minimum, this combination of features. And this approach can be traced back to Bruner and Goodman’s important work of over 60 years ago.

Now, what does all of this imply for a strong modularity theory, any theory that commits to informationally encapsulated perceptual modules? If the above discussion is successful, then the standard alternative strategies fail to deflect the discussed cases as genuine evidence for the cognitive penetration of perception. In each case, whether it is a desire, value, belief, or some other higher-level mental state, there is evidently *some* cognitive state (internally) influencing experience. These cases are best described as meeting the conditions of (CP). And therefore, as we will now argue, perceptual systems are not informationally encapsulated.

Recall that perceptual systems are paradigms for modular systems. And recall further that if one is an encapsulated 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. Therefore, any legitimate case of the cognitive penetration of experience undermines the alleged informational encapsulation of the relevant perceptual systems, and in turn challenges any theoretical architecture of those systems that commits to informational encapsulation 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. We have defended three sets of studies against the relevant alternative interpretations. And so perceptual systems—in these cases *vision*—are not informationally encapsulated. Any modularity hypothesis about the architecture of perceptual systems that commits to the necessity of informational encapsulation (and by implication: cognitive impenetrability) for modularity is therefore threatened.

To clarify our critique, it will be useful to briefly consider a hypothetical defence for the encapsulated modularity theorist 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, which means that some components of perceptual systems may be strongly modular.[22]

To this, the modularity theorist might add that the modularity approach is a conceptual framework for modelling parts of the mind. As such the framework is valuable if and to the degree that it usefully explains *some* features of the human mind. And, the above critique notwithstanding, hypothesizing encapsulated modules does successfully explain significant aspects of perceptual processing. Consequently, cases of cognitive penetration of perceptual experience do not pose a serious threat to modularity understood in this way.[23]

Reply: the fact that some aspects of perceptual processing can be explained by encapsulated modularity does nothing to save a modular *architecture* of perception. 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-

---

[22] We thank XXXX for pressing us to consider this reply for modularity theory.

[23] We thank XXXX for pressing us to consider this response.

level components in the overall visual system. In fact, it may be 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 informational encapsulation and thus a *strong* notion of 'module'. But note that the *scope* of modularity on such a view, that is, the kinds of systems to which the conceptual framework can be successfully applied, 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.

It is important, moreover, not to overemphasize a characterization of (perceptual) modularity as concerning only perceptual processing. This is because an interest in mental architecture is guided not merely by goals of psychological modelling. Another crucial issue of relevance is epistemic: a modularity theory of perception promises a preferable epistemology, one where perceptual systems rapidly deliver perceptual representations in a way not prone (or less prone) to errors introduced by the cognitive agent. As Fodor puts the point, the "function of perception is to deliver to thought a *representation* of the world." And since here the goal is to represent "[n]ot the distant past, not the distant future and not…what is very far away…it is understandable that *perception* should be performed by fast, mandatory, encapsulated, etc. systems…" (Fodor 1985: 5; emphasis added). The systems in question are sub-personal modules, but the representations they provide or give rise to are personal-level experiences. Given the epistemic role that such representations are supposed to serve, and the supposed epistemic advantage of modular perceptual systems, the modularity theorist should be no happier with evidence for penetrated experience than he is with evidence for unencapsulated perceptual processing.[24]

Where does this leave the view? The claim that some individual low-level circuits are encapsulated and thus strongly modular is largely uncontroversial among cognitive scientists. And the claim that some sub-systems 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 turn, these weakened claims are insufficient to secure the putative epistemic benefit of modular perceptual systems. In short, one cannot save a modular architecture of perception by appeal to encapsulated perceptual components or sub-systems. To do so would be to opt for strength of modules over scope, in turn undermining the modularity hypothesis as an *architecture* of perceptual systems.

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

A number of recent theorists have 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 tends 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, informational encapsulation is too strict a requirement on the modularity of perception,

---

[24] Fodor makes similar suggestions elsewhere; see, for example, his discussion of "perceptual identifications" (1983: 68-71). And Pylyshyn (1980) makes similar commitments, claiming that the reliability of perception requires cognitive impenetrability. For further discussion of the epistemic consequences of cognitive penetrability, see Lyons 2011; Siegel 2012, 2013; Stokes 2012, 2013.

then it makes sense to not require it of higher-level cognitive systems. Weakening modularity in this way, however, comes with significant costs to any modular account of cognition. First, it weakens the explanatory value of modular architectures. Second, it threatens the internal coherence of modularity 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 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, proposing modular architectures is that they explain behavioural dissociations between cognitive functions. A cognitive task *A* is said to be dissociated from cognitive task *B* when 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 in *A*. Cognitive scientists generally hold that dissociations are signs of functional independence, and will often hypothesize the existence of cognitive modules on the basis of these behavioural patterns. If *A* is observed to fail when *B* does not, then one may infer that *A* involves a subsystem, or module, *M*, that *B* does not recruit. When *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* each involves a subsystem, 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. In this case, a double dissociation between face recognition and visual object recognition—i.e. observing patients with intact visual objects recognition but impaired face recognition, and patients with intact face recognition but impaired visual object recognition—suggests that the system used to recognize faces is not identical to the system used to recognize objects, and that each of the two systems have at least one subsystem, or module, that the other doesn't have (Coltheart 1999). This, however, does not mean that the face recognition and visual object recognition systems are functionally independent from each other, since they evidently share some of their subsystems (e.g. the subsystem responsible for low-level visual feature analysis). Rather, what the double dissociation suggests, in this case, is that each system is functionally independent from at least one of the subsystems, or modules, of the other system—i.e. there is a module used for face recognition that plays no role in object recognition and there is a module for object recognition that plays no role in face recognition.

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, and there is an emerging consensus that the inference should be understood as an inference to the best explanation, where one infers the existence of cognitive modules on the grounds that this hypothesis best explains a set of dissociation data, since the hypothesized modules would produce the relevant dissociations if they were damaged separately (Shallice 1988, Coltheart 2001).[25] The status of the inference as a central methodological tool, however, is very much a matter of debate. Various authors have argued, on both theoretical and empirical grounds, that

---

[25] See Shallice 1988 for a detailed discussion of this methodology.

the existence of a double dissociation between subjects' performance 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; Machery 2012; Plaut 1995; Van Orden, Pennington, & Stone 2001). We take no side in this debate. Instead, we question whether an inference to the best explanation is supported *when modules are assumed to be unencapsulated*.

Let us suppose, then, as the weakened modularity theory we're considering does, that modules are *not* encapsulated. Suppose, for example, that the above-mentioned face and object recognition modules are not encapsulated, that they both often need to compute over information made available by other systems in order to perform their tasks. This means that a double dissociation between face and object recognition could occur *even if* both alleged modules remained intact (i.e. were not damaged). This would occur, for instance, if both modules need to compute over information normally made available by other systems and damage to these other systems (or damage to some pathways between them and the two modules) prevents the availability of the needed information. In this case, the double dissociation could no longer be taken as strong evidence that the two modules are functionally independent (i.e. that they can be damaged separately), as both modules could fail to perform their tasks for reasons that have nothing to do with a failure of their respective mechanisms. But since it is the assumed functional independence of cognitive modules, in the functional modularity inference, that is supposed to explain the existence of dissociation data, it is therefore difficult to see how the face and object recognition module hypotheses could, in this case, *best* explain the observed double dissociation between face and object recognition.[26]

---

[26] The alleged theory of mind module is another case, one where unencapsulated modularity would fail to explain behavioural dissociations between mind reading and other cognitive capacities. See Gerrans and Stone (2008) for a discussion of this case.

By contrast, a double dissociation between face and object recognition is adequately explained by encapsulated modularity. Importantly, because the face and object recognition modules would in this case be informationally encapsulated, their normal functioning would thereby not depend on information made available by other systems. A double dissociation between face and object recognition would thus suggest that the modules themselves have been separately damaged.[27]

This point about the explanatory weakness of unencapsulated modularity is worth further emphasis, since it suggests that there is conceptual tension between the notions of unencapsulation and modularity.

To see this, consider a minimal conception of modularity. As Carruthers suggests, a module must be, at the very least, a "dissociable functional component" (Carruthers 2006: 2). This minimal conception of modularity is what gives the functional modularity inference its theoretical force. Behavioural dissociations are explained by modular architectures—and are thus signs of functional independence between cognitive mechanisms—since dissociability (at the level of mechanisms) *implies* functional independence and modules are dissociable systems. We agree that a minimal notion of modularity should include dissociability, since without it a modular architecture would reduce to functional decomposition. And functional decomposition—understanding the mind in terms of functional components and sub-components—is uncontroversial as an approach, except perhaps in some connectionist quarters.

On the one hand, therefore, a cognitive system *S* is considered functionally independent from another system *O* if *S* and *O*'s function can be dissociated. This means that *S*'s function is not affected by what happens to *O*, and that *S* can be modified (or damaged) without affecting *O*'s function.

---

[27] This is not to say, however, that such an encapsulated modular architecture is the only plausible explanation of the double dissociation between face and object recognition. Even with encapsulated modularity, the functional modularity inference remains abductive (Shallice 1988).

On the other hand, *S* is considered unencapsulated relative to *O* if *S* needs to compute over information made available by *O* in order to perform its task. This means that *S* depends on *O* for its normal functioning. *S* is thus functionally dependent on *O*, and is therefore *not* dissociable from *O* (since dissociability implies functional independence). In sum, if functional independence is understood in terms of dissociation, as the minimal conception of modularity suggests, then *S* cannot both depend on information provided by *O* and be dissociable from *O*.

An illustration may help. Carruthers explains the minimal conception of modularity with the following analogy.

> 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 autonomous: the CD player may require the amp to deliver its output, but it does not need to compute over information made available by the amp in reading data off of a

CD. In other words, in performing its task, 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 analogy 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 together, exchanging input and output, and often asymmetrically. But unlike the hi-fi components, an unencapsulated module *M*, as per the massive modularity theorist, will often depend for its normal operation on other components in the system. And this means that *M* will not, in these cases, be dissociable from these other components. Here again *M* cannot both depend on information provided by other systems and be dissociable from them.

This, finally, is the second horn of our proposed dilemma, which challenges modularity theorists that expand the scope of modularity *by* weakening the strength of modules so as not to require informational encapsulation. Weakening modularity to this degree weakens the explanatory value of modular architectures, which in turn weakens the functional modularity inference. Second, the very notion of an unencapsulated module appears to be at odds with the core of modularity: conceptualizing modules as dissociable functional components. Recall further that this weakened modularity may be partly motivated by—in addition to broadened scope—acknowledgement of the apparent failure of encapsulated modularity to explain various perceptual phenomena. This was the first horn of our dilemma: perceptual systems are not encapsulated modules if perception is cognitively penetrated. And there is compelling empirical evidence for phenomena best explained by cognitive penetration. This concludes our proposed dilemma for modular architectures of the mind. Narrow scope-encapsulated modularity theories are challenged by the first horn, broad scope-unencapsulated modularity theories, by the second horn.

III. Functional independence without modularity

Over the past century and a half, a large body of neuropsychological data, primarily in the form of dissociation data, indicates that there are specialized neural circuits in the brain and that there are stable relations between these circuits and particular cognitive functions.[28] In fact, modular and non-modular theorists alike see specialization within the brain as an undisputed fact. Both sides, therefore, would agree in substance with 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). What is at issue is how best to characterize this specialization.

Fodor, as discussed, characterizes the specialization of perceptual systems within the strong (encapsulated) modularity framework. According to this approach, individual brain areas can be ascribed specific perceptual functions when brain areas constitute "domain-specific computational systems characterized by informational encapsulation, high speed, restricted access, neural specificity, and the rest." (Fodor 1983: 101). It is when brain areas can be characterized in this way that, according to Fodor, we should expect to find stable (i.e. lawful) relations between structure and function. Or, to put it differently, we should expect to find stable relations between particular brain areas and specific cognitive functions when brain areas can perform their computational functions independently of other brain areas (by virtue of being encapsulated).

By contrast, Fodor was much less optimistic about the prospect of finding stable

---

[28] More recently, functional neuroimaging data (e.g. fMRI, PET), in the form of selective activations of brain areas for certain tasks, also point to the wide range of specialized neural circuits in the brain, although the methodology in this case differs from the standard behavioral dissociation logic in neuropsychology.

structure-function relations in the case of *un*encapsulated computational systems.

> Consider, by contrast, [unencapsulated] systems, where more or less any subsystem may want to talk to any other at more or less any time. In this case, you'd expect the neuroanatomy to be relatively diffuse. At the limit, you might as well have a random net, with each computational subsystem connected, directly or indirectly, with every other; a kind of wiring in which you get a minimum of stable correspondence between neuroanatomical form and psychological function. (Fodor, 1983: 118).

On Fodor's model, you get neural specificity, and thus stable structure-function relations, only when cognitive systems perform their computations autonomously and locally. In the case of unencapsulated systems, the computational and informational resources needed to perform the task at hand are distributed across a wide range of systems, which is why we should not expect to find stable relations between unencapsulated computational systems and specific cognitive functions.

In the previous section, we argued that there is tension between the minimal conception of modules as dissociable functional components and the idea that modules can be unencapsulated. The argument was that unencapsulated systems are not functionally independent. And since dissociability implies functional independence, unencapsulated systems cannot be dissociable.

These claims – Fodor's claim that stable structure-function relations require encapsulation and our claim that modules cannot be both unencapsulated and dissociable – are intimately related. Both stable structure-function relations and dissociability, it seems, require functional independence, but unencapsulated systems, as we saw, are not functionally independent.

We now argue that functional independence can be defined in such a way that unencapsulated systems are functionally independent. We then argue that the resulting notion of functionally independent system, while considerably weaker than the minimal conception of a module, is sufficiently strong to characterize stable structure-function relations.

To see how a system *S* can be functionally independent from another system *O* without being encapsulated from *O*, consider the case of Broca's area (BA). This area is involved in language processing (production and perception of speech), and in order to contribute to this cognitive capacity it needs to compute over information that is processed and made available by other areas, one of which is the superior temporal sulcus (STS) which processes and stores phonological representations (Hickok & Poeppel 2007). BA is thus unencapsulated relative to the STS. (Notice also that since BA's contribution to language processing would be affected if the STS were damaged, it is therefore not dissociable from the STS.) Nevertheless, there is a sense in which BA is functionally independent from the STS.

To see this, consider the distinction between the low-level computational operations, or "workings", performed by BA, and the higher-level cognitive "uses" to which it is put (Bergeron 2007, Anderson, 2010). We know that BA is put to a number of linguistic and non-linguistic uses—for example, it is involved in both musical and linguistic syntactic processing, in object manipulation, and in action sequencing and action perception (Maess *et al*. 2001, Nishitani et al. 2005). This, in turn, has been interpreted as evidence that BA's contribution to these various cognitive uses could be performed by a "reusable" set of low-level computational operations, or workings—for instance, sequencing operations on a wide range of inputs (Fiebach & Schubotz 2006), or the processing (detection, extraction) of hierarchical structures in a wide range of cognitive domains (Tettamanti & Weniger 2006). The important lesson is this: BA's "function" can be interpreted in two different ways depending on whether one is referring to its local workings or to its higher-level cognitive uses.

In the light of this distinction, we can now specify the sense in which BA is functionally independent from the STS. BA is functionally independent from the STS *with respect to sequencing operations* in the sense that BA performs these operations and has the capacity to perform them even if the STS failed to compute anything. Generalizing:

> (FI) A system *S* is *functionally independent* from another system *O with respect to working W*, iff *S* performs *W* and has the capacity to perform *W* even if *O* failed to compute anything.[29]

And accordingly:

> (FI system) A system *S* is an *FI system with respect to working W*, iff *S* performs *W* and has the capacity to perform *W* even if no other systems computed anything.

A few points are in order. First, to say that *S* has the *capacity to perform W* even if another system *O* failed to compute anything is to say that *S* possesses the right kind of machinery to perform *W* given that it is provided with the right kind of information. So, the second conjunct of the condition for functional independence must be read counterfactually.

In the case in which *S* performs *W* over information provided by *O*, this means that *S*

---

[29] This formulation is adapted from the analysis of isolability provided by Lyons (2001): "A substrate *S* is *isolable with respect to task T* iff *S* performs task *T* and could do so even if nothing else computed any (cognitive) functions"(289).

has the right kind of machinery to perform *W* over the kind of information that *O* normally provides. For example, if we assume that BA does in fact perform sequencing operations over phonological representations made available by the STS, then to say that BA has the capacity to perform sequencing operations even if the STS failed to compute anything (and thus failed to make available the relevant information) is to say that BA has the right kind of machinery to perform these operations over the kind of information that the STS *normally* provides.

Second, the notion of FI system is considerably weaker than the *minimal* notion of a module as a dissociable functional component. This is because FI is considerably weaker than the traditional definition of functional independence in terms of dissociation. Indeed, functional dissociability implies FI but FI does not imply functional dissociability—recall that BA is functionally independent from the STS (in FI's sense) even though it is not dissociable from the STS, since its contribution to speech production would be affected if the STS were damaged. In fact, the general characterization of cognitive components as FI systems cannot, in any non-trivial way, count as modular. To see this, suppose that modules are no more than FI systems or networks of FI systems. Suppose, for example, that BA plus the STS, plus some other systems constitute the speech production module. This is to *assemble* a module out of FI systems: decomposing the mind into cognitive capacities and attributing these capacities to an assemblage of systems. By this method, any identifiable cognitive capacity could be turned, trivially, into a module, since any identifiable cognitive capacity could, under our proposal, be implemented by an FI system or a network of FI systems. What is trivial is not that the proposed notion of FI system could potentially be applied to the functional decomposition of any identifiable cognitive capacity: this is exactly what a general notion of cognitive component should do. What's trivial, instead, is calling any FI system or network of FI systems capable of implementing a cognitive capacity a ‘module’; labeling each such cognitive component a ‘module’ adds nothing of theoretical import to the functional decomposition approach in cognitive science.

There is, in addition, another aspect of FI that makes it weaker than the minimal conception of modularity. In order for a system *S* to be functionally independent from another system *O* (according to FI), the requisite condition applies only to one of the two levels of functional specification—i.e. it applies to workings (at the local level), but *not* to cognitive uses (at the higher systemic level).

For example, assuming again that BA does in fact perform sequencing operations over phonological representations made available by the STS, BA is functionally independent from the STS *with respect to its sequencing operations* (local workings), but it is *not* functionally independent from the STS *with respect to speech production* since both areas are jointly put to this cognitive *use*. To see this, consider a modification of (FI) where the working/use distinction is not made. Replacing ‘working’ with the more general term ‘function’ we have:

> (FI*) A system *S* is *functionally independent* from another system *O with respect to function F*, iff *S* performs *F* and is capable of *F*ing even if *O* failed to compute anything.

Substituting ‘BA’ for *S* and ‘STS’ for *O* we have:

> BA is *functionally independent* from the STS *with respect to function F*, iff BA performs *F* and is capable of *F*ing even if the STS failed to compute anything.

In this case, the condition is satisfied if ‘*F*’ refers to sequencing operations, but it is not satisfied if ‘*F*’ refers to speech production.

Importantly, the distinction between these two levels of functional specification is rarely operative in modular theorizing. For example, neuropsychologists typically attribute cognitive *uses* (as opposed to workings) to brain areas on the basis of dissociation data. Thus, Broca's area has been characterized as a speech production module; the temporoparietal junction has been characterized as a theory of mind (or belief attribution) module; the fusiform face area (FFA) has been characterized as a face recognition module, and so on.

This, in fact, should not be surprising since the functional modularity inference rests on behavioral data that are derived from an analysis of the performance of brain-damaged patients on various cognitive tasks. As such, these data consist of the specification of the *behavioural consequences* of the (mal)workings of various cognitive components, which means that they will naturally be expressed in ways that capture one or more of the cognitive *uses* of these components within the larger cognitive economy.

Given that this is how neuropsychological functional specification has typically been carried out (i.e. characterizing the specialization of a particular brain area in terms of the its cognitive uses), Fodor's pessimism about the prospect of finding stable relations between unencapsulated systems and specific cognitive functions is understandable. Indeed, this form of functional specification amounts to characterizing structure-*use* relations, but unencapsulated systems, as Fodor pointed out, will tend to have a wide range of cognitive uses across different domains. The picture is quite different, however, if we conceive of unencapsulated systems as FI systems. FI systems are characterized as structure-*working* relations—they characterize the specialization of brain areas in terms of their cognitive workings. And since the workings of an FI system are both functionally independent (in FI's sense) and stable across its different cognitive uses, this form of functional specification allows for the characterization of stable relations between unencapsulated systems and their workings.

Unencapsulation, therefore, is not an obstacle to the identification of stable structure-function relations (contra Fodor), as long as the target is structure-working, as opposed to structure-use, relations. It is the functional independence of a system (in FI's sense), not encapsulation, which determines whether a stable relationship can be established between that system and a particular function. So the cognitive penetration of perceptual experience is also not an obstacle to the functional decomposition and modeling of (unencapsulated) perceptual systems.

This is not to say, however, that the functional specification of Fodorian modules and unencapsulated FI systems can be approached in the same way. Besides encapsulation, these two kinds of systems will typically differ in another important way. Fodorian modules, because they are encapsulated, are constrained on the range of information they can access in the course of their processing. They perform their computations over a restricted class of inputs, which in the case of perceptual systems is a narrow range of distal properties (e.g. colours, faces). The specification of Fodorian modules will thus typically involve domain-specific (or content-specific) structure-function relations (e.g. colour-processing and face-recognition modules). The specification of unencapsulated FI systems, by contrast, will tend to involve domain-neutral (or content-neutral) structure-function relations, since the workings of these systems will typically contribute to a wide range of cognitive uses across different informational domains.

Modular theorists of all stripes have considered domain-specific functional specification to be one of the major strengths of modular theorizing. Knowing, for instance, that a particular area of the brain specializes in face-recognition, or belief attribution, would tell us something quite specific about how these capacities are effected within the larger cognitive economy. Now, what our discussion of FI systems indicates is that domain-neutral specification can be equally informative about the organization of cognition, but in a different way. The specification of domain-neutral structure-working relations can account for the contribution of a given brain structure to a wide range of (sometimes seemingly unrelated) cognitive capacities. And by the same token, this kind of functional specification can expose common structural and computational principles underlying seemingly unrelated capacities. For example, there is something quite informative about the fact that there appears to be common cognitive mechanisms for object manipulation, action sequencing, musical processing, and speech, and that these mechanisms are performed by a set of reusable neural structures (in Broca's area). In fact, the reuse of neural circuitry for various cognitive purposes appears to be a central organizational principle of the brain[30], which means that the specification of reusable FI systems may have wide-ranging implications for the study of cognitive architecture.

We therefore agree with both Fodor and massive modularity theorists with respect to the broad explanandum: specialization within the mind (and the brain) in the form of stable relations between brain structures and particular cognitive functions. We have argued, however, that both encapsulated (Fodorian) and unencapsulated modularity are inadequate as approaches to the study of cognitive architecture. And here we have diagnosed the inadequacy of the modular explanans in part by re-focusing the broad explanandum to structure-*working* relations.

Encapsulated modularity appears incapable of explaining well-evidenced cognitive-perceptual relations, namely, cognitive effects on perceptual experience—and thus processing—supposed to be incompatible with encapsulated perceptual systems. Unencapsulated modularity is both too weak to adequately explain what it is supposed to explain best, and is at odds with the core conception of modularity. This is the

---

[30] Anderson (2010) provides an integrated review of recent neural reuse theories. In particular, see Vittorio Gallese's "neural exploitation hypothesis" (2008), Susan Hurley's "shared circuits model" (2008), Stanislas Dehaene's "neuronal recycling" theory (2005) and Michael Anderson's "massive redeployment" hypothesis (2007).

dilemma for modularity architectures of the mind. Cognitive specialization, we suggest, would be better and more generally explained by a different framework, one based on non-modular functionally independent systems.

Acknowledgements

References

Anderson, M.L. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology*, 21(2): 143-174.

Anderson, M. (2010) 'Neural reuse: A fundamental organizational principle of the brain,' *Behavioural and Brain Sciences*: 33: 245-266.

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.

Bergeron, V. (2007) Anatomical and functional modularity in cognitive science: shifting the focus. *Philosophical Psychology* 20(2): 175-195.

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

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.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The'neuronal recycling' hypothesis. In: S. Dehaene, J. Duhamel, M.D. Hauser and G. Rizzolatti (eds). *From Monkey Brain to Human Brain*. Cambridge, MA: MIT Press.

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.

Deroy, O. (2013) 'Phenomenal contrast without the cognitive penetrability of perception,' *Philosophical Studies*. 162, 87-107

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,' *Behavioural and Brain Sciences* 17: 43-104.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J. and Gibson, T. (2009) Structural intergration in language and music: Evidence for a shared system. Memory and Cognition 37(1): 1-9.

Fiebach, C. J. & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex*, *42*(4), 499-502.

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

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

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

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, 3(3-4): 317-33.

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

Gerrans, P. and Stone, V. E. (2008) 'Generous or parsimonious cognitive architecture? Cognitive neuroscience and theory of mind'. *British Journal for the Philosophy of Science* 59: 121-41.

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.

Hansen, T., M. Olkkonen, S. Walter & K.R. Gegenfurtner. 'Memory modulates color appearance.' *Nature Neuroscience* 9 (2006) 1367-8.

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.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8, 393-402.

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.

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences*, 31: 1-58.

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.

Levin, D. and Banaji, M. (2006)'Distortions in the Perceived Lightness of Faces: The Roleof Race Categories,' *Journal of Experimental Psychology: General* 135: 501-12.

Lyons, J. (2001) 'Carving the Mind at its (Not Necessarily Modular) Joints,' *British Journal for the Philosophy of Science* 52 (2):277-302.

______(2011) 'Circularity, Reliability, and Cognitive Penetrability of Perception', *Philosophical Issues* 21(1): 289-311.

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.

______ (2012) Dissociations in neuropsychology and cognitive neuroscience. *Philosophy of Science*, 79: 490–518.

Macpherson, F. (2012) 'Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism,' *Philosophy and Phenomenological Research* 84 (1):24-62.

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001) Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4(5): 540-545.

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,' *Behavioural and Brain Sciences* 22(3): 385-6.

Nishitani, N., Schürmann, M., Amunts K. and Hari, R. (2005) Broca's region: From action to language. *Physiology* 20:60–69.

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

Payne, K. (2001). 'Prejudice and Perception: The role of automatic and controlled processes in misperceiving a weapon.' *Journal of Personality and Social Psychology* 8: 181-92.

Payne, K., Shimizu, Y., Jacoby, L. (2005) 'Mental control and visual illusions: Toward explaining race-biased weapon misidentifications.' *Journal of Experimental Social Psychology* 41: 36–47.

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,' *Behavioural and Brain Sciences* 3:111-132.

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

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

Siegel, S. (2011) 'Cognitive Penetrability and Perceptual Justification,' *Nous* 46 (2).

______(2013) 'The Epistemic Impact of the Etiology of Experience,'*Philosophical Studies* 162 (3):697-722.

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

Stefanuci, J.K., and D.R. Proffitt. (2008) 'Skating down a steeper slope: Fear influences the perception of geographical slant.' *Perception* 37: 321-323.

______(2009) 'The Roles of Altitude and Fear in the Perception of Height.' *Journal of Experimental Psychology* 35: 424–438.

Stokes, D. (2012) 'Perceiving and Desiring: A new look at the cognitive penetrability of experience.' *Philosophical Studies* 158: 479-92.

_____.(2013) 'The cognitive penetrability of perception' *Philosophy Compass*

Tettamanti, M. & Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex*, *42*(4), 491-494.

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.

Witt, J.K. and T.E. Dorsch. (2009) 'Kicking to bigger uprights: Field goal kicking performance influences perceived size.' *Perception* 38: 1328-1340.

Witzel, C., Valkova, H., Hansen, T., Gegenfurtner, K. (2011) ‘Object knowledge modulates colour appearance’ *i-Perception* 2: 13-49.

Wu, W. (forthcoming) ‘Visual Spatial Constancy and Modularity: Does Intention Penetrate Vision?’ *Philosophical Studies*.

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