Back to the Fodor-modules
The Modularity of Mind Revisited

Abstract:

I revisit Fodor's (1983) notion of "module" and defend it against views weakening the criteria to consider a process modular. Many approaches to modularity (e.g., Barrett & Kurzban's, Carruthers', Cosmides & Tooby's, and others) are unsatisfying because the notion put forward is uninformative. I will explain why we should understand Fodor's notion of module as identifying a homeostatic property cluster (cf. Boyd, 1991). This account should be integrated within a broader view of cognition.

4159 words

I. Introduction

In this paper, I discuss the notion of "module", widely used in cognitive science. This discussion is not a central in the field despite the enormous influence of Fodor's (1983) book, especially in philosophy and in psychology where the discussion has been the liveliest. Although mostly peripheral, I believe the discussion should be put to the forefront again since it is especially important in a context where various accounts of the architecture of mind are suggested. One family of accounts of the architecture of mind, dual-process theories (Evans, 2008; Kahneman, 2011; Stanovich, 2011), also seems to build upon a lot of Fodor's assumption but do so in a relatively uncritical way. The goal of this paper is to start a new discussion of some of these issues.

I first explain Fodor's (1983) modularity of mind thesis and how it is a way, although far from unproblematic, to approach the architecture of mind where modules account for peripheral cognition and where central cognition is not (at all) modular. I then explore how this notion of "module" has been used by evolutionary psychologists to explain parts of central cognition as well as peripheral cognition. Next, I present a different, philosophically informed, take on modularity, where all of the mind's processes are seen as being modular. This view is defended mostly by Carruthers (2006, 2013). I finally suggest that the notion of "module" understood along the lines developed by Fodor should be reintroduced in the literature, but seen as identifying a natural kind following Boyd's (1991) account of natural kinds as homeostatic property clusters (HPC).

II. Fodor-modules

Fodor insists that modularity can only explain peripheral cognition, viz. it cannot explain any aspect of central cognition. In other words, for Fodor (1983, 2000), central cognition is not modular to any extent. By saying that modularity explains peripheral cognition, Fodor means that modularity applies to various input systems (e.g., perceptual systems) and that is not fit to explain other parts of cognition. He does this by distinguishing, in functional terms, between transducers (the retina and optical nerve, the ear drum and auditory nerve, the skin's sensory nerves, etc.), input systems (or compilers: they mediate between the transducer and central cognition; they transform the information into a format that the central system can process) and central processors.

In this paper, a "Fodor-module" will be a module as Fodor defines it in his book: they should possess nine characteristics that I will cover shortly. I will also discuss them critically as I recognize some problems and tensions within this account. It would be a loss to hide the potential for interesting discussions regarding each characteristic. These characteristics are as follows: their operation is (1) domain specific, (2) mandatory once activated (and proceeds automatically), (3) fast and (4) informationally encapsulated; the processes of modules are (5) not centrally accessible (only their output is), they have (6) shallow outputs, (7) exhibit specific "breakdown patterns" and (8) characteristic pace and sequencing (they have an ontogenetic timetable, viz. specific developmental characteristics), and they are (9) associated with a fixed neural architecture. However, it is crucial to remember that Fodor states explicitly that modules may miss some of these characteristics: I will say more on this last point in Section V, where I will go into some details of my proposed modifications of his view.

The idea I want to emphasize below is simple: Fodor's account is interesting to explain *some* cognitive processes and we should keep it (or some version of it). Even if this account offers no explanation for other cognitive processes, those processes *that are* explained by Fodor-modules are explained in interesting ways. Of course, Fodor limits his account to peripheral cognition, mostly input systems, yet even those input systems might be more diverse than his account of modularity allows for, viz. some input systems exhibit some *non*-modular traits (and many central processes or systems have modular traits): these differences between different kinds of modular processes are worthy of exploration, something that is harder to do within frameworks that do not use Fodor-modules.

One of the best known and most discussed characteristic of modules is their domain specificity. Modules, according to Fodor, are specialized to respond to certain inputs, and usually have their own sensory transducers. They are limited to a particular type of inputs, and this would be why they are as efficient as they are. Understood in this way, domain specificity is problematic on a number of levels.

The way to determine how many modules there are in the mind depends mostly on what is considered to be a domain, viz., it depends on the grain used to analyze, and it also depends on what modules are considered to be. And, once the proper grain is found (if there is a way to do so; cf. Prinz (2006, pp. 27–30) for a sceptical perspective), the jury might still be out to find the proper description of a given process (Atkinson & Wheeler, 2004). Vision, for example, likely involves many modules, but should horizontal line detectors and vertical line detectors be counted as two types of modules, or is it more interesting to use the more general class of *edge detectors*? The description of the visual system will likely change in important ways depending of the answer we give here. The general idea still remains important: a module answers to (or is activated by) certain classes of inputs, and determining how to find the proper way and the correct level of analysis is largely up to empirical research.

Modules are automatic and mandatory: visual or auditory illusions provide a good example of this characteristic. If the appropriate stimulus is presented and seen or heard, the illusion will be seen or heard. Fodor presents three examples to illustrate his idea: first, if someone hears a given utterance in a known language, she will hear a sentence and give it meaning; second, any object perceived is perceived in a three-dimensional space; third, touching a surface entails feeling it. The first example he gives is certainly not without its problems, but this characteristic is otherwise fairly straightforward: it is not possible for a subject who has no disabilities not to see when his eyes are opened, not to feel when he touches an object, not to hear when a noise or not to taste or smell when certain molecules come in contact with the appropriate receptors, and the same goes for other types of higher level modules such as speech recognition (if I hear an utterance of English, I will attribute meaning to the utterance).

Modules are fast: once activated, a module usually produces its output well under a quarter of a second. The two important aspects this reveals according to Fodor (1983, p. 63) are the contrast between the speed of the modules' processes as opposed to how slow central processes can be, and the strong link between speed and their mandatory operation.

As Fodor explains, some of the computational problems we solve automatically, such as being able to identify an object against its background, are not necessarily easier or harder computational problems *per se* than solving a long mathematical equation. He even goes further by suggesting that "[t]his dissimilarity between perception and thought is surely so adequately robust that it is unlikely to be an artefact of the way that we individuate cognitive achievements" (Fodor, 1983, p. 63), something Carruthers disagrees with (cf. Section IV). The second idea, that "processes of input analysis are fast *because* they are mandatory" (Fodor, 1983, p. 64), is related to the fact that an automatic process does not involve any decision making process[1]. Reflexes are faster than deciding to do a given action because, simply put, "making your mind up takes time" (Fodor, 1983, p. 64).

The outputs of Fodor-modules are shallow. The notion of shallowness, as applied to modules, is ambiguous, both in Fodor (1983, pp. 86–97) and in the literature more generally. He sees it as the difference between "observation and inference" (1983, p. 86), shallow outputs being those directly observable whereas non-shallow outputs being those that can be inferred (e.g., we can directly observe that a traffic signal is red but can only infer that it is meant to signal the obligation of stopping). However, he remains vague as to exactly what he means by this, and Fodor's vagueness has given rise to a number of distinct interpretations.

Prinz, for instance, rejects the idea that "[s]hallow outputs are outputs that do not require a lot of processing" (2006, p. 25) because it is not precise enough to create any meaningful categorization (what is "a lot" is not clear). Carruthers prefers to understand the idea that outputs of modules are shallow as meaning that they are nonconceptual (2006, p. 4), and this is certainly a strange interpretation of shallowness (e.g., Fodor would probably not agree with this idea, as even the most basic categories, the primitives, are "concepts" for him; cf. Fodor (1998, Chapter 2)). Faucher and Poirier (2009) prefer to explain shallowness in terms of "basic categories", viz. categories that do not use background knowledge, categories that are simpler representations (Faucher & Poirier, 2009, p. 287). This view is less controversial, and more in line with Fodor's project generally. This means that we should understand shallowness as meaning that the modules' outputs are not theoretically charged which seems an interesting way to understand modules' shallowness in the contemporary context. By combining aspects of the three views discussed above, we obtain an account of shallowness as the fact that the outputs of modules involve no reconceptualization using background knowledge because the processes are strongly encapsulated, and this absence of reconceptualization is the reason the outputs of module do not require much processing.

Modular processes are not centrally accessible, viz. we do not have introspective knowledge of their workings. Only their outputs are accessible to other modules or to central cognition. This latter characteristic is analogous to informational encapsulation: the inaccessibility characteristic specifies that the representations within a module are not accessible to central processes, while informational encapsulation tells us that modules cannot have (direct) access to the content of central cognition (or that of other modules). For example, visual modules do not have access to our knowledge that a given picture is an optical illusion (and we cannot consciously affect the inner workings of the module), hence the illusion persists and there is no *direct* way of affecting it.

Fodor argues modules might be associated with fixed neural architectures (Fodor, 1983, p. 98), and that these modules also exhibit specific ontogenetic sequencing. Indeed: during development, some capacities appear at specific moments and in characteristic ways. Language acquisition is the best known case, but there seems to be such developmental constancy in folk physics, folk psychology, mathematics, etc. Piaget's work is the best known on some of these questions and, even if his framework is opposed to domain specificity, the neopiagetians' framework is not (e.g., (Gopnik & Meltzoff, 1998; Karmiloff-Smith, 1992).

As I already mentioned, many processes have these characteristics according to Fodor, but the modularity of mind applies only to peripheral cognition. He argues, and insists on this idea, that central processes are not and could not be modular. For example, central processes are argued to be quinean, viz. processes of the central system are taken to be potentially sensitive to the whole set of beliefs held by the subject (holism), making these processes unencapsulated. Fodor's account has been widely debated, and many disagree with his view, stating it is too restrictive to explain cognition (Carruthers, 2006; Samuels, 2006), sometimes going as far as rejecting the very idea of modularity altogether (Prinz, 2006).

[1] "Because these processes are automatic, you save computation (hence time) that would otherwise have to be devoted to deciding whether, and how, they ought to be performed." (Fodor, 1983, p. 64)

Some of the discussion surrounding the notion of "module" has even proceeded to ignore how Fodor views module even while making explicit references to it to justify the very use of the term.

This other account of modularity has risen within evolutionary psychology: some authors have argued for the massive modularity hypothesis, in which even central cognition might be, at least partially, modular, an idea against which Fodor (2000) has argued since it is a radical departure from his proposed model. What is often not said in these debates is that they are sometimes more about how to define "module" rather than about modularity or the search for modules, Fodorian or not.

III. Evolutionary psychology and the functional approaches

Evolutionary psychologists have used the notion of module in the context of the massive modularity hypothesis. The idea, against Fodor's (1983), is that there is more to "the modularity of mind" than just an explanation of the peripheral systems. According to Barrett, the *only* essential characteristic we should consider when defining "module" is functional specificity (Barrett, 2009, p. 779). If Barrett's definition is accepted, any process of the mind, if it has one (or more) specific function, will be counted as a module, without further consideration as to how it works. This, of course, might be seen as problematic since the notion of module would thereby lose its very substance: "module" would then become an uncontroversial notion that does not add anything to traditional "boxology" in psychology (Faucher & Poirier, 2009).

The problem encountered here is the difficulty of defining what a "module" is. It remains hard to define this term as it is central to many theories, and this difficulty appears clearly with Barrett's minimal characterization. This notion is used to talk about a large number of cognitive phenomenon and "not only have authors used the term *modular* to refer to different concepts, but even explicit definition of the term by some researchers has been insufficient to avoid subsequent misunderstandings by others" (Barrett & Kurzban, 2006, p. 642). Moreover, there seems to be "no agreement on a workable characterization of modules for evolutionary psychology" (Downes, 2010, sec. 3). Still, in all instances, modules are seen by evolutionary psychologists as "adaptations, specific to a domain most of the time" (Faucher & Poirier, 2009, p. 296, my translation). They were selected, through natural selection, for solving adaptive problems such as detecting cheaters in social exchange and recognize kin (these are some of the most common and discussed examples).

If the view of modules as merely adaptations were generally agreed upon by evolutionary psychologists, the notion of module, as defined in this functional account, would lack the explanatory power of a richer notion of module. A very general characterization of module will encompass many distinct processes but, once we identify one such process as a module, we will have very few details on the process in question. A richer account however would provide us with much more information once a module is identified. The analogy I have in mind is the following: in chemistry, if we identify a substance as a non-metal, we will be able to know some of its characteristics, but if we identify the same substance as oxygen, which has a richer definition (including that it is a non-metal, but also its precise chemical structure), we will know a good deal more about the substance in question.

Of course, as Barrett and Kurzban indicate, modularity in general can help "direc[t] the search for specialization", especially in a framework where evolution has a role to play, as it "constrains the hypothesis space regarding plausible functions" (2006, p. 643). Yet, this notion is far from the one we began this discussion with and it has nothing left in common with Fodor-modules (or even Fodor's account more generally). The original evolutionary psychology research program put forward by Tooby & Cosmides (1992) was ambiguous, to say the least, about what modules are and this lead to a series of discussions culminating in a complete dissociation with Fodor's (1983) notion in order to adopt a notion closer to Barrett's (2009):

> […] Fodor's (1983) concept of a module is neither useful nor important for evolutionary psychologists. For evolutionary psychologists, the original sense of module – a program organized to perform a particular function is the correct one, but with an evolutionary twist on the concept of function. (Ermer, Cosmides, & Tooby, 2007, p. 153)

In other words, on this account, "module" does not mean much more than a process that has a particular function. If it is the case, and evidence suggests it is, Faucher and Poirier propose to evolutionary psychologists that they should "simply stop talking about *massive modularity*, and rather talk in terms of a mind massively constituted of *adaptive structures*" (2009, p. 307, my translation and emphasis) because there might still be some use to Fodor-modules or similar structures (with less and / or modified characteristics). Modules, as defined by Barrett (2009) and by Ermer *et al.* (2007), have a lot less explanatory power and this is an important loss if we are to explain how the mind works.

IV. Carruthers' take

Fodor is clear: we should not (and we could not) understand the mind as being *only* modular. Nonetheless, Carruthers (2006) suggests that, with a weaker account of what a module is, we can have a framework where the mind is only composed of modules. Of course, he agrees that peripheral cognition is modular, but he goes further by arguing that central cognition also is *entirely* modular. A major difference between Carruthers' approach and many of the accounts previously discussed is that he argues for massive modularity, without relying on a seemingly insubstantial characterization such as Barrett's, clearly identifying many traits that modules should possess. What is gained by calling every cognitive process "a module" is, as I will suggest, a strategy that lacks justification.

In Carruthers' massive modularity account, *all* cognitive processes are either modular or emerge from the interaction of modular processes. To do so in a plausible way, he loosens the definition of "modules" by removing some of its characteristics and redefining others. As we will see shortly, the notion he ends up with, even if richer than Barrett's (2009), ends up being so inclusive that it might not be much more informative.

Briefly, for Carruthers, modules are processing systems, usually associated with a functional domain, that are frugal in their operations and are more or less strongly encapsulated (he introduces a distinction between *wide-scope* and *narrow-scope* encapsulation), and by and large, only the outputs of a modular process will be available to other processes (Carruthers, 2006, pp. 62–63). However, his account is mostly liberal and allows a lot of variability in each of the characteristics attributed to modules.

Carruthers begins by rejecting some of the characteristics of Fodor-modules, since they would be incompatible with *any* account that argues that some modules are part of central processes. The shallowness of the output is the first rejected by Carruthers, but he also discards speed. Following these modifications, he takes as plausible domain-specificity, the mandatory and innate character of modules and the neural specificity characteristics, before modifying what is meant by encapsulation and, then, adding frugality. I will not, here, discuss the reasons motivating Carruthers' choices: the main thrust of his argument is that he wishes to allow for every cognitive process, peripheral or central, to be modular to at least a certain extent. For example, the rejection of "speed" as a characteristic is clear: modules are fast, according to Fodor, but this characterization only makes sense when the modules' speed is compared to the speed of central processes, viz. modules are faster than central cognition. Without any such comparison, since both peripheral and central cognition are entirely modular for Carruthers, it makes little sense of maintaining that modules are fast (or slow) (2006, p. 9).

Interestingly, this particular characterization of modules does not mean all modules only satisfy the minimum to satisfy Carruthers' idea of module. Some modules may actually have all the characteristics of Fodor-modules: this possibility is not excluded. But, and this is a point Carruthers emphasizes, in the context of the massive modularity hypothesis, some of the characteristics attributed to these "strong" modules must be removed in order to include all possible processes of the mind in the set of modules, even when they do not fit perfectly in Fodor's take on modules.

In the end, Carruthers proposes his weakened notion of modules in order to integrate all cognitive processes and assign to them some of the characteristics of modules in a meaningful way. He does not offer a reason to adopt this position however: the goal is to have a complete account of the architecture of mind in terms of interactions between modules, of which there could be many types. Fodor's notion, however, is replaced without any suggestion to the effect that it could be kept, e.g., as "Fodor-modules" or as other subtypes of modules, to account for more specific kinds of processing.

To use again the chemical analogy made earlier, it is as we had a number of nonmetals (cognitive processes), some of which we could identify as oxygen (Fodor-modules), others as arsenic (another type of modules), but we refused to make the distinction between oxygen and arsenic in order to only have the one notion, non-metal, to explain all of non-metal chemistry.

V. Back to the Fodor-modules

Let's sum up: Fodor has suggested a very narrow conception of what a module is supposed to be, a notion that is not without problems but that has proven to be useful in many areas of cognitive science. In the 30 years since his proposal, many philosophers and psychologists have suggested that we should weaken the notion in order to characterize more cognitive processes as being modules. In this paper, so far, I have resisted this move since I believe it will deprive us of an important and useful tool: the notion "Fodor-module". My argument does not address how we should label things: perhaps Carruthers is right in claiming that talking about modules for nearly every cognitive process is useful (and has appeal). I doubt this, but where I disagree with him is that I believe we should keep the notion of Fodor-module because it is explanatory useful while he seems to think that it should be thrown to the dustbin of history. It does not have to be the only kind of module we end up talking about in cognitive science: it would just be crucial to be precise enough and to use all available tools correctly. The ambiguity of what the term refers to or even the relative unclarity of what "massive modularity" is supposed to mean helps no one.

There remains one very important issue I want to discuss: Fodor's own discussion of modularity is made in terms that could be made more plausible and clearer by applying Boyd's (1991) notion of HPC to specify how modules can be natural kinds. This approach offers very powerful tools to adjust the initial notion and make sense of some remarks Fodor made in the 1983 book: he stated that some modules might miss some characteristics he identified, and that some characteristics could vary in strength and degree. It might even be the case that some of the characteristics are, in the end, optional or present only in a handful of cases. We might even want, at some point, to remove completely some characteristics, or add new ones. This is something that we can only discover and begin to understand by taking "Fodor-module" as a serious subject of empirical inquiry: even if the notions of ontogenetic timetable or fixed neural architecture are appealing and seem to be very useful, they might end up being more misleading than helpful (e.g., Stanovich, 2004 drops them). It might also be the case, however, that they rather provide insights to understand parts of cognition that are still poorly explained, as well as parts of central cognition—if Samuels' (2006) argument that central processes are modular to a certain extent is right.

As Buckner states: "[t]he HPC approach provides a way to distinguish categories suitable for empirical study from those that are not without relying upon essences" (2013, p. 3), and it can be used to modify the way we characterize the cluster as we see fit when we make discoveries about why it seems to be stable—usually in terms of shared properties and common underlying mechanisms. Within such an approach, we can find "the maximal class of items in which a significant number of scientifically interesting properties cluster due to the operation of at least one shared causal mechanism" (Buckner, 2013, p. 3) and, even if we are yet to find such a shared causal mechanisms in the case of Fodor-modules, this leads to a much more plausible understanding of Fodor-modularity.

Given this possibility, it seems that it might still be too early to reject modularity as Fodor understood it altogether. Even better: it might provide us with exactly what we need to start understanding better many of the complex properties of the automatic, unreflective and low-effort class of cognitive processes identified in dual-process frameworks as Type 1 or System 1 processes.

References

Atkinson, A. P., & Wheeler, M. (2004). The Grain of Domains: The Evolutionary-Psychological Case Against Domain-General Cognition. *Mind and Language*, *19*(2), 147–176. doi:10.1111/j.1468-0017.2004.00252.x

Barrett, H. C. (2009). Les modules “en chair et en os.” In T. Heams, P. Huneman, G. Lecointre, & M. Silberstein (Eds.), *Les mondes darwiniens. L’évolution de l’évolution* (pp. 779–786). Paris: Éditions Matériologiques.

Barrett, H. C., & Kurzban, R. (2006). Modularity in Cognition: Framing the Debate. *Psychological Review*, *113*(3), 628–647. doi:10.1037/0033-295X.113.3.628

Boyd, R. (1991). Realism, Anti-Foundationism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, *61*, 127–148.

Buckner, C. (2013). A property cluster theory of cognition. *Philosophical Psychology*, 1–30. doi:10.1080/09515089.2013.843274

Carruthers, P. (2006). *The architecture of the mind : massive modularity and the flexibility of thought*. New York: Oxford University Press.

Carruthers, P. (2013). The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú. Retrieved from http://faculty.philosophy.umd.edu/pcarruthers/The%20Fragmentation%20of%20Reasoning.pdf

Downes, S. M. (2010). Evolutionary Psychology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2010/entries/evolutionary-psychology/

Ermer, E., Cosmides, L., & Tooby, J. (2007). Functional specialization and the adaptationist program. In S. Gangstead & J. Simpson (Eds.), *The Evolution of Mind: Fundamental Questions and Controversies* (pp. 153–160). New York: Guilford Press.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Faucher, L., & Poirier, P. (2009). Philosophie: Modularité et psychologie évolutionniste. In J.-B. Van der Henst & H. Mercier (Eds.), *Darwin en tête! L’Évolution et les sciences cognitives* (pp. 275–308). Grenoble: Presses universitaires de Grenoble.

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge (MA): MIT Press.

Fodor, J. A. (1998). *Concepts: where cognitive science went wrong*. New York: Oxford University Press.

Fodor, J. A. (2000). *The mind doesn’t work that way: the scope and limits of computational psychology*. Cambridge (MA): MIT Press.

Gopnik, A., & Meltzoff, A. N. (1998). *Words, thoughts, and theories*. Cambridge (MA): MIT Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Toronto: Doubleday Canada.

Karmiloff-Smith, A. (1992). *Beyond modularity : a developmental perspective on cognitive science*. Cambridge (MA): MIT Press.

Prinz, J. J. (2006). Is the Mind Really Modular? In R. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 37–56). New York: Blackwell.

Samuels, R. (2006). Is the Human Mind Massively Modular? In R. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 37–56). New York: Blackwell.

Stanovich, K. E. (2004). *The robot’s rebellion : finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.