Why the Separate-Modifiability Constraint does not Preclude a Non-Modular Mind

Thesis for M.A. in Philosophy at Brandeis University

William Edgecomb

Introduction

In the cognitive and biological sciences, where much of our task is to reverse engineer the mechanics and organization of complex functional systems (e.g. minds, organisms), it is sometimes useful to identify those design characteristics—call them *adequacy constraints*—that are necessary for a system to achieve minimally adequate standards of performance and stability. More precisely, adequacy constraints are valuable because they allow further inferences about the nature of the system under investigation—hypotheses not consistent with the relevant adequacy constraints can be ruled out.

In his ‘Argument from Design’, Peter Carruthers uses precisely this kind of appeal to adequacy constraints to persuade us that the mind must be *massively modular* in its organization (i.e. composed of many, many, functionally dissociable components).[1, 2] The centerpiece of the argument is the *separate-modifiability constraint*. This adequacy constraint, which by Carruthers’ lights applies to complex functional systems generally (henceforth just *designed systems*), claims the following: for a given system, it must be feasible for the agents that effect its phylogenetic change—namely natural selection—to make targeted alterations to particular functions without thereby disrupting the mechanics that underpin the many other functions of the system. The crux of Carruthers’ argument is that because of how it compartmentalizes the realizers of a system’s various functions, and as a result buffers overall functionality against the

[1] Carruthers (2006).

[2] According to Carruthers, the Argument from Design derives ultimately from Simon (1972).

perturbations of phylogenetic change, a massively modular organization alone can satisfy the separate-modifiability constraint. And so the mind, as every designed system must be, is massively modular in organization.

But while massive modularity is clearly one way for designed systems to achieve adequate stability across evolutionary change, Carruthers provides little reason to believe it is the only way. It will be my purpose to argue that in fact, it is entirely plausible for systems with highly integrated and highly multifunctional components—that is, *non-modular systems*—to satisfy the separate-modifiability constraint. The prospects are strong particularly for those non-modular systems where (1) short-run phylogenetic changes yield only small, incremental changes to a system's functional behavior, or (2) sets of functions with highly integrated mechanics exhibit *loose coupling*, such that an alteration to one function in the set will tend to result in alterations of different magnitudes to the other functions in the set. My task then will be to explain why conditions (1) and (2) facilitate functional stability in non-modular systems, as well as to explore the sorts of design features that would allow (1) and (2) to obtain in the first place. If my reasoning is on track, there is ample room for non-modular designed systems to satisfy the separate-modifiability constraint, and thus little reason to accept Carruthers' inference to a massively modular mind.

The structure of the paper will be as follows: In Section I, I will introduce Carruthers' preferred construal of the massive modularity hypothesis (the claim that mind specifically is massively modular) and explain how the Argument from Design figures into his overall defense of the view. In Section II, I will articulate and expand upon the Argument from Design and its principal claim that a non-modular design to the mind is precluded by the separate-modifiability constraint. In the process some important details will be filled in about what, for our purposes, is required of a system before it qualifies as massively modular. In Section III I will rebut the argument laid out in Section II. And in Section IV, I will reevaluate the state of the controversy, in light of the arguments made here. The source for all of Carruthers' views examined in this paper is his 2006 book, *The Architecture of Mind: Massive Modularity and the Flexibility of Thought*.[3]

I. Massive modularity as many, many, functionally dissociable components

What does the massive modularity hypothesis amount to? Carruthers' strategy for articulating the view is to lay out the best arguments in its favor, and for each, to draw out the version of the hypothesis the argument best supports. Ultimately, the account he opts to defend is a composite, combining the (purported) implications from a few key arguments. For that full account, I refer you to *The Architecture of Mind*. Since our interest here is limited to Carruthers' Argument from Design, so too, the only account of massive modularity that will concern us is the one purportedly established by that argument.

And what does that account of the massive modularity hypothesis amount to? The details will emerge only later, when we recapitulate the Argument from Design in Section II. But for now, we can state its central claim: the mind is composed of many, many, *functionally dissociable components*, where a component is functionally dissociable to the extent that one, it has a particular function, and two, its operations are to a significant degree independent of the operations of the rest of the system. As Caruthers sees it, this notion of functionally dissociable component captures the ordinary sense in which we understand consumer electronics to be modular. He speaks of the hi-fi system as an example:

> 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-

[3] Peter Carruthers, *The Architecture of Mind: Massive Modularity and the Flexibility of Thought*, New York: Oxford University Press, 2006.

> 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.[4]

Notice the role that independent component operations plays in the example. It is precisely the fact that the component parts (e.g. speakers, tape-deck, CD player) are not greatly intertwined in their operations that allows some parts to be interchangeable with others, and some parts to function even if others are damaged or missing.

At a minimum then, Carruthers takes the Argument from Design to warrant the claim that the mind is composed of a very great many functionally dissociable components. This is a decidedly weak way of construing the massive modularity hypothesis, and the weakest way it gets presented in *The Architecture of Mind*.[5] Indeed, a number of authors have complained it is so weak as to be uncontroversial, or at least not controversial enough.[6] Here is one way to put the objection: neurons are, arguably, functionally dissociable cognitive components. Each neuron has a particular function (namely to induce an action potential above a certain threshold of excitation), and each conducts its operations with a significant degree of independence. But obviously, so the objection goes, it would be trivial to conclude that the mind is massively modular on the mere basis of its many neurons. Therefore, to avoid triviality, the massive modularity hypothesis must claim something more substantial than that the mind is composed of many, many, functionally dissociable parts.

For our purposes, we can set aside this objection. That is because the version of the massive modularity hypothesis that is the target of our critique, the one Carruthers takes to be warranted by the Argument from Design, is considerably more robust than the one assumed by

[4] *Ibid*., 2.
[5] For a much more robust notion of cognitive module, see Jerry Fodor's seminal book, *The Modularity of Mind* (1983).
[6] See Cowie (2008), Wilson (2008), Prinz (2006), and Samuels (2006).

the objection. To be sure, composition by many functionally dissociable components remains the core commitment of the view. But as we shall see in the following section, Carruthers takes the separate-modifiability constraint to warrant further conclusions about the mind's organization, most saliently that the mind's components must be organized in a many-tiered hierarchy. The result is a hypothesis that is both interesting and controversial.

II. The separate-modifiability constraint and the Argument from Design

The separate-modifiability constraint is at the heart of the Argument from Design. To get a feel for what the constraint amounts to, it will be helpful to first say something about designed systems (or, complex functional systems), the class of entities to which the constraint applies. For our purposes, what makes a system a designed system is that it performs many, many, functions. The paradigmatic category of designed systems is the *organism*. From metabolism to habitat selection to gamete formation and beyond, organisms employ a huge array of functions in the service of survival and reproduction. Other examples of designed systems include biological entities that occur within organisms, such as organs, tissues, and cells, while others still include multi-organism collectives such as the colonies of eusocial insects. Some robots, computer programs, or other artificial systems may qualify as well.

We can assume that designed systems are subject to a selection process, and that they undergo evolutionary change as a result. The type of selection process will vary with the type of system in question. For biological entities, the principal means of selection is natural selection, through which genotype mutations that yield fitness-enhancing phenotypic changes become increasingly prevalent in populations. For artificial designed systems on the other hand, selection is for the most part driven by decisions made by a designer or engineer. Through this agent-guided process of selection, artificial systems undergo evolutionary change that is in many respects similar to the biological case, with adjustments to functionality occurring in a more or less gradual fashion (consider the successive generations of Apple's iPhone, for instance).

For any evolutionary alteration of a designed system, we can consider each of the system's many functions and ask how it was effected by the change. Sometimes an alteration to a system's mechanics will improve one or more of the system's functions—that is, increase the function's net contribution to overall fitness. For other functions, the effect of evolutionary change might be detrimental, such that those functions end up contributing less to overall fitness than they did previously. With respect to other functions still, an evolutionary alteration might be fitness-neutral, either because those functions are entirely unaffected, or because the way in which they are effected yields no change in their net contribution to fitness.

So designed systems have many, many, functions, and undergo evolutionary change as the result of a selection process. Furthermore, phylogenetic alterations can have either positive, negative, or neutral fitness-effects with respect to each function of a system. For Carruthers, the separate-modifiability constraint follows as a relatively straightforward consequence of these features of designed systems. Broadly speaking, the constraint is a limitation on what, for designed systems, constitutes an adequate manner of response to selection pressures. In particular, it says that the functions of a designed system must be separately modifiable, such that evolutionary adjustments can be made *just to those functions under selection for change*—that is, without in the process affecting the many other functions of the system. In Carruthers' words,

> "Evolution needs to be able to add new functions without disrupting those that already exist; and it needs to be able to tinker with the operations of a given functional sub-system—either debugging it, or altering its processing in response to changes in external circumstances—without affecting the functionality of the remainder."[7]

[7] Carruthers, *The Architecture of Mind*, 21-22.

The thought is that if the functions of a designed system were somehow not separately modifiable, such that a change to one function would tend to precipitate a cascade of changes to other functions, then the overall functionality of that system would be disastrously unstable across phylogenetic change. Whenever those functions under selection are improved through evolutionary alterations, many others would be disrupted as an inadvertent byproduct, and the system would struggle to maintain any already well-tuned elements of its functionality.

That the separate-modifiability constraint holds is particularly plausible in light of two very likely assumptions. The first is that the mechanics that underlie most functions are finely tuned and highly sensitive to disruption. To put it more precisely, there are always going to be vastly more ways in which altering a function's mechanics is detrimental to the function rather than positive or neutral. If this assumption is correct, then designed systems would do best to limit the number of pathways by which their functions could be effected in the course of evolutionary change. Avoiding linkages between functions that would force selection to act on many functions at once, as per the separate-modifiability constraint, is a clear advantage in this respect. The second assumption is that designed systems must be to a significant degree amenable to evolutionary adjustments. This rules out the possibility that designed systems can have many collectively alterable functions, and at the same time circumvent the costs associated with this design by severely blunting the extent to which any function is alterable at all.

The gist of the separate-modifiability constraint should now be clear. Now let us notice an important detail: the constraint claims not just that the separate modifiability of functions confers a fitness benefit, or that it is an essential element of optimal system design. Rather it makes the stronger claim that designed systems *need* to have separately modifiable functions. How does Carruthers justify separate modifiability as a bona fide requirement? He does not do so explicitly, but it is easy to fill in the argument. Presumably, the idea is that designed systems without separately modifiable functions, to the extent that they could come into existence even briefly, would be readily eliminated in the selection process. Crippled by their phylogenetical instability, they would struggle to adapt to changing circumstances over evolutionary time, and they would be easily outcompeted by systems that had already implemented a design architecture with separately modifiable functions.

To see how the separate-modifiability constraint figures into Carruthers' overall argument, it is instructive to notice that it is a token of a more general type of constraint. It is, to remind you, an adequacy constraint. Thus its role is to classify a design characteristic as required for all members of a class of designed systems, on the grounds in the absence of that attribute, no member of the class could feasibly achieve base-level standards of performance and stability. For the separate-modifiability constraint, the class of designed systems at issue is *all designed systems*. The attribute required is *the separate modifiability of functions*. And the facet of performance or stability that would be compromised in the absence of this attribute is *the capacity make evolutionary alterations to particular functions without damaging overall functionality as a consequence*.

Identifying adequacy constraints has a clear discursive value: if we know that a type of designed system must have a particular attribute, then we can eliminate from consideration any hypotheses on system design that are incompatible that attribute being attained. It is in precisely this way that Carruthers' Argument from Design aims to leverage the conclusion that the mind must be massively modular in its organization. The following is the crucial step in the argument, and the target of our critique in Section III: *only designed systems that are massively modular can satisfy the separate-modifiability constraint*. If this premise is true, then Carruthers has a strong case for the massive modularity of mind, and principled grounds for rejecting proposals of non-modular cognitive architecture.

How then does Carruthers defend this major premise, that massively modular designed systems alone can satisfy the separate-modifiability constraint? Recall that Carruthers is not using the Argument from Design to justify some preconceived notion of what it is for the mind to be massively modular. Rather, his strategy is to let loose the argument, and then favor whatever version of the hypothesis it happens to support best. It should not be surprising then that the notion of module that precipitates from the Argument from Design—that of the *functionally dissociable component*—is a perfectly tailored solution to the separate-modifiability constraint. What makes it so? Remember the two essential properties of functionally dissociable components. One, each performs a particular function. And two, the operations of each component are to a significant degree independent of the operations of the rest of the system. These properties naturally facilitate the separate modifiability of functions. To the extent that functions are realized within independently operating, discrete component modules, natural selection has a ready means to adjust any given function—it need only make alterations to a function's proprietary module. What's more, the functional changes to the system will be limited, more or less, to the particular function realized by the module undergoing alteration. This is not only because modules are, on Carruthers' proposal, functionally individuated. It is also because they have limited interactions with the rest of the system, thus constraining the degree to which altering one module will have downstream effects on the behavior of the system's other components. Putting it all together, this is how Carruthers explains the significance of modules:

> Since the properties of modules are to some significant degree independent of one another, both they and the developmental pathways that lead to them can have

> distinctive effects on the overall fitness of the organism. But by the same token, since modules are separately modifiable, natural selection can act on one without having to make alterations in all (which could have potentially disastrous effects). So evolution can tinker with the separate components of the overall organism, at many levels of organization, responding to particular evolutionary pressures by factoring overall fitness in distinctive fitness-effects of component modules…only a modular organization can enable this to happen…[8]

And later,

> The prediction of this line of reasoning, then, is that cognition will be structured out of systems that are to some significant degree dissociable, and each of which has a distinctive function, or set of functions,[9] to perform.[10]

Worth emphasizing, the argument here is supposed to establish *massive* modularity of mind, not just that the mind is modular to some interesting extent. At least insofar as the massive modularity hypothesis claims merely that the mind is made of many, many modules (Carruthers' weakest construal), it is easy to see why the argument should have this virtue. Since modularity achieves separate modifiability by mapping functions onto modules in a (more or less) one-to-one fashion, the number of modules in a system will be proportional to the system's number of functions. And, it is paradigmatic of designed systems that they have many functions; hence so too must they have many modules.

However, as noted previously, this weakest construal of massive modularity is extremely minimal, so much so that it has elicited worries of triviality. We should hope then that the Argument from Design recommends further, more controversial inferences about what massively modular organization amounts to. Indeed it does, and we shall turn to this matter in a moment. But first, let us summarize the core thread of argumentation in Carruthers' Argument from Design.

[8] *Ibid*., 14.
[9] That Carruthers thinks modules often perform multiple functions is a significant caveat to his view, and will be discussed later.
[10] *Ibid*., 17.

So far, we have established the following as major premises:

(1) The separate-modifiability constraint holds. Thus the functions of designed systems must be independently alterable in the course of selection for phylogenetic change.

(2) Only designed systems with a massively modular organization can satisfy the separate-modifiability constraint.

To derive the conclusion, Carruthers invokes one additional, minor premise, which I assume to be true:

(3) The mind is a designed system.[11]

Conclusion: Therefore, the mind is massively modular.

We will soon turn to a rebuttal of this argument, but not before briefly examining the stronger version of the massive modularity hypothesis that the Argument from Design allegedly warrants. This augmented version of the hypothesis maintains the central notion of module as functionally dissociable component, as well as the claim that cognitive modules are present in great quantities. What it adds is the claim that modules are organized in a many-tiered hierarchy, such that when we look at functional organization at any grain of analysis, the modules we find will themselves be composed of functional sub-modules (and those sub-modules will in turn be composed of functional sub-sub-modules) and so on until bottoming out at the ground-level neural or sub-neural processes.

Carruthers provides details about the shape and size of the mind's organizational hierarchy that we need not get into.[12] I just want to emphasize one thing. He thinks that owing to the great complexity of mid to high-level functions (e.g. mate selection, practical reasoning), the hierarchy must be deep, in the sense of having many layers (As he says, "And whenever the function performed is complex, the sub-system in question should itself decompose into an array

[11] See pages 15-17.
[12] See pages 17-21.

of sub-subsystems, and so on."[13]). It is straightforward to see how the Argument from Design lends itself to this conclusion. As we understood the argument earlier, designed systems such as minds and organisms need to be organized along modular lines in order to insulate their many functions from the disruptive vicissitudes of phylogenetic change. Now we need simply recognize that higher-level functions (or at least their incarnate instantiations) are *themselves* designed systems, and repeat the argument with them as the targets for analysis. That is, we could plausibly argue that functions employ a huge array of (sub-)functions, that functions are subject to a selection process, and that as result, their subsidiary functions require the protection conferred by a modular organization no less than do minds, organisms, or other paradigmatic designed systems. Repeat this argument for all the functions of the mind that are sufficiently complex, and we get the result that the mind should be modular not just at one level of organization, but through and through. This makes for a robust and controversial construal of the massive modularity hypothesis.[14]

III. The Argument from Design rebutted

(III.i) *The first strategy for rebuttal*

There are two principal strategies for resisting the Argument from Design. The first is to reject what I have labeled premise (1), and argue to the contrary that designed systems need not have functions that are separately modifiable. In this vein, it is commonly argued that while separate modifiability confers a benefit with respect to phylogenetic stability, this benefit is but a single feature across a much greater cost-benefit landscape. Namely, there are benefits to integrated organismal designs that we need to consider, as well as costs that fall to designs

[13] *Ibid.*, 20.

[14] See, for example, Van Orden, Holden, and Turvey (2003) and Kello and Van Orden (2009) for an account of cognition that, while it might countenance neurons as Carruthers-style modules, would clearly reject his account of modularity at the higher levels of cognitive organization.

prioritizing separate modifiability, namely modular designs. For example, Schulz (2008) argues that it is sometimes better for groups of character traits to be collectively alterable, namely when the traits in question serve fitness best when present in fixed proportions (e.g. the optimal degree of *aggressiveness* might be directly proportional to an animal's *fighting strength*). And, Hansen (2003) argues that pleiotropy, whereby individual genes influence not just one but a variety of character traits, may actually facilitate evolvability insofar as it increases the variational potential for the traits affected. Also Krohs (2009), drawing from perspectives in engineering, emphasizes the costs associated with modular designs, among them a protracted ontogenetic process, as well as increased material and energetic requirements.

Indeed, even Carruthers is somewhat amenable to this sort of critique, as he too identifies some costs associated with modularity. Like Krohs (2009), Carruthers calls attention to energetic constraints. He states, "the brain consumes energy at about eight times the rate that would be predicted from its mass alone…So adding extra processing power doesn't come cheap."[15, 16] And, he admits that this consideration does not seem to favor a modular brain architecture, which requires many, many, cognitive subsystems. As he says, "To some extent this resource constraint pulls in the opposite direction from the constraint of separate modifiability…If minimizing energetic costs were the major design criterion, then one would expect that the fewer brain systems there are, the better."[17]

Carruthers thinks that the design of the mind is a compromise of these (among other) competing evolutionary pressures, with the separate-modifiability constraint favoring more cognitive subsystems, and resource constraints favoring fewer. His view is that both constraints

[15] *Ibid*., 23.
[16] Relatedly, Carruthers also highlights costs associated with increased head size, such as a higher rate of maternal death during labor.
[17] *Ibid*., 23.

can be reasonably well satisfied so long as modules are conservative in their use of resources. Specifically, Carruthers thinks modules achieve efficiency by engaging in massive sharing of parts, and by (in many cases) performing a multitude of functions.[18] Carruthers does not take this to weaken his claim to a massively modular mind in any significant way, but it seems to me the opposite is true. The more that modules are multifunctional, and the more that they share parts, the less that they can remain functionally dissociable in the way Carruthers' view demands. Most obviously, the more functions a module takes on, the less it can be said to be functionally specialized. And the more that parts are shared between modules, the less that modules can remain independent in their operations. Notably, whenever a single module with shared parts is destroyed, the other modules with which it shares parts will be damaged too, and in direct proportion to the number of parts shared.

In any case, while I think there is merit to arguments that cast doubt on the separate-modifiability constraint as a major shaper of biological and cognitive design, this is not where I will focus my efforts. Instead, my contribution will be to motivate the second principal strategy for resisting the Argument from Design. Here premise (2) is rejected, the premise that claims massive modularity necessary for satisfying the separate-modifiability constraint. On this line of response then, which to my knowledge has not been discussed in the philosophical literature, it is argued that designed systems organized in a highly integrated, non-modular fashion can in fact satisfy the separate-modifiability constraint.

(III.ii) *The second strategy for rebuttal*

Before delving into the argument here, let us briefly state what we mean by an *integrated* or *non-modular* design architecture. Keeping with Carruthers' framing of the debate, let us simply allow that a designed system is integrated to the extent that it lacks functionally

[18] See page 21 and pages 23-24.

dissociable components. Thus, a designed system is integrated to the extent that it lacks components that one, perform a particular function, and two, are to a significant degree independent in their operations. Or to put it positively, an integrated system is marked by components that participate in many functions, and whose operations are also greatly intertwined with the operations of the remainder of the system. Also worth noting, an integrated system may have many functions that are not mappable to any organizational unit that can properly be classified as a component. While such functions would likely be the product of interactions *among* components, the components at play would in these cases not cohere to form a superordinate component.

It is not hard to see why, *prima facie*, Carruthers takes it that integrated systems cannot satisfy the separate-modifiability constraint. To make alterations to a function, changes need to be made to the function's underlying mechanics. But if groups of functions are greatly entangled in their mechanics, then it seems any tinkering to the system will affect a multitude of functions, even if only one is the target of (directional) selection. This is certainly a reasonable concern. We might say that insofar as the separate-modifiability constraint holds, satisfying the constraint is a substantial obstacle for designed systems with an integrated functional organization. But is it such a stringent difficulty that we can rule out the possibility that there could be such systems, as Carruthers argues? It is my contention that Carruthers has rushed to judgment on this issue. As we shall see in what follows, it is plausible that integrated systems can be organized in such a way as to satisfy the separate-modifiability constraint.

Indeed, there are several design strategies integrated systems could implement in service of attaining separately modifiable functions. Here is the first strategy: *organize functional mechanics so that short-run phylogenetic alterations yield only small, incremental changes to* *functional behavior*.[19] To see the efficacy of this design strategy, consider a system with two functions A and B that are integrated in their mechanics and alterable only jointly. More specifically, suppose that short-run evolutionary adjustments beneficial to one function will necessarily result in detrimental changes to the other. Suppose also that when such joint alterations to A and B occur, the degree of change to each function will be of similar magnitude. Here's the key: if joint alterations to A and B are very modest in intensity, then for whichever function is hurt by an alteration, the resulting cost to fitness will be similarly modest, and hence tolerable. Natural selection could therefore make targeted alterations to one function without seriously destabilizing the other as a consequence.

There are a couple of miscellaneous items worth noting here. First, it is something of a matter of interpretation whether strictly speaking, the separate-modifiability constraint is satisfied in the example above. After all, I have stipulated that functions A and B are modifiable only jointly. Still I think we could plausibly say that the separate-modifiability constraint is satisfied, for satisfaction could simply require that the degree of joint alterability be kept below a certain minimum threshold. Second, the viability of this design strategy, where phylogenetic alterations are linked but modest, is greatest when the intensity of selection pressure is higher for some traits in a set of linked traits than for the others. That way, while tweaking by natural selection will result in a balance of beneficial and detrimental modifications, still it can deliver a net gain to fitness so long as the changes made favor the traits under the most intense pressures for selection.

How might the design strategy under discussion be implemented? I think the basic idea can be captured with two general organizational principles. One: *let individual functions be*

[19] While (seemingly) absent from the contemporary philosophical discussion on modularity, the interest of the biological community in this sort of organismal design strategy can be traced back at least to Conrad (1990), where its virtues are touted at length.

distributed over a sizable portion of a system. And two: *do not let any element of a function's mechanics be indispensable to the function's proper behavior*. These two principles are really driving at the same point. Genotypes are not thoroughly reshuffled from generation to generation, but rather change only incrementally. Thus, in order for a function to remain robust against genetic changes, it need only be able to withstand more or less isolated genetic disruptions. The above principles exactly embody a means for functions to remain robust in this manner. By spreading themselves into many nooks and crannies of their systems' architectures, and by avoiding elements that are absolutely necessary for proper functional behavior, functions can vastly reduce their vulnerability to disruptive genetic changes.[20]

Of course, the design strategy we've been discussing has its downsides. For one, all things being equal, it would be preferable if functions were entirely separately modifiable—that is, modifiable without necessitating any harmful alterations to another function or functions. Furthermore, while dampening the rate of phylogenetic change is good for functions being changed for the worse, the flipside is that it slows the rate of evolution for those functions that are benefitting from selection. Thus there are some reasons to doubt that designed systems adopting this strategy would be sufficiently evolvable.

Still, it could be countered that instantiating functions within complex, distributed networks—for which only slow and gradual phylogenetic change is possible—has benefits that outweigh its costs. For instance, this design type might facilitate finely tuned movements through phenotypic space, across evolutionary time, in a way not possible for systems prioritizing speed of adaptation. But this discussion would take us too far afield. And at any rate, there is at least

[20] These two principles are very much in line with the increasingly popular hypothesis that cognitive processes are self-organizing, or "soft-assembled." See for example Kello and Van Orden (2009).

one other, arguably better way for integrated systems to satisfy the separate-modifiability constraint, which we shall turn to now.

The second design strategy: *organize systems so that sets of functions with highly integrated mechanics exhibit loose coupling, whereby an alteration to one function in the set will tend to result in alterations of dissimilar magnitude to the other functions in the set.* For illustration, let us once again consider two functions A and B that are highly integrated in their mechanics. Let us suppose function A is subject to directional selection, and function B is subject to stabilizing selection. The virtue of loose coupling between A and B is this: if A can be altered via natural selection while only necessitating comparatively small changes to B, or none at all, then the separate-modifiability constraint would be straightforwardly satisfied. Thus when a system exhibits the right sort of loose coupling relations between functions, natural selection is provided a means to act on individual functions without destabilizing the others with which their mechanics are integrated.[21]

Let's take a look now at how loosely coupled functions might be implemented in an integrated system. One way is pretty straightforward—it might be that for a set of integrated functions, the functions in the set depend on such different properties of the shared underlying mechanics that alterations significant to one function may for the others be neutral, or almost neutral. Here is a non-biological scenario to illustrate the point: suppose you are a computer engineer with a laboratory in the northern climes of Scandinavia. The computer system you are working on is designed to process huge amounts of data, and as a consequence is big and bulky, and also produces a lot of heat. Indeed, as you see it at least, producing heat is one of its

[21] An interesting question that is outside the scope of this paper: to what extent are the two design strategies presented so far compatible? That is, as a means for integrated systems to achieve separate modifiability, to what extent is it beneficial, or counterproductive, to combine the strategy of slowing phylogenetic change with that of utilizing loosely coupled functions?

functions. For you deliberately arranged the laboratory so that the supercomputer would heat up the workspace, in lieu of a traditional heating system. And, importantly for our purposes, the computer's heat-generating function relies on the same underlying mechanics as does its principal function of processing big data. That is, what drives the data-processing is precisely the same electrical signals and churning away of the hardware that makes the computer generate heat (certainly, there is no data-processing module side by side with a heat-generating module!). Now imagine that you want to make significant alterations to the way the computer processes data. Does the integration of the data-processing function with the heat-generating function imply that the former cannot be altered without disrupting the latter, á la the Argument from Design? No, because the data-processing function is sensitive to differences in the mechanics at a much finer grain than is the heat-generating function. The result is a loose coupling between the functions, such that there are many, many ways in which the character of data-processing could be altered without affecting heat-generation in any appreciable way.

Perhaps the brain is host to integrated functions that are loosely coupled in a similar sort of manner. Suppose, for example, that there are in fact networks of neurons that are highly functionally integrated. How does such a network coordinate its activity so that it serves the right function (or functions) at the right time, given the demands of the moment? Part of the answer might be that the network has a number of what we might call *functional modes*, or states of the network that serve various respective functions, and are activated and maintained only under the right parameters. However, the parameters required to activate and maintain a particular functional mode would likely be somewhat broad, or permissive, in order to give it adequate stability and to facilitate its activation. In the parlance of dynamical systems theory, we might call such functional modes *basins of attraction*. If this basic picture is right, the kind of network in question would plausibly exhibit loose coupling. For although there would likely be certain key alterations to a network's mechanics that would yield significant changes to the character of some functional mode or another, by and large, most functional modes will be tolerant of most possible changes to the network. Put simply, there should be alterations we could make to the network that affect one or a few functional modes significantly, but without much affecting the others.

So far, the examples of loose coupling we have examined presumed a kind of static relationship between functions. That is, for our examples, when one of the coupled functions underwent an evolutionary alteration, the other did not need to compensate in any way in order to retain its functional character; it was simply not affected by the change. However, a propensity for dynamic compensation is plausibly a very effective means for integrated functions to achieve loose coupling. Here's a non-cognitive example of how it might work: suppose that some chemical substance S plays a crucial role in two separate cellular metabolic processes, both of which take place freely in the cytoplasm. Let us say that at present, both processes perform optimally at the same concentration of S. Now let us suppose that circumstances change, and a selection pressure presents itself that favors a lesser concentration of S, but for only one of the two metabolic processes. Say the needed evolutionary adjustment is made—now the cell maintains substance S at lower concentrations. My idea is that the metabolic process that favors more of S than is currently present can utilize compensatory mechanisms to retain optimal or near-optimal performance. For instance, it could be that when S is maintained below a certain threshold, an enzyme is called upon to catalyze the appropriate S-involved chemical reactions. That way, the metabolic process ostensibly disadvantaged by the alteration can actually maintain a rate of reaction equivalent to when S was at its original concentration. This would be a clear example of loose coupling—the two metabolic processes would be loosely coupled with respect to evolutionary changes that affect the concentration of S.

Compensatory mechanisms that buffer functionality against evolutionary perturbations work by exploiting the multiple realizability of functions. Basically, since there are multiple ways to configure the underlying mechanics of a function, designed systems can simply switch between the different configurations as needed, depending on the particular disruptions to mechanics taking place over evolutionary time. This seems to be a quite workable solution to the separate-modifiability constraint, and we should not be all surprised if it is widespread in the biological world. Again, neural processes may provide a compelling example. Notice, for instance, just how many ways there in which neural activity gets affected or modulated. There is both ionotropic and metabotropic synaptic transmission, either of which can be excitatory or inhibitory. There are receptors of various sorts that can be added or subtracted from postsynaptic neurons. There are freely diffusing gaseous neuromodulators (most famously nitric oxide), as well as hormonal neuromodulators that freely cross the blood-brain barrier. There are glial cells that help to coordinate neural activity in various ways, for example by building myelin sheaths around axons, which facilitate proper timing of action potentials. The list goes on. Here is the takeaway: neural processes have ample resources with which to implement a strategy of adjusting their mechanics to compensate for phylogenetic disruptions. Perhaps we should be surprised only if such a compensatory strategy was *not* employed by functions rooted in neural activity.

IV. Reevaluating the controversy

We will now discuss an objection Carruthers would likely raise, and then turn to reevaluate the state of the controversy more generally. The objection I have in mind derives from what Carruthers has to say about the architecture of artificial systems, such as computer programs and applications of artificial intelligence. He argues that artificial systems of sufficient complexity need to have a modular architecture, and for the same reason that biological systems do (i.e. in order to preserve the separate modifiability of functions). And more crucially for our purposes, he takes it that the ubiquity of modular architectures in artificial systems constitutes further evidence for the necessity of modularity in biological designs. He explains,

> …many programming languages now require a total processing system to treat some of its parts as 'objects'…This enables the code within the 'objects' to be altered without having to make alterations in the code elsewhere, with all the attendant risks that would bring…And the resulting architecture is regarded as well nigh inevitable (irrespective of the programming language used) once a certain threshold in the overall degree of complexity in the system gets passed.[22]

And,

> The fact that human designers of intelligent systems have converged on modular organization is evidence that the human mind, similarly, will be modular in design.[23]

Here's one more way we could put the objection: if integrated design architectures could viably satisfy the separate-modifiability constraint, then there would be at least some examples of human-made, integrated designed systems. But there are no such examples; therefore integrated systems are unfit to satisfy the separate-modifiability constraint.

The objection is far from convincing. For no effort has been made to rule out alternative explanations as to why integrated design architectures have, for the most part at least, eluded human engineers. That integrated designs are simply inadequate is just one possibility. But another quite plausible scenario is that there are cognitive and technological barriers that have so far stood in the way of integrated designs making their breakthrough. Cognitively, it might be that modular designs are more intuitive than their integrated counterparts, and can be understood

[22] Carruthers, *The Architecture of Mind*, 22.
[23] *Ibid*., 23.

and worked with in a way that puts a lot less demand on our mental resources. Even Herbert Simon, from whom Carruthers claims the Argument from Design ultimately derives, agrees that integrated systems might be extremely difficult for humans to grasp. As he says,

> If there are important systems in the world that are complex without being hierarchic ['hierarchic' here means *modular*, roughly], they may to a considerable extent escape our observation and understanding. Analysis of their behavior would involve such detailed knowledge and calculation of their elementary parts that it would be beyond our capacities of memory or computation.[24]

In this instance, Simon is referring to our ability to identify and understand biological integrated systems, rather than our ability to engineer integrated systems ourselves. But the point is the same: the complexity of component interactions is greater for integrated systems than modular ones, and this makes the former type much harder for us to grapple with mentally.

So far, it has been implicit in my response that in order to accomplish much of anything as engineers, humans need to understand the functional mechanics of the systems they are working on. If this is right, and if it also turns out that integration is widespread in the biological kingdom, it should be no mystery why organisms have it and human artifacts don't: natural selection is not constrained by a requirement that it comprehend its designs, and so is much freer to develop integrated systems. Still, cognitive barriers to engineering integrated systems could perhaps be circumvented, but probably only by breaking down technological barriers as well. What we need, arguably, is an engineering methodology that simulates natural selection. That is, we need a way to search the space of possible artifactual designs that relies less on the predictions, insights, and plans of human engineers, and more on a guided process of tinkering and trial and error. With such a strategy for engineering, which would likely require the assistance of sophisticated and powerful computational algorithms, design solutions could be

[24] Herbert A. Simon, "The Architecture of Complexity" *Proceedings of the American Philosophical Society*, 106.6 (1962): 477.

landed upon on the mere basis that they do what is desired of them. It would not be a constraint that anyone could have antecedently predicted the solution, or indeed, that anyone even understand how the resulting system works.[25]

Now let us ask more generally what we should make of the controversy surrounding the Argument from Design, in light of the arguments made here. I think the main takeaway is that the Argument from Design is inadequate as *a priori* grounds for the massive modularity hypothesis. As I hope to have shown, armchair reflection is enough to formulate plausible hypotheses as to how integrated designed systems could satisfy the separate-modifiability constraint, and hence remain phylogenetically stable. But of course, the considerations raised here are a far cry from proving that the mind is functionally integrated to any interesting extent. I take it that that matter is an empirical one, and that the jury is still out. I would even concede that if modularity turned out to be biologically ubiquitous, it is possible that the requirement of separate modifiability remain a crucial determinant of that fact. What I would not accept, however, is that Carruthers has given us an adequate account of why separate modifiability can be achieved only via a massively modular organization. At the very least, he owes an explanation as to why the possibilities for integrated designs discussed in this paper wouldn't also suffice to satisfy the separate-modifiability constraint, or why those designs wouldn't be otherwise viable.

V. Conclusion

The major premises of Peter Carruthers' Argument from Design, which aims to establish that the mind is massively modular, are as follows: (1) there is an important constraint on the functional organization of designed systems. Namely, designed systems must be organized so that selection can make targeted alterations to particular functions—that is, alterations that do not

[25] Indeed, there are great breakthroughs being made in the field of evolutionary robotics, which both aims to simulate natural selection, and clearly seems like a platform amenable to integrated designs. See for example Philippides et al. (2005).

inadvertently disrupt the many other functions of a system. We've called this the *separate-modifiability constraint*. And, premise (2): the separate-modifiability constraint is satisfiable only by designed systems organized along massively modular lines. While I have touched on a few reasons we might be skeptical of premise (1), the primary purpose of this paper has been to challenge premise (2). I have argued that integrated or non-modular systems can quite plausibly satisfy the separate-modifiability constraint, at least supposing they implement one of two overall design strategies. The first strategy is to limit the degree to which short-run evolutionary adjustments can affect a system's functional behavior. And the second strategy is to utilize functions that are integrated but still *loosely coupled*, or disproportionately effectible by alterations to their shared underlying mechanics.

References

Carruthers, Peter. *The Architecture of Mind: Massive Modularity and the Flexibility of Thought*. New York: Oxford University Press, 2006.

Conrad, Michael. “The Geometry of Evolution.” *Biosystems*, 24.1 (1990): 61-81.

Cowie, Fiona. “Us, Them, and It: Modules, Genes, Environments and Evolution.” *Mind & Language*¸ 23.3 (2008): 284-292.

Fodor, Jerry. *The Modularity of Mind*. MIT Press, 1983.

Hansen, Thomas F. “Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability.” *BioSystems,* 69.2 (2003): 83-94.

Kello, Christopher T., and Guy C. Van Orden. “Soft-assembly of sensorimotor function.” *Nonlinear Dynamics, Psychology, and Life Sciences, 13.1 (2009): 57-78*

Krohs, Ulrich. “The Cost of Modularity.” In *Functions in Biological and Artificial Worlds: Comparative Philosophical Perspectives*, edited by Ulrich Krohs and Peter Kroes, 259-276. MIT Press, 2009.

Philippides, Andy, and Phil Husbands, Tom Smith, and Michael O’Shea. “Flexible Couplings: Diffusing Neuromodulators and Adaptive Robotics.” *Artificial Life*, 11.1-2 (2005): 139-160.

Prinz, Jesse. J. "Is the Mind Really Modular?" In *Contemporary Debates in Cognitive Science*, edited by Robert J. Stainton, 22-36. Blackwell Publishing, 2006.

Samuels, Richard. "Is the Human Mind Massively Modular?" In *Contemporary Debates in Cognitive Science*, edited by Robert J. Stainton, 37-56. Blackwell Publishing, 2006.

Van Orden, Guy C., John G. Holden, and Michael T. Turvey. “Self-Organization of Cognitive Performance.” *Journal of Experimental Psychology: General*, 132.3 (2003): 331-350.

Schulz, Armin W. “Structural Flaws: Massive Modularity and the Argument from Design.” *The British Journal for the Philosophy of Science*, 59.4 (2008): 733-743.

Simon, Herbert A. “The Architecture of Complexity.” *Proceedings of the American Philosophical Society*, 106.6 (1962): 467-482.

Wilson, Robert A. “The Drink You Have When You’re Not Having a Drink.” *Mind & Language*¸ 23.3 (2008): 273-283.