Synthese (2018) 195:11–33
https://doi.org/10.1007/s11229-015-0713-5

Making mechanism interesting

Alex Rosenberg

Received: 9 December 2014 / Accepted: 3 March 2015 / Published online: 30 May 2015

Abstract I note the multitude of ways in which, beginning with the classic paper by Machamer et al. (Philos Sci 67:1–25, 2000), the mechanists have qualify their methodological dicta, and limit the vulnerability of their claims by strategic vagueness regarding their application. I go on to generalize a version of the mechanist requirement on explanations due to Craver and Kaplan (Philos Sci 78(4):601–627, 2011) in cognitive and systems neuroscience so that it applies broadly across the life sciences in accordance with the view elaborated by Craver and Darden in *In Search of Mechanisms* (2013). I then go on to explore what ramifications their mechanist requirement on explanations may have for explanatory "dependencies" reported in biology and the special sciences. What this exploration suggests is that mechanism threatens to eliminate instead of underwrite a large number of such "dependencies" reported in higher-levels of biology and the special sciences. I diagnose the source of this threat in mechanism's demand that explanations identify nested causal differences makers in mechanisms, their components, the components further components, and so forth. Finally, I identify the "love–hate" relationship mechanism must have with functional explanation, and show how it makes mechanism an extremely interesting thesis indeed.

Keywords Mechanism · Autonomy · Causation · Functionalism

This paper was inspired by arguments about the role of mechanism in developmental biology advanced by Charbel El-Hani in "Downward determination as a propensity changing noncausal relation" (El-Hani 2013).

A. Rosenberg (✉)
Department of Philosophy, Duke University, Durham, NC 27708, USA
e-mail: alexrose@duke.edu

> Thinking about mechanisms gives a better way to think about ones ontic commitments. Thinking about mechanisms offers an interesting and a good way to look at the history of science. Thinking about mechanisms provides a descriptively adequate way of talking about science and scientific discovery. Thinking about mechanisms presages new ways to handle some important philosophical concepts and problems. In fact if one does not think about mechanisms one cannot understand neurobiology and molecular biology. (Machamer et al. 2000, p. 25)

So concludes the most cited article in *Philosophy of Science* for the period 2003–2010: "Thinking about mechanisms."[1] One might be pardoned for hoping that a stronger set of theses might be found in the peroration of such a widely cited article. It is hard to disagree with these limited claims about the value of thinking about mechanisms. It is difficult to identify theses that might conflict with them. The conclusion of this paper suggests that the authors are "cagey" enough to recognize that claims about how science proceeds are subject to counterexamples and they don't want to be caught out over-generalizing.

Here is a much more interesting thesis defended by one of the authors cited above (along with a co-author):

> **3M** [for "model-to-mechanism-mapping requirement]
> In *successful* explanatory models in cognitive and systems neuroscience
> (a) the variables in the model correspond to components, activities, properties, and organizational features of the *target mechanism* that produces, maintains or *underlies* the phenomenon and
> (b) the (perhaps mathematical) *dependencies* posited among these variables in the model correspond to the (perhaps quantifiable) *causal relations* among the components of the *target mechanism*. (Craver and Kaplan 2011, p. 611, emphasis added)[2]

In short, explanation in cognitive and systems neuroscience should be causal[3] and should invoke mechanisms.

This claim is admirably clear but very narrow in scope, and intentionally so. The authors qualify it immediately: "This principle is restricted to cognitive and systems neuroscience and so allows that there are legitimate nonmechanistic forms of explanation," but presumably not in cognitive and systems neuroscience. However, they immediately broaden the initially qualified claim to include a much wider domain (in a footnote to this restriction): "…we see no reason to exempt all of cognitive science from the explanatory demands laid out [in **3M**] (Craver and Kaplan 2011, p. 611).

The question immediately arises: why restrict this thesis that explanations identify mechanisms merely to cognitive and systems neuroscience, or even cognitive science? Is there something special about these domains that demands a search for mechanisms? Is there something about some or all other areas of biological science that excuses their

[1] http://www.pitt.edu/~pkmach/

[2] I have italicized terms that will be important hereafter. The term 'dependency' will carry much weight.

[3] In what follows I will assume that causal explanation identifies causal difference makers. More on this in Sect. 3.

successful explanations from the demand that they provide mechanisms? On the face of it, if any discipline should be exempt from demands on successful explanations such as **3M**, it should be cognitive and systems neuroscience. After all this is the domain of the "hard problem," of qualia and consciousness, phenomena that resist causal explanations more firmly than all others. If there are good reasons cognitive and systems neuroscience must honor 3M surely the rest of the life sciences should also have to do so. High-level "dependencies," as Craver and Kaplan call them, are everywhere in biology, and for that matter in the special sciences. An argument that establishes a thesis such as **3M** in the domain Craver and Kaplan identify should secure support for the wider application of such a thesis.

Indeed, a more recent work by two of the authors of "Thinking about mechanisms" does broach a much stronger thesis than their original paper:

> Across the life sciences *the* goal is to open black boxes and to learn through experiment and observation which entities and activities are components in a mechanism and how those components are organized together to do something that none of them does in isolation. (Craver and Darden 2013, p. 3)

And

> …biology has become *the* search for mechanisms…Biologists look for mechanisms because they serve the three central aims of science: prediction, explanation, and control. (Craver and Darden 2013, p. 6, emphasis added)

Further,

> One cannot understand biology…without understanding…how mechanism schemas are constructed, evaluated, and revised (p. 10).

How seriously are we to take the repeated use of the definite article in these statements? Do the mechanists assert that *the principle* goal of biology is the search for mechanisms? They do not! Or do they? Craver and Daren write, "Science…is not defined as the search for mechanisms; still much of biology is in fact driven by the search for mechanisms (Craver and Darden 2013, p. 7)." How much? Most of biology? Almost all of biology? The most important parts? What about the so-called special sciences? Are they significantly different in the roles they accord to mechanisms? Why stop at **3M?** Mechanism cries out to be generalized.

> What would a generalized version of **3M** look like? Here is a first approximation:
> **M.** In *successful* explanatory models in the life sciences (including behavioral and social science)
> (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism(s) that produces, maintains or underlies the phenomenon and
> (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism(s).

Notice how little has to be varied from **3M** to expand its claims: simply substitute 'life sciences' for 'cognitive and systems neuroscience,' and pluralize 'models' to accom-

modate the likelihood that some models will be realized by the working together of multiple distinct mechanisms. **M** also presumes that the life sciences all proceed by constructing explanatory models.[4] To reiterate the question put here to mechanists: Why not generalize from **3M** to**M**? Are there considerations that prohibit or excuse mechanists from doing so? Are there considerations that require they do so? The details of answers to these questions would make mechanism (much) more interesting.

In what follows I try to answer at least the first of these two questions on behalf of the mechanists, and to explore the strengths and limits of possible answers to the question.

1 Mechanism and interlevel causation

What if anything might obstruct the claims of **M**, that explanation of "dependencies" in biology proceeds by, requires adverting to, is ultimately conveyed by uncovering underlying mechanisms? Probably the most serious qualm about embracing such a thesis is its implications for the autonomy of higher-level[5] explanations—a widely held bulwark against "reductionism"[6] in biology and the special sciences. The 3M thesis stated above appears to be the requirement that explanation in cognitive and system neuroscience be reductive—that higher-level models, especially ones involving functional capacities be "cashed in" for lower level component mechanisms plus their organization that provide a causal process realizing the model.[7] Generalizing this claim to M threatens the view that there are *autonomous* causal explanations in biology and the special sciences that are adequate, that the "dependencies" these explanations report do not require improvement, deepening, or that their character may even be harmed by such extensions.[8]

One venerable antireductionist objection to **3M** and its generalization to **M** stems from claims of multiple realizability of the kinds in higher-level biology and the special

[4] Let's not demand a very strong or contentful notion of 'model,' but simply accord the title 'model' to whatever is recognized in a special science as an explanatory representation of a regularity or regularities, what **3M** labels a "dependency".

[5] A word about the use of 'level' in this paper. I follow mechanists in not elucidating this term. As Craver (2007) notes, the notion cannot be adequately defined in terms mechanists generally employ such as component, constitution, composition, spatial demarcation or size. For purposes of this paper, a level will be the domain of a scientific (sub)discipline described by the proprietary vocabulary of that (sub)discipline.

[6] The term 'reductionism' is of course widely used by non-philosophers to label what philosophers more precisely lable as 'eliminativism'—the replacement of higher-level level explanations by lower level ones. Reductionism strictly so called holds that lower level explanations preserve whatever explanatory power a higher-level explanation has, while improving them in various ways—increasing their precision, underwriting the reliability of their dependences, etc.

[7] Cf. Piccinini and Craver (2011), Craver and Kaplan (2011) argue that these dependencies are not explanatory at all.

[8] Cf. Kitcher (1984): "plugging a molecular account into [the cytological] narratives [of meiosis] would decrease the explanatory power of those narratives." Quoted in Rosenberg and Arp, p. 218. Notice the expression "plugging in" that makes it clear Kitcher's objection is to reductionism, not eliminativism. See footnote 1.

sciences. If multiple mechanisms realize the antecedent property and the consequent property of a special science dependency, then **3M** and**M** will require too much: the provision of a vast disjunction of realizations or implementations of the antecedent and the consequent kind by inputs and outputs of an equally vast disjunction of mechanism. This will not be a serious problem for mechanism when *explananda* are singular events in which what is to be explained is brought about by the operation of a particular mechanism. In such cases the task is to identify the particular mechanism that was in play, whether or not a "motely" of other mechanism might also have done the job. But **3M** and **M** are about explaining 'dependencies'—regularities, repeated sequences reported in higher-level sciences. Biology is rife with "model systems"—specimens from which the scientist aim to generalize. Even in these cases the objection to **M** from multiple realizability may be overplayed: in many parts of biology, at least, where there is a disjunction of underlying mechanisms, they differ from one another along tractable dimensions, sometimes even quantitatively, in ways that can be neatly managed by explanations that honor requirements like **M**. In fact this will almost always be the case in the biological realm, since convergent evolution on identical functions by widely different mechanisms is rare.[9] (this is a point to which we shall return). In any case, hereafter I will assume that multiple realizability *per se* is not an insoluble problem for **M**.

The threat that mechanism really faces is the prospect that imposing **M** prevents mechanism from doing justice to a lot of explanatory dependencies that have emerged in the life sciences and the special sciences. Mechanists need to find a way to impose **M** on many domains of inquiry while at the same time according the dependencies these domains have discovered and report real explanatory authority. Mechanists don't want to be eliminativists. Can they avoid it?

To avoid the threat of eliminativism, mechanists need to show how higher-level models and the explanations they figure in have real explanatory force independent of a specification of the mechanisms that underlie them. One way to do this might be to insist that high-level causation is autonomous from lower level causation. If we could show this consistent with **3M** or other variants of **M**, the mechanist will be able to honor the explanatory power of higher-level models by holding that these models identify the distinct higher-level causes. The danger in this strategy for mechanism (or any one seeking to defend antireductionism) is that it may commit one to an "emergentist" conclusion: that higher-level causes are not only distinct from lower level causes but not even explainable by them.[10] Such a conclusion would be hard to reconcile with **3M** or **M**. The mechanist would have to admit that there are higher-level causes not fully explained by lower level mechanisms while demanding that such mechanisms be provided. The obvious retort by mechanism's skeptics would be "Why are the lower level mechanisms required if the higher-level processes are not explained by them?"

[9] Here and elsewhere in this paper the pertinent notion of function is that of the 'selected effects' etiological concept. For a detailed exposition and defense of the ubiquity of selected effects functions see Neander and Rosenberg (2009).

[10] This is of course what functionalists hold. See for example Weiskopf (2011a) for a recent statement that I treat briefly below.

Some mechanists have made proposals that could thread this needle: reconciling autonomous higher-level causation with requirements like **M** without making such higher-level causation mysterious. Writing with Bechtel, another strong advocate of mechanism, Craver has advanced such an argument seeking to show that mechanism is compatible with higher-level causation. If this is right, then dependencies that obtain at higher-levels may have a role in at least some fully adequate explanations at their levels. In that case, explanations honoring requirements such as **3M**, and its generalization, **M**, could be treated as improvements, deepening explanations already adequate by the standards of the special sciences that formulate them.

What is more, if demands like **3M** or **M** can be satisfied by explanations that don't involve upward causation from lower levels to produce higher-level causation, it will remain open to hold that the higher-level causal sequences really are autonomous from lower level causal sequences. This would be the best assurance of the autonomy of explanations that report higher-level causal sequences from deeper explanations that appeal to the mechanisms that bring about the higher-level regularities.

Bechtel and Craver in fact seek to do this. Beginning with the unargued assumption that higher-level causation obtains, they have claimed that mechanism rules out both downward causation from wholes to parts and upward causation from parts to wholes (Craver and Bechtel 2007, p. 548, and also Bechtel 2008). Instead, what appears to be "interlevel" (upward or downward) causation is always just a combination of "constitution"[11] plus *intra*level (same level) causation. Thus, high-level causation is preserved, independent of lower level causation and provides a basis for autonomous higher-level explanations. Demands such as **M** are preserved since the higher-level causation, though independent of lower level causation, consists in lower level causation plus (hereafter '+') constitution of higher-level objects and processes by their mechanical components suitably arranged. Craver and Bechtel write, "We assume that there are higher-level causes and, further, that all higher-level causes are fully explained by constitutive mechanisms." (p. 548). So, distinct higher-level causes exist but they are fully explained in ways that honor **M**. Just what we need.

This proposal raises some hard questions for mechanism. Consider the question how exactly can higher-level causation result from lower-level causation together with the constitution of wholes by the mechanisms that compose them? For example, how can there be causal regularities about Mendelian genes? One obvious answer is that causation at the lower level among the macromolecular assemblies that together constitute the Mendelian genes, give rise to the causal regularities at the higher-levels. How does lower level causation together with constitution create a novel, distinct and different explanatorily autonomous causal process at the higher-level? If lower level causation + constitution can' create distinct and different upper level causation, then the latter must be strongly "emergent" (i.e. metaphysically novel *sensu* Kim 2006 and not just epistemically unexpected). That's what distinct and different come to. This would surely be an unattractive conclusion for defenders of **M** or for that matter **3M**. Honoring the demands of **M** would not provide a complete explanation

[11] 'Composition' would be a more accurate term than constitution. I will use the mechanists' term with the understanding that what is really involved isn't the sort of 'constitution' issues that interest metaphysicians dealing with identity problems.

of the dependencies in higher-level models, since it would not capture the distinct causal process the higher-level models identify. But if there is nothing distinctive about higher-level causes it's tempting to say that lower level causation + constitution may simply create the illusion of upper level causation, and with it the illusion of upper level explanation. Why call the upper level causal processes illusory? If lower level causation + constitution suffices for upper level causation, then there is so to speak nothing left for upper level causation to do. Processes at the higher-level make no additional, distinctive, separable, additional, needed contribution to lower level causation in effecting higher-level regularities.[12]

Downward causation would give us a reason to accord real existence to upper level causes. Upward causation would accord distinct existence to upper level effects as well. But Craver and Bechtel describe such intralevel causal claims as "erroneous" (Craver and Bechtel 2007, p. 555) inferences from constitution + intralevel causation. Causation at higher-level j consists in causation at lower level i + constitution of the higher-level j causal mechanisms by their level i components suitably arranged. Notice that on Craver and Bechtel's view, it will turn out that causation everywhere consists only in causation at some basement level, plus the constitution of every other causal process everywhere out of basement level causation. Of course, if there is no basement level, Bechtel and Craver will have to recon with a "causal drainage" result (Block 2003; Kim 2003).

In any case, intralevel causation + constitution doesn't actually capture the phenomenon Bechtel and Craver think "erroneously" described as upward causation. To see the problem consider the mechanistic process described erroneously or not by the expression "upward causation:" Consider for example, the digestive system and the circulatory system, each composed of a large number of components. Changes in the digestive system cause changes in the circulatory system. This happens when an input to at least one component of the digestive system brings about a sequence of changes in its other components. As the immediately effected component of the digestive system changes, there is an instantaneous "Cambridge" change in the entire system. But this change is not immediately followed by a change in the circulatory system or any of its components. That latter occurs only when some (other) component of the digestive system has an output that counts as an input for some component of the circulatory system. This change in a component of the circulatory system is also accompanied by a "Cambridge" change in the entire circulatory system. But it is only once (at least the initial) output of the (last component of the) digestive system has worked its way through the components of the circulatory system that changes in the latter system as a whole count as the relevant effects of changes in the digestive system. This is the process that the expression "upward causation" is intended to identify. Is it erroneously so-called, as Bechtel and Craver allege? Instead of calling it erroneous "upward causation" it seems more appropriately described as intralevel (between components) causation + composition. The trouble with such a description for mechanism is that it makes no real room for higher-level causation at all, and thus provides no basis for explanatory autonomy of higher-level dependencies.

[12] This is the threat philosophers of mind face when they seek to accord causal force to intentional states while embracing physicalism.

Of course upper level causation would be distinct from lower level causation if there is some further causal factor involved—some upper level conserved quality, some "biff" or "umph," some productive power and its exercise distinct from lower level causation and created by lower level causation + constitution. This upper level property would distinguish upper level causation from lower level causation and from the sort of upward causation described above. But as a basis for conferring autonomy to higher-level explanations, its price is too high. For then there would be (metaphysically) emergent properties of whole mechanisms not constituted by properties of and relations among their components, something no mechanista countenances. We shall see exactly why this is so in section 3 below.

Preserving any explanatory autonomy for higher-level "dependencies," to use the language of **3M**, by denying interlevel causation is going to be difficult for mechanists. If mechanism is to grant that the higher-level dependencies are causal while requiring that their causality be grounded in causal relations among their mechanical components, a quite different approach will be required.

2 Why mechanism threatens to eliminate instead of underwrite

Actually, mechanism's problems with the preservation of higher-level dependencies are much graver than the merely philosophical puzzle of reconciling autonomous higher-level causation with **M** and its instances. To see how grave consider the mechanista's unusually strong credo regarding the unity of biology:

> The integration of biology is forged by building mechanism schemas that span many different levels, bridge across many different time scales, and that satisfy evidential constraints from many areas of biology (chemistry and physics too). From the perspective of a given phenomenon, one can *look down* to the entities and activities composing it. One can *look up* to the higher-level mechanisms of which it is a component. One can *look back* to the mechanisms that come before it or by which it developed. One *can look forward* to what comes after it. One *can look around* to see the even wider context with which it operates. The adequate explanation of many biological phenomena requires describing a temporally extended and multilevel mechanism. This is why many fields, working at multiple levels, often must integrate their work in the discovery of mechanisms. (Craver and Darden 2013, p. 163)

The word 'many' appears three times in this paragraph. One might pause here and ask about the qualification: why is the claim made about "many," and not "most" or "all" fields of biology? After all, one can "look" in all the directions specified in every field of biology and the special sciences. Are the mechanists just being "cagey"? Are there some biological phenomena to which **M**, our generalization of **3M** doesn't apply? Are their some fields that are not to be integrated with others by the discovery of mechanisms, or not to be integrated at all? Are there some mechanisms that are free from the constraints of chemistry and physics, along with those of biology? These are not rhetorical questions. They are raised by the qualifications of this passage, and the

failure to address them would make mechanism a claim so vague and cautious that it lessens its interest for the philosopher and reduces its guidance for the scientist.

But there is a crucial issue to be faced even by the weaker claim that many, though perhaps not all, fields are integrated by mechanisms. It is by no means clear that many of the higher-level dependencies of the special sciences can actually be integrated by mechanism. Rather, trying to satisfy the demands of **M** may lead to the realization that the dependencies at higher-levels are not causal at all, and therefore are not causally explanatory at all. Imposing requirements like **M** on special sciences are likely to eliminate some of their distinctive explanations altogether. This may be a much more disturbing outcome to most philosophers than the failure of demands like **M** allow for the autonomy of higher-level explanations.

What the provision of mechanisms in accordance with **M** shows is that many higher-level regularities generally fail to identify the causally relevant properties whose instantiation actually accounts for their *explananda* and so the dependencies these regularities report are not explanatory at all.

To see why, start with the following higher-level dependency.

> *Ceteris paribus*, an indigenous group's skin color darkness is inversely proportional to distance from the equator.

Here is a dependency uncovered long ago and given explanatory weight in a number of social scientific explanations. The difficulty it presents to **M** illustrates starkly the problem mechanists face preserving much received social science as explanatory. Acting under the orders of **M**, evolutionary anthropologists must set out to seek the mechanism that underlies this skin-color dependency. To begin with they have to recon with the fact that there are counter-examples to the dependency; Inuit peoples in the arctic, Berbers in the Atlas Mountains of Algeria, Cambodians in Indo-China. These counterexamples immediately suggest that distance from the equator may not be the difference maker for skin color. There is still another bit of data at least relevant to if not undermining of the skin color dependency: in every indigenous group, males are on average darker than females. This correlation of skin color and sex immediately suggested (wrongly) to European anthropologist (especially ethnocentric, not to say, racist ones) another dependency: women's lighter skin depends on male sexual preference, one that is itself explained by sexual selection, thus by a mechanism as **M** requires, albeit a completely wrong mechanism.

Here are the actual mechanisms that cause differences in skin color darkness. There are at least two:

First, unlike other mammals, humans need to synthesize vitamin D (especially D3) because they don't usually ingesting enough. There is a mechanism of vitamin D-synthesis that is driven by energy delivered by ultraviolet radiation from the Sun. The rate of synthesis depends on the number of photons that strike relevant molecules under the skin. Too much vitamin D is harmful as is too little. Accordingly natural selection has acted on skin color to regulate the synthesis of vitamin D by fine-tuning skin pigment to absorb photons and reflect them, depending on the evolutionary environment—very sunny has selected for darker skin, much less sunny has selected for lighter skin. The underlying Mendelian genetic mechanism itself is underlain by still another suite of mechanisms that disjunctively and conjunctively involve a large

number of alternative and interacting molecular genetic pathways, mainly regulatory ones that control the expression of structural genes, the ones directly responsible for the synthesis of melanin. The full details of the mechanism of optimizing ultraviolet radiation for vitamin D production is a matter of considerable but manageable multiple realizability.

Second, humans require folate for a variety of metabolic processes. However too much folate production results in reduced sperm motility and therefore in reduced male fertility. On the other hand insufficient folate results in significant birth defects and fetal viability. However, folate's precursor, folic acid, is subject to photodegradation, as a direct function of ultraviolet radiation. Accordingly there will be selection for a mechanism that regulates ultraviolet exposure of the folate synthesis systems to maintain different levels in males and females. In consequence, males of every group will be on average darker than females, since they would be harmed by too much folate, while females will be lighter owing to a need for more folate.

But what of the Inuits, Berbers, Cambodians and a dozen more groups whose skin color is a counterexample to the distance-from-the-equator dependency? A little additional information shows that none of them are counterexamples to the actual-mechanism explanation, and each of them confirms it. Inuits secure more than sufficient quantities of vitamin D from the high fat-content of their meat- and fish-rich diets (and suffer vitamin D deficiency diseases when their diet is changed by modernization); Berbers live in mountains that are much cloudier than surrounding deserts and therefore have much lower levels of penetrating ultraviolet; Cambodians migrated north from southern areas of Laos only relatively recently in human history, too recently for natural selection of optimal skin color to have caught up yet. So on for all the other exceptions to the original dependency.

The time required for modulation of skin color by selection is estimated to be quite short: on the order of 10,000 years in some cases. Owing to our African origin the process begins with all *Homo sapiens* having dark skin. Any one shade of skin color is multiply realizable by different combinations of a large number of structural and regulatory genes. There several genes that are parts of a complex network that controls melanin production alone. Among these genes a large number of mutations are possible (and have been actual), any one of which can (and has) effected skin color. All of these mutations have been channeled by local environments into mechanisms of skin pigment production that regulate ultraviolet transmission through the skin to produce levels of vitamin D and folate that are optimal for males and females depending on local sun-light levels at the surface of the earth.[13]

So, the dependency that skin color darkness is inversely proportional to distance from the equator turns out to roughly be correct, even though the presumptive explainer in the dependency—distance from the equator—does not identify a causal difference.

[13] It is important to notice that though natural selection plays an important role in this explanation, it is not among the mechanisms bringing about differences in skin color, for the simple reason that natural selection is not a mechanism, it is a process. One very obvious reason to disqualify natural selection as a mechanism derives from the mechanist's recognition that mechanisms are always mechanisms for something or other, whether that some thing is an outcome or effect or whether it is the process in which the mechanism engages. But natural selection is famously not for anything at all, there is no output it is organized to attain. See Garson (2013). The fact that natural selection is not a mechanism will be important hereafter.

These explainers are respectively the "ultimate" process of natural selection that drives the evolutionary outcome and the "proximate" physiological mechanisms that drive the production and maintenance of skin color in individual *Homo sapiens*. If the presumptive explainer of skin color—distance from the equator—isn't the actual cause, doesn't figure in the mechanisms of skin color production, what work is it doing and how can its explanatory role be preserved? You may ask, why should it be preserved at all, now that we know the real mechanisms?

Why suppose closeness to the equator plays any role in explaining skin color darkness? Well, the dependency does support a few counterfactuals about actual and possible groups of *Homo sapiens*: In the nearest possible worlds to ours, where a group indigenous to equatorial climes in the actual world is moved to the antipodes, their skins still dark. But that's because just moving a group can't change their skin color production mechanisms at all, and so can't change their actual skin color immediately; similarly, while no group has occupied Antarctica in the actual world, in close possible worlds a (possible) group of *Homo sapiens* that does indigenously occupy Antarctica has light skin color. But this is because it has the same mechanisms as other *Homo sapiens* do, and therefore has subject to selection for light skin color. Note the two counterfactuals in question are true entirely owing to the operation of the same evolutionary mechanism and the same physiological mechanisms in these close worlds as operate in the actual world. They are not true owing to any facts about distance from the equator. In every nomologically possible world, organism with a fixed mechanism of skin production will continue to have the same amount of ultraviolet radiation absorption and reflection rates no matter where they are placed with respect to the equator, or for that matter the orbit of the planet and the spectrum of its star.

When we try to honor **M**, to explain the distance-to-the-equator dependency that obtains in the actual world by the relevant mechanism, we discover that the antecedent of the dependency doesn't have anything causal to do with its consequent at all. What does bring about its consequent are facts about the evolutionary and physiological mechanisms that bring about skin color optimal for local ultraviolet exposure. "Distance from the equator" doesn't describe the causal difference maker in skin color. All it does is describe a distribution or cline in the incidence of a quantitative trait. If we substitute for the description of that cline, a description of the distribution of differences in the mechanism of skin pigmentation as ambient ultraviolet radiation vary, we will have identified the causes of skin color differences.

What explanatory role is left for 'distance from the equator'? Well, it does identify the cline and does so in possible worlds very close in evolutionary history and physiological mechanisms. But by following the dictates of **M**, we haven't in a way underwritten its explanatory power, shown why distance from the equator causally explains skin color. We have not located a mechanism that takes distance from the equator as an input and gives skin color darkness as an output. Satisfying **M** has resulted in the elimination of the distance-from-the-equator dependency in favor of one that does identify causal difference makers for skin color:

Skin color varies as the local amount of ultraviolet radiation, *ceteris paribus.*

Here the *ceteris paribus* clause does real work because with the details of the evolutionary and the physiological mechanisms in hand, we can start to explain exceptions, seeming counterexamples to the real dependency. By simply following the mecha-

nism that **M** demands we uncover inputs to the mechanism such as vitamin D-rich diet, heavy persistent cloud cover, recent local migrations, the genetics of albinism, that result in the counterexamples to the distance-from-the equator dependency.

What we can't do is treat the original claim, that skin color varies as the instance from the equator as reporting a dependency at all, certainly not a causal dependency. **M** demands its elimination from the stock of explanatory dependencies.

Well, good riddance you might say, no one should ever have taken it seriously to begin with. The trouble is the same argument (though with less of the mechanistic detail already filled in) can be run for many higher-level dependencies in biology and for many of the dependencies about *Homo sapiens* in the special sciences. For example, any regularity about a particular species is subject to exactly the same problem: it can't be accepted by **M** as explanatory at all. The regularity will be shown by steps that **M** demands, to fail to identify a causal property, a difference maker of the sort functional stratification requires. Try explaining why all swans are white or ravens black or for that matter why all humans have 23 chromosomes, *ceteris paribus* in a way that honors the demands of **M**.

The problem wont really be multiple realizability—the fact that there are many different ways of being a swan, a raven, a human, white, black or 23 chromosomed. There may be, but these ways will probably differ from one another in ways we can manage. We may be able to order different mechanisms along a manageable number of dimensions in which differences in mechanism don't make a difference in output for a given fixed input. For example, we can do this with the myriad possible mutations among the many genes that produce skin colors. This is why multiple realizability isn't much of a problem for reductionism—the thesis that lower level explanations preserve, while deepening higher-level ones.

The problem that the demands of **M** seems to reveal is that higher-level kinds rarely pick out the actual causal difference makers for their consequents. The explanatory kind terms of much of higher-level biology—systematics, paleontology, ecology, and the special sciences are more like "distance from the equator" than they are like "ultraviolet radiation modulation."

Consider the alleged dependency reported by *All swans are white, ceteris paribus*. The regularity does support a few counterfactuals: had a particular actual swan been born at a different time or kept in the dark, or fed a richer diet, it would still have been white; similarly any possible but non-actual offspring of two actual swans who don't mate would have been white too. But following out the demands of **M** reveals that being a swan, being a member of the species *Cygnis olor* isn't the cause of its whiteness at all. Consider the counterfactual that if swans had evolved in a significantly different environment, they would still be white. In light of the existence in the actual world of black swans and black-necked ones, this counterfactual is obviously false. It's false owing to the fact that the mechanism that makes most swans white might not have evolved in such circumstances. And this in turn suggests strongly that being a swan is not the cause of swans' whiteness, but that its some mechanism that produces white pigment in their feathers that do all the causal work. Notice how much more robust are the counterfactuals about any organisms with such physiological mechanisms even in worlds with wildly different evolutionary histories than ours.

More examples would be tedious, but symptoms of the problem are obvious in the multitude of *ceteris paribus* generalizations of these disciplines. When we follow **M**'s demand to identify the mechanisms that underlie many of these "dependencies," we discover they are not dependencies. We uncover the identities of the real causal difference makers that result in them, we discover the sources of their *ceteris paribus* clauses, and we explain the nature of their exceptions, and their counterexamples.

Now mechanism has become much more interesting. Adopting demands such as **M** threaten to undermine many dependencies at higher-levels of biology and the special sciences. At least it does so if as **M** evidently requires, the explanations are causal ones—ones that in the words of **3M** show the variables in the "dependencies…correspond to (perhaps quantifiable) causal relations among the components of the target mechanism" (Craver and Kaplan 2011, p. 611).

We need to reconsider the ambitions of mechanism as described in the passage from Craver and Darden (2013, p. 163) above. Mechanism's demand is certainly not going to "integrate" biology by "spanning many different levels." "Looking up" to higher-levels may reveal that there are no mechanisms at these levels. "From the perspective of a given phenomenon, one can *look down* to the entities and activities composing it," only to discover that the given higher phenomenon is epiphenomenal or worse entirely misdescribed. "Many fields, working at multiple levels, often must integrate their work in the discovery of mechanisms (Craver and Darden 2013, p. 63)." But doing so may eliminate previously accepted explanations instead of underwriting them.

3 What exactly makes mechanism interesting?

Why is it that many of the dependencies of biology and the special sciences are fated to elimination by an explanation that satisfies **M**? Why are their antecedents mostly like 'distance from the equator'? Why do their antecedents not identify the real causal difference makers for their consequences? The answer is at the same time obvious and seriously problematic for any attempt to reconcile mechanism with the explanatory autonomy of anything on offer in the so-called special sciences as well as higher-level biology. Mechanism is a very strongly physicalist thesis, with implications that make the preservation of higher-level explanations and autonomous dependencies of the special sciences that trade in nonphysical kinds unsustainable.

Mechanism starts with the uncontroversial assumption, broadly construed, that scientific explanation is causal. In particular, it accepts that the causes that figure in causal explanations are causal difference makers. But there is more to mechanism than causal explanation. If mechanism is right, if **M** is to be imposed across the board, at least in the biological domain, then the explanation of higher-level dependencies will bottom out (at least for the moment, if not forever) in "physical" mechanisms responsible for the physical changes. Almost all of these will be motions of the components of the successive layers of models that compose the system, the set of mechanisms or modules whose dependencies are to be explained.

Bechtel is typical of mechanists when he writes: " In most biological disciplines, both the phenomena themselves and the operations proposed to explain them can be *adequately characterized* as involving physical transformations of material sub-

stances…"(2008, p. 23, emphasis added).[14] Of course, as Craver and Darden write, "it's not part of our [the mechanists'] view that all explanations must bottom out in some privileged set of fundamental entities and activities (such as elementary particles and strings). Biological explanations rarely need to descend to the depth of quantum physics. As currently understood, most biological mechanisms are otherwise *insensitive to differences in particular details of the components at such very small size scales* [emphasis added]." Two things are clear in this passage: first the mechanists commitment to the adequacy of higher-level explanations—they need not "bottom out" in physics; second, their strong commitment to physicalism—differences between description are differences in abstraction versus detail of implementation.

This is a remarkably strong thesis. **3M** and **M** commit mechanists to the thesis the causal difference makers that explanations cite at every level of organization will also have to be physical properties. That is, in so far as there are levels of inquiry, with their own proprietary causal properties, each of these properties will differ from lower level properties of mechanisms that realize, implement, instantiate the higher-level dependency in only one way: they will have to be abstractions from lower level properties, abstractions that prescind from the details of their implementation. Such details of implementation are given by the succession of models of mechanisms and their components that honor principles such as **M** and **3M**. As Piccinini and Craver write, "…that's how mechanistic explanation generally works; it focuses on the mechanistic level most relevant to explaining a behavior while abstracting away from the mechanistic levels below (Piccinini and Craver 2010, ms. p. 44)."

The mechanists are fond of boxology, the invocation of boxes—black, grey, transparent. But the way they do this makes their commitment to physical difference clear. Craver and Darden (2013, pp. 89–90) write:

> A superficial, phenomenal model…describes the behavior of the mechanism without describing how the mechanism works…Incomplete schemas [that reveal the mechanism responsible for the phenomena are best thought of as explanation sketches. They have black boxes for components for which not even a functional role is known…. Sketches may also have grey boxes, for which a functional role has been conjectured…. The goal in providing a complete description of a mechanism is to fill in black and grey boxes…*every description bottoms out*

[14] As befits a mechanista, Bechtel is slightly cagey about the generality of this claim. When it comes to the mind, he waffles slightly:

> The performance of a mental activity also involves material changes, notably changes in sodium and potassium concentrations inside and outside neurons, but the characterization of them as mental activities does not focus on these material changes. Rather, it focuses on such questions as how the organism appropriately relates its behavior to features of its distal environment…The focus is not on the material change within the mechanism, but rather on identifying more abstractly those functional parts and operations that are organized such that the mechanism can interact appropriately in its environment. Thus mental mechanisms are ones that can be investigated taking a physical stance (examining neural structures and their operations) but also, distinctively and crucially, taking an information processing stance. (2008, p. 23).

Has Bechtel surrendered a commitment to mechanism in this consession to information processing? Cf. Weiskopf (2011b).

at some point where the gain in detail makes no difference to the researcher [emphasis added].

They go on to note that mechanism's methodology is "a version of the familiar childhood game of iteratively asking "Why?," except in this case we ask, "And how does that work?" (p. 90) Note the italicized phrase. Identifying boxes and making them transparent is a matter of increasing the detail of the causal claim, not cashing one causal claim in for a new and different one.

A couple of examples will help see this. Optimizing the synthesis of vitamin D3 and folic acid is the causal difference maker in skin color. **M** in effect demands we identify the mechanisms that implement the realization of this property in particular circumstances.[15] These mechanisms will presumably identify somatic (or germ line) genes (depending whether the explanation is ontogenetic or phylogenetic) in terms of their products and pathways of melanin synthesis. **M** will require that the causal difference makers for these mechanisms and their components be characterized by lower level properties that give the details of their molecular implementations, and so on, as far down as a science is require to go in characterizing successive difference makers implementing higher-level properties.

For another example, consider Kaplan and Craver's discussion of the difference-of-Gaussians (DOG) model of organization of spatial receptive fields in vision. The DOG model is a well-established mathematical model of the organization of retinal ganglion cells. Retinal ganglion cells transform the graded membrane potentials generated by photoreceptors into output signals with a regular Gaussian distribution around the relevant ganglion cells. The Gaussian distribution systematizes a variety of experimental visual data. **3M** demands that the model be implemented, that neuroscience provide the specific way this causal difference maker—having output signals with regular Gaussian distributions—is realized. **3M** requires a more detailed physical description of the same difference maker. Kaplan and Craver report three hypothesized possible implementations, three different ways of realizing the same difference maker, and argue that one is a better candidate than the others: bipolar cells with narrow dendritic fields pooling inputs only from the centermost cone receptors in the population of photoreceptors. **3M** demands that if this is the mechanism that physically realizes the DOG model, then its implementation details—bipolar cells having narrow dendritic fields—needs to be given further implementation details via their components and organization, and so on until we move beyond details that neuroscience concerns itself with. At this point of course, molecular biology takes up the challenge of meeting **3M**'s demands.

The point is not just that since mechanisms are physical systems, as are all their components. The only way science can honor a demand such as **M** and **3M** is by identifying the same purely physical difference makers for its dependencies with more and more detail of implementation, detail that is explanatorily irrelevant at higher-

[15] As noted above, these mechanism will be disjunctive owing to the complex disjunctive pathways that move from different genes (and the mutations that produce them) to optimum vitamin D and folate synthesis. Multiple realizability begins at the first level of implementation in this case.

levels, and therefore so to speak gets crossed out, while leaving the same difference maker described in successively more abstract terms.

Mechanism enjoins us to fill in black boxes, turning them grey, and eventually transparent. Mechanism's boxology is sophisticated of course, but it rests on a very strong factual assumption: that nature is thoroughly modularized. These models will, as the passage from Craver and Darden (2013) cited above indicates, will cut across many different "levels" (in the sense of footnote above): modules will be composed of diverse components, of many different sizes from the genome to the ecosystem interacting across different time scales from the microsecond to the geological epoch, operating through physical, chemical, macromolecular, physiological, evolutionary and other processes. But **M** and **3M** do require thoroughgoing modularization just because of their insistence on iterated componential explanation. Any domain in which M and 3M are enforceable explanatory requirements will be will be domains in which modularization obtains. Mechanism have components, their components have components, and so on, all the way down. This is the basis of mechanism's boxology. But why suppose that mechanism's boxes—black, grey, and transparent—are ubiquitous, or wide spread, or frequently to be found in nature or even exist at all? There is some basis for this conviction in the life sciences: the pervasive role of natural selection, which requires modularization. Darwinian processes can only operate to produce the cumulative adaptations so widespread in the domain of biology if adaptations are independent of one another. Otherwise improvements in one trait that are selected for may result in increasing maladaptation in other traits. The evident perfection of so many biological traits is testimony to natural selections power to modularize (Lewontin 1978).

But it has not escaped the reader's notice, of course, that Darwinian processes that produce adaptations are thereby producing functions. Indeed, on the selected effects etiological conception of function, that is what functions are—adaptations. (Cf. footnote 9 above). Nor has it escaped mechanista's attentions that something counts as a mechanism only if it is a mechanism for X, where X is either some activity that X performs—digestion, respiration, meiosis, oxidative phosphorylation, or X is the effect or outcome of the mechanism's operation. But there is one more thing that is clear about X whether it is a process the mechanism engages in or the outcome of that process. X is almost always describable as a function of the mechanism in question.

So, if mechanism requires modularization, and modularization is driven by natural selection packaging processes into modules by shaping them to be or to confer adaptations, and if the only causal difference makers that demands such as **3M** and **M** countenance are physical differences, mechanism really does turn out to be a very interesting thesis, a rather strong factual claim as well as a controversial philosophical one.

The only way the explanation of a higher-level dependency can satisfy **3M** or more broadly **M** is when the kinds that figure in the dependency differ from those in the mechanistic explanation by being more abstract ones from which irrelevant details of implementation have been eliminated. For example, 'force' is a more abstract kind than 'gravitational force' or 'electromotive force.' 'Acid' is a more abstract kind than 'sulfuric acid.' Newton's second law does not specify what kinds of forces produce accelerations. So, the functional kinds of biology and the special sciences will have to differ from the physical kinds of their mechanistic explanations only along the dimension of abstractness versus detail. Under what conditions will this relationship between

kinds obtain? This is where multiple realizability comes into the picture. When the various ways in which a functional kind can be realized by physical mechanisms is not very diverse, or when the different physical mechanisms differ from one another along manageable dimensions (as for example they do in DNA sequences that realize structural or regulatory genes for specific protein products), it is arguable that the abstraction relation holds between functional kinds and physical ones. Otherwise it does not hold.

The upshot for demands such as **3M** or **M** is obvious. They can't tolerate much convergent evolution: convergent evolution produces functional kinds that are instantiated by radically different realizations, with quite different details of implementation, so different from one another that they may not share an abstract physical kind in common. When an environment begins to impose strong evolutionary constraints on pre-existing mechanisms, it may result in unmanageable multiple realizability. For example, it is sometimes said that the wing evolved independently 40 times in evolutionary history. If there were interesting dependencies about wings, demanding their explanation satisfy **M** would be hopeless since differences in the mechanism that explain flight don't differ merely in details of implementation. Consider birds' wings, insect wings, Pterosaur wings and bat wings. Of course the only generalizations about all wings are analytic truths. Vision is a more serious problem, especially for a claim such as **3M** about systems neuroscience. Suppose there were a great deal of convergent evolution in adaptation for vision that started from very different structures, in humans and birds for example. Then there would be dependencies about vision that really are multiply realizable (for a detailed argument to this conclusion see Weiskopf 2011a). Would they present serious difficulties for mechanism?

The emergence of multiply realized adaptations might not so troubling to mechanism, if it turned out that natural selection is a mechanism. For the explanation of the process that brought them about would be a thoroughly mechanistic one: if it were then uncovering the details of a course of convergent evolution would certainly honor a demand such as **M**. But natural selection is not a mechanism (Garson 2013). It is certainly not a mechanism for bringing about multiple realizations of the same functional adaptation. So, if it had produced a great deal of radical convergent evolution, a great deal more than seems to have characterized our planet, it would make life difficult for mechanism. More important, mechanists don't wish to deny that natural selection brings adaptations into existence, whose explanations and whose causes are, as Mayr (1988) long ago reminded us, "ultimate" and not proximate ones.[16] Insofar as natural kinds are individuated by their causes and effects, the kinds natural selection produces will presumably be ones with differential effects on reproduction and differential causes in variation and selective retention. But it looks like **M** and **3M** will require mechanists to repudiate "ultimate" explanations. And this for two reasons. First, ultimate explanations of dependencies will not be mechanistic. Second, mechanism demands that functional kinds turn out to be physical ones: the functionally individuated predicates in which higher-level dependencies are expressed "correspond to components, activities, properties, and organizational features of the target

[16] The ultimate explanation for the eye-spot of a moth's wing is given by an adaptational account of camouflage. By contrast, the proximate explanation is given by an account of regulatory and structural somatic gene expression.

mechanisms." That is, **M** requires that the terms, predicates, descriptions that report the functional facts are really just descriptions of (relations among) abstract physical properties. In effect, mechanism treats terms that describe the explananda of ultimate evolutionary explanations as descriptions of the explananda of proximate physical explanations. Indeed, if natural selection is also a mechanism, then the kind-terms in which its explanada are expressed also turn out to be just terms picking out abstract physical kinds.

Sometimes mechanists show awareness of their commitment to this thesis. Thus Picccinini and Craver insist that "Functional properties are an undetachable aspect of mechanistic explanations. Any given explanatory text might accentuate the functional properties at the expense of the structural properties, but this is a difference of emphasis rather than difference in kind. The target of the description is in each case a mechanism" (Piccinini and Craver 2010, p. 17). So, mechanism requires functional individuation to get started: mechanisms are always mechanisms for X, whether X is what they do or what they attain, and where X is a function. But mechanists don't take functions seriously as autonomous irreducible properties. They can't. They are committed to treating functional descriptions as descriptions of physical properties, mechanisms.

All this is just what philosophers of psychology and philosophers of biology have been arguing against for a couple of generations, at least as far back as Fodor (1974). If the mechanism of natural selection operating on this planet did not allow for convergent evolution, for mechanisms that differ in structure but have the same selected effects, the same functions, then the mechanists would have a significant factual argument that as a matter of fact functional terms as a matter of contingent fact pick out homogeneous classes of physical mechanisms in abstract terms that prescind from the details of implementation. They would in effect have a factual argument against Fodor and other exponents of multiple realizability. Do the mechanists hold to **M** as roughly an empirical hypothesis or resting on one, about the rarity of radical convergent evolution?

If they do, then mechanism does turn out to be an interesting theory, one making a strong factual claim about the nature of reality. And there will certainly be philosophers of psychology who demur from this factual claim.

David Weiskopf provides a recent and fairly prominent example of someone arguing in exactly this way, and in cognitive and systems neuroscience. He writes that functional models "provide legitimate explanations even when they are not sketches of mechanisms." For example, in some cases a "complex connectionist network as a whole carries out a complex function but the subfunctions into which it might be analyzed correspond to no separate parts of the network. So there is no way to localize distinct functions, but these network models are still explanatory…(Weiskopf 2011a, p. 33)." Further, cognitive models "need not map model entities onto real world entities, or model activities and structures onto real-world activities and structures. Entities in models may pick out capacities, processes, distributed structures or other large scale functional properties of systems (p. 35)."[17]

[17] Weiskopf (2011a) provides examples of genuinely different mechanisms constrained so heavily by evolutionary forces as to converge on the same functional outputs for given inputs in visual perception. Insofar as natural selection can only target real difference makers, these cases represent a challenge to mechanista's assimilation of a function to merely an abstract physical difference maker.

Mechanists must live in an uneasy "love–hate" relation with functionalism. Mechanism needs functionalism to get started: mechanisms are always mechanisms for X. One begins an explanatory task by identifying the process or product of the mechanism and then one seeks an explanation for it. But they cannot really take functionalism seriously. They obviously can't allow that the functional explanations in the special sciences are autonomous. That is a patent violation of **3M** and **M**. Mechanists cannot deny that functional descriptions identify the causal difference makers, since they insist that what functions describe are physical mechanisms and their components. Giving up functional analysis is to give up boxology. Merchanistas don't advocate that either. But the package of theses to which they are committed turn out to make mechanism an interesting thesis indeed, a factual claim about the rarity of convergent evolution, and a thesis that rules out ultimate explanations, and indeed evolutionary ones if evolution turns out not to be a mechanism.

4 Mechanism and functionalism: making mechanism really interesting

But now what do we do when we turn to disciplines in which the causal variables invoked by the proprietary variables don't appear to be selected effect biological functions at all. What will a demand such as **M** make of any explanations that are accepted in these disciplines?

Consider one of Jerry Fodor's favorite dependencies in a special science:

Gresham's Law: Bad money drives good money out of circulation, *ceteris paribus.*

At first bluish one would think Gresham's law is not likely easily to survive demands like M. Trying to identify the mechanisms responsible for this "dependence" is likely to lead to the conclusion that it isn't one at all, that it's a prime candidate for elimination, along with dependencies such as the distance to the equator theory of skin color. After all, without the *ceteris paribus* clause this regularity fails to support relevant counterfactuals and is no causal dependency, it is no law. For example, in the German hyperinflation of 1923, bad money—the stuff the government printed—was driven out of circulation since sellers would not take any amount of it for any commodity. Instead they would only accept "good" money—foreign notes, gold and silver coins, etc. This is the kind of thing that could lead a mechanist to accept with equanimity the failure to provide Gresham's "law" with a mechanistic model in accordance with **M**.

Of course we can add clauses to Gresham's law to exclude cases such as this: "Bad money drives good money out of circulation, if they are exchanged at the same price." But this qualification comes dangerously close to making the law a necessary truth and so depriving it of its explanatory power.

Suppose, however, that Gresham's law really is an explanatory dependency of economics. Then, we will need to honor **M**'s demand that a mechanism be provided for Gresham's law. The number and complexity of the mechanisms required by any explanation of Gresham's law will be very large: mechanisms that underwrite and explain the existence of money, differences between them, institutions of exchange, price systems. Each of these mechanisms will have components that are themselves complex combinations of the "mechanism" of rational choice operating within several different institutions in the economy. It is very difficult if not impossible to characterize

the antecedent of Gresham's dependency in just those mechanistic terms that reveal why "bad money" is the difference maker for the dependency's consequent. Once anyone tries to specify these mechanisms in accordance with **M**, the complaint of the exponent of the autonomy of higher-level explanations begins to sound more reasonable than it was regarding the explanation of meiosis in terms of molecular biology: "plugging a molecular account into [the cytological] narratives [of meiosis] would decrease the explanatory power of those narratives." *Mutatis mutandis*, plugging in all the mechanisms required to make the difference that bad money makes in driving good money's going out of circulation would decrease the explanatory power of the dependency.

Mechanists may shrug their shoulders at this complaint. After all, without the mechanisms, the dependency is pretty weak, relatively useless as a guide to monetary policy, and insufficient to explain the details of any case in which buyers and sellers discover the divergence between the face value and the exchange value of a coin or note. But then, this "don't care" attitude is one they will probably have to adopt for many dependencies in the special sciences. Substitute the actual mechanisms for these dependencies, we can explain what the dependency purports to explain with far greater precision and detail.

But suppose mechanists were committed to saving Gresham's law, and the host of other "dependencies" like it in the social sciences, as I presume most mechanista's would wish to do. Then they are going to have to accept a very strong, very substantive, highly controversial view about the special sciences, the dependencies they have uncovered so far and the ones that they may uncover in the future.

Consider the assumption in cognitive and systems neuroscience that motivates **3M**: according to the mechanist, there it's the fact that we can identify functions—selected effect functions—which serve as explanation sketches that are eventually filled out by specifying mechanisms. And the mechanist has the assurance of Darwinian processes at work phylogenetically, ontogenetically and developmentally, to shape mechanisms that achieve these functions and they do so by modules—relatively independent mechanisms. After all, the only alternative to selected effects' processes in producing functions are (a) Devine design and construction, (b) human design and construction, and (c) strong teleology. Since each of these has been ruled out as causal processes in the biological domain, the only source of function left is adaptation—the Darwinian process that produces the appearance of purpose without its reality.[18] All this is of course somewhat controversial in cognitive and systems neuroscience. But it will be much more so in the social and behavioral sciences. And yet applying **M** in these domains will require even stronger theses.

Recall the fraught love/hate relationship between mechanism and functionalism. For **M** to apply everywhere in the way **3M** is claimed to apply in systems and cognitive neuroscience, the causal difference makers in a discipline's dependencies will have

[18] It bears repeating that here the pertinent notion of function here is the 'selected-effects sense, and not for example, the 'causal role' sense. 'Causal role' functions are of interest in the life sciences only to the extent that they have an explicit or implicit evolutionary etiology, i.e. are also selected effects. See Neander and Rosenberg (2009).

initially to be functionally individuated, and their functional character will have to be the result of some kind of Darwinian process or human artifice.

Now, the proprietary vocabulary of all the special sciences is functional at least in this sense: they identify the kinds they describe mainly in terms of their effects (and more rarely in terms of their causes). This is for several reason: first, most nouns in most languages identify their referents in terms of their effects, often their effects on us; second, the items of interest in the special sciences are ones that are significant to us, that have effects on us. That's why we identify them and seek explanations of their existence. Functional explanations are the stock in trade of the social and behavioral sciences.

Imposing **M** across the board in the special sciences is going to make the problems **3M** faces in cognitive and systems neuroscience seem quite manageable by comparison. In cognitive and systems neuroscience there seems a great deal of reason to suppose that functions were produced by Darwinian processes, that they won't be unmanageably multiply realizable. Mechanism may feel comfortable dispensing with the strategy of Darwinian ultimate explanation as anything more than a preliminary heuristic for locating dependencies to be explained mechanistically. It's true that most of the ultimate explanations of organismal traits and behavioral dispositions may turn out to be "just so stories" or at best ones we can't prove to be otherwise. Mechanism may be able to live with a self-imposed limitation on taking evolutionary explanations seriously in the life sciences even as it takes them as starting points. Can it do anything like this in the special sciences? It will have to where recourse to human artifice—conscious design and implementation of functions is ruled out. And this will almost always be the case.

Gresham's law is replete with functional kinds. Indeed, the functional kinds in which it trades—'money'—'good' and 'bad', "driving out, " "circulation," are themselves composed of other functional kinds—"buying," "selling," "bank-note," "coin." And Gresham's law itself can't operate except against the background of institutions such as the market-price system, something else that exists only because it fulfills a function, one that no individual could intentionally have established or effectively manage (cf. the collapse of centrally planned economies). In biology where **M** works there are boxes, modularized, compartmentalized, nearly decomposable units carved out by processes that we hypothesize give the boxes functions. Perhaps there are not as many as a thoroughgoing application of **M** requires. But there are in the biological domain a fair number of boxes. As a bit of thought about Gresham's law will reveal, the kind of modularization required for an explanation honoring **M** to provide any illumination is not on the cards in economics or sociology to say the least. For there to be functions that enable us to "reverse engineer" the mechanisms and their components that realize their dependencies, there have to be only a limited number of processes that produce functions, and a limited number of mechanisms available to package together into any given function. We have already seen that this may not always be the case throughout biology, where there is only one process that can produce functions and modules that realize them: Darwinian natural selection. At a minimum, it would take a strong thesis of Darwinian cultural evolution about every function discharged by a social institution, practice, rule, norm, or the groups that employ them, to underwrite a demand such as **M** imposes on the special sciences. And even then, honoring **M** is going to preserve

almost nothing among the received hypotheses, dependencies, models and theories in the special sciences. Most of them will pretty much go the way of the dependency between skin color darkness and distance from the equator.

5 Conclusion

Like cautious philosophers in general, mechanists want to accept the actual explanations of the special sciences as adequate on the authority of the special scientists who advance them. Who are we philosophers to be skeptical of what the specialists in economics or anthropology praise as satisfying their adequacy criteria on explanations? The trouble mechanists face is reconciling the autonomy of these explanations with mechanism. In particular, it may be difficult to vindicate the dependencies that the special sciences and even some parts of higher-level biology identify by revealing how their antecedents are causal difference maker for their consequents.

Mechanists won't be able to save higher-level dependencies when the higher-level kinds aren't successively more abstract versions of lower level physical kinds from which irrelevant details of implementation have been expunged. Under what conditions will higher-level explanatory variables plausibly be treatable as identifying such mechanisms and their components at high levels of abstraction? Only when these variables identify functions, in particular, selected-effects functions not resulting from extreme convergent evolution. However, when a Darwinian explanation gives us any confidence that a higher-level dependency is explanatory, **M** will require that we treat the "ultimate" Darwinian explanation as merely the preliminary first step to a quite different purely physical explanation. Whatever the plausibility of this approach in cognitive and systems neuroscience, mechanists owe an account of how it works elsewhere or why it doesn't have to. And if mechanists can't provide a convincing reason why **M** doesn't apply throughout the life sciences, they need to give us a reason why **3M** applies in systems and cognitive neuroscience.

Whether they can do so or not, mechanism turns out to be a much more interesting than mechanists originally suspected.

Conflict of interest I have no conflict of interest to disclose.

References

Bechtel, W. (2008). *Mental mechanisms*. New York: Psychology Press.
Block, N. (2003). Do causal powers drain away? *Philosophy and Phenomenological Research*, *67*, 133–150.
Craver, C. (2007). *Explaining the brain*. New York: Oxford University Press.
Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *2007*(22), 547–563.
Craver, C., & Darden, D. (2013). *In search of mechanisms*. Chicago: University of Chicago Press.
Craver, C., & Kaplan, D. M. (2011). The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective. *Philosophy of Science*, *78*(4), 601–627.
El-Hani, C. (2013). *Downward determination as a propensity changing noncausal relation*. Ms.
Fodor, J. (1974). Special sciences and the disunity of science as a working hypothesis. *Synthese*, *28*, 97–115.
Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science*, *80*, 317–333.

Kim, J. (2003). Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research*, *67*, 151–176.
Kitcher, P. (1984). 1953 and all that: A tale of two sciences. *Philosophical Review, 93*, 335–373; Quoted in Rosenberg, A., & Arp, R. (2009). *Readings in the philosophy of biology* (p. 218). New York: Wiley.
Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, *151*(3), 347–354.
Lewontin, R. (1978). Adaptation. *Scientific American*, *239*, 212–228.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.
Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge, MA: Harvard University Press.
Neander, K., & Rosenberg, A. (2009). Are homologies (selected effect or causal role) function free? *Philosophy of Science*, *76*, 307–334.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analysis as mechanism sketches. *Synthese*, *183*(3), 283–311.
Rosenberg, A., & Arp, R. (2010). *Philosophy of biology: An anthology*. New York: Wiley.
Weiskopf, D. (2011a). Models and mechanisms in psychological explanation. *Synthese*, *183*, 313–338.
Weiskopf, D. (2011b). The functional unity of special science kinds. *British Journal for the Philosophy of Science*, *62*, 233–258.