Extended Cognition, Extended Selection, and Developmental Systems Theory

Rob Rupert, U. Colorado, Boulder

Presented at the 2011 meeting of the International Society for History, Philosophy, and Social Studies of Biology

I. Extended Cognition and Developmental Systems Theory

The hypothesis of extended cognition (HEC) holds that in a significant proportion of actual cases, human cognitive processing is partly constituted by states outside the human organism; on this view, much of our thought literally takes place beyond the boundary of skin and skull. Intended as a thesis in the philosophy of cognitive science, acceptance of HEC is meant to entail a scientific revolution – a paradigm shift away from the purported Cartesian basis of orthodox computationalism.

There has been no shortage of critical responses to HEC. Some detractors worry that HEC provocatively extends the honorific label ‘cognitive’ to external, causal factors, *sans* empirical or theoretical payoff. According to this critical perspective, proponents of HEC fail to explain why their research program amounts to anything more than the identification of causal interactions between an internal cognitive system and external noncognitive factors (tools used by the internal cognitive system, for example) – the sort of thing long countenanced by those who are not inclined toward HEC. After all, throughout its history, cognitive science has recognized *some* external contributors to the production of intelligent behavior – from the vision-facilitating light provided by the sun to the nutrition provided by food, from the graphemes on the page that reflect the light that provides input during reading to the parental speech that affects cognitive development. Perhaps, in keeping with the overarching situated movement in cognitive science, researchers should focus on a wider range of external contributors, and perhaps HEC helps to direct researchers’ attention to such factors. But the moderate, interaction-oriented view, bruited above, serves just as well to move cognitive science in the requisite direction. What additional theoretical or practical work does it do to claim that the external factors are *genuinely cognitive*?

A straightforward way to resolve this dispute might appeal to a well-supported demarcation criterion – a mark of the cognitive (Adams and Aizawa 2001) – that distinguishes the genuinely cognitive causal contributors from the remainder. This approach might favor HEC; it might turn out that among the external causal contributors, some – the sun, for instance – are humdrum external objects and states that provide input to the cognitive system, while others have a more theoretically privileged, distinctively cognitive status – of the sort traditionally conferred only on neurally realized computational states.

In other work (Rupert 2009, 2010), I have argued that the most promising demarcation criterion defers to what might broadly be called an ‘architecture’: the set of mechanisms that co-contribute, in interdependent, overlapping subsets to the production of a wide variety of forms of intelligent behavior. A persisting collection of integrated elements is the central theoretical construct of cognitive science, whether computational, connectionist, or dynamicist. Of course the collection of elements and mechanism takes different form in different models and modeling traditions, but throughout all of this variation, the overarching idea of an integrated, persisting system that “manages” interaction with the passing parade of external objects and processes is a constant. (Partly in order to avoid the term’s association with one particular computationalist tradition (Pylyshyn 1984), I prefer to speak of the cognitive system, rather than the cognitive architecture.) Causal contributors to intelligent behavior abound; the genuinely cognitive ones are states of elements and mechanisms that are part of the persisting, integrated cognitive system. Furthermore, I suspect it is rare in humans that external mechanisms appear as part of this integrated set; such cases certainly won’t lead to a revolution in the study of language acquisition, short-term memory, or perception, for example. Most of the external objects that actually contribute in problem-solving are either devices used to produce only a narrow range of cognitive explananda or are one-off, or at least short-lived, resources.

With the stage set and my cards on the table, I move to the primary question of this paper: Does evolutionary biology – developmental systems theory (DST), in particular – bolster the case for HEC? Might, for instance, DST successfully apply a principle for packaging biological units that also supports HEC? Such questions arise naturally, given the dialectical parallels between HEC and DST. Proponents of both positions appeal to symmetry, arguing that internal factors (neural or genetic) do not play the privileged role claimed for them by the dominant approaches in their respective fields. External factors contribute causally and nontrivially to the production of the phenomena of interest – either intelligent behavior or the distribution (and changes in the distribution) of phenotypic traits, as the case may be. The manner or importance of that contribution does not differ from that of the internal contributors, so the reasoning goes. Thus, we should place internal and external factors on par with regard to their theoretical roles – say, as bits of cognitive processing or as part of the replicating units the lineage-trees of which expand or contract as a consequence of natural selection. Moreover, proponents of both views face the task of delimiting the set of causal factors possessing the theoretical status normally accorded only to internal resources, while excluding other causal factors (although the Big Bang causally influences all biological and cognitive phenomena, no one, so far as I know, wants to treat it as part of the unit of inheritance or as a part of cognitive processing).

Such big-picture parallels tantalize, but, one should wonder whether DST has something concrete to offer to HEC. Comfort to HEC-theorists might come in one of two forms. Perhaps the proponent of HEC can draw strategic inspiration from DST, the latter providing a model of theoretical reasoning after which the HEC-theorist can pattern her own. Alternatively, the HEC-theorist might extract something more robust from the debate surrounding DST. In particular, the HEC-theorist might hope that, in light of the genuine insights of DST, evolutionary biology has settled on a taxonomic principle that can be applied directly so as to deliver extended cognitive systems.

II. Inheritance, Intelligence, and Systems Individuation

Both the HEC-theorist and the proponent of DST emphasize the broad range of causal factors that account for the explananda of interest. The HEC-theorist points to the wealth of external causes that contribute to the production of intelligent behavior. Proponents of DST draw our attention to the wealth of external factors that contribute to the appearance of species-typical traits – song patterns in a population of birds, for example, or size of fire ant queens (Griffiths and Gray 2004, pp. 412-413) – and more generally to the resemblances and difference in phenotypes from one generation to the next, and changes in the distribution of such traits over evolutionary time.

Recall now the delimitation problem outlined above: Why consider the relevant external causes to be part of single unit of theoretical importance in the science of interest (cognitive science or biology)? Here the response of the DST-theorist rests on a particular kind of theoretical grounding, a claim about inheritance:

The developmental systems view argues that we should redefine *inheritance* so that every element of the developmental matrix that is replicated in each generation and which plays a role in the production of the evolved life cycle of the organism counts as something that is inherited…Lineages can be selected for having good symbionts or being imprinted on a good host [external factors], and these features can persist for evolutionarily significant periods of time. (Sterelny and Griffiths 1999, p. 97)

The DST-theorist addresses well-recognized biological explananda: the distribution of phenotypic traits, including similarity from one generation to the next, and the changes in the distribution of those traits over time via natural selection. At least as importantly, in its explanation of these biological phenomena, DST invokes the standard explanatory strategy of evolutionary theory – an appeal to inherited differences in fitness – but adds the innovative claim that the unit of inheritance often includes external, nongenetic material (Griffiths and Gray 2004, 411-412). Proponents of DST argue, perhaps convincingly, that the orthodox, gene-centered view simply misappraises the empirical situation: the units of that get recreated, from one life cycle to another, frequently stretch far beyond the gene to the nongenetic chemical properties of the organism and into the environment. Moreover, such expansive units can also be subject to selection pressures. All of this innovation takes place *within the framework of widely accepted evolutionary theorizing; it applies the apparatus of inherited differences in fitness*. The DST-theorist accepts the well-entrenched biological concepts and argues that the empirical facts support a surprising application of these concepts; this leads to some important changes in evolutionary theorizing, but not a revision of the fundamental theoretical construct: inherited difference in fitness.

In contrast, the HEC-theorist has no such theoretical footing with reference to which she can support her proposed innovations. Her case rests on no well-entrenched theoretical framework. It must be granted that external causes and the structure of the external environment contribute significantly to the production of the intelligent behavior; but by the application of which well-entrenched and well-motivated theoretical construct do such external causes become theoretically unified with internal causal contributors to the production of intelligent behavior? Inheritance does not stand anywhere near the theoretical center of cognitive science. The concepts of computation and representation have been at the heart of much work in cognitive science, but these concepts are contested and their importance rejected or downplayed by many advocates for HEC. Thus, although the HEC-theorist might take some general inspiration from DST (“we would like our approach to succeed in *that* way”), the project peters out when the HEC-theorist turns to implementation.

As advertised, though, there is a second way in which the debate over DST might provide grist for the HEC-theorist’s mill. The project of which DST is a part presupposes a characterization of developmental systems themselves. As Griffiths and Gray say:

We have shown that natural selection can act on populations of developmental systems and give rise to adaptation, but in doing so we have assumed that developmental systems are the sort of things that can be counted, that they have clear boundaries, and that they do not overlap so much that they cannot be distinguished from one another. We now turn to justifying this assumption. (2004, 423)

Griffiths and Gray go on to offer some suggestions regarding such a principle of individuation, to do with shared fate, more on which below.

Before considering their specific proposal, though, and its relation to HEC, I should note a potential source of confusion. As is clear from the quotation above, Griffiths and Gray mean to characterize the developmental system, which they describe elsewhere in their paper as “the entire matrix of interactants involved in a life cycle” (ibid. 419-420). Nevertheless, immediately after introducing the issue of the individuation of developmental systems in general, Griffiths and Gray move to a discussion of a less than fully general issue – “what makes a repeated assembly a developmental system in its own right” (ibid., 423). I say “less than fully general,” because developmental systems include many resources (adult birds’ songs, pheromones in a nesting environment, and so on) that do not seem to be parts of repeated assemblies; thus, one might think a further or more general principle is at play in those cases. (Perhaps Griffiths and Gray would fall back on the individuation of developmental systems as the package of resources that reliably reproduces itself through a life cycle, but as we’ve seen, this inheritance-based approach is of no use to the HEC-theorist.) Perhaps, then, Griffiths and Gray propose to rely on shared-fate only with respect to the narrower question of what makes something a developmental system when it is part of the subclass of developmental systems the components of which are assemblies of smaller organisms (such as the cells that make up a body or the ants that make up a colony).

Exegetical niceties aside, Griffiths and Gray have offered a principle for piecing together parts into a biologically relevant unit – whether they mean to cover only organisms, only superorganisms, or the complete range of developmental systems. This principle should be of interest to HEC-theorists, especially given Griffiths and Gray’s announced intention: to characterize developmental systems, which category includes systems constituted by single organisms, traditionally conceived, together with nonorganisms in their environment (chemical gradients of pheromones, for example, surely aren’t organisms). This kind of case is central in the literature on extended cognition; thus, it’s worth giving the proposal a run for its money, even if Griffiths and Gray have unwittingly changed the subject from that of individuating developmental systems in general to the individuation only of superorganisms or iterated collection of organisms into assemblies.

As indicated, Griffiths and Gray propose a criterion of shared fate (borrowed from Sober and Wilson): “The ants in a nest and the cells in a human body have a shared fate not just with respect to one part of their activities, but with respect to all of them” (ibid., 423). Griffiths and Gray enrich this idea with further theoretical resources – the concepts of trait groups and of the suppression of competition – but the notion of shared fate is clear enough. Almost certainly, my liver and heart will not contribute to the reproduction of a new life cycle unless they both contribute. Bracketing recent work with artificial organs, the reproductive fate of humans hearts is normally tightly wedded to the reproductive fate of human livers.

Does application of this criterion further the goals of the HEC-theorist? If it yields extended biological systems, this would seem to enhance HEC’s plausibility substantially, for doing so would show that biological systems of the right grain exist; moreover, if such systems produce intelligent behavior in real time (it is not simply that cognition *now* is affected historically by external developmental resources that are part of the system subject to a shared fate), the inference to HEC seems irresistible.

Let us assume for the moment, then, that various components constitute a single cognitive system when those elements all contribute to the production of intelligent behavior in real time and they share a common fate. The problem for the HEC-theorist is that most of the (actual) systems central to the debate do not satisfy the shared-fate criterion; these systems exhibit significant asymmetries among their components, analogous to asymmetries resulting in reproductive separability in the biological context. In lieu of some more compelling criterion for the individuation of biological systems, these asymmetries undermine the case for HEC currently at issue.

Consider the example of an artist using a sketch pad to create drawings via a feedback loop (Clark 2003, 76-77). The artist begins a sketch by making preliminary figures. The results of these early strokes impinge on the organism, causing her to see the artistic possibilities a new light and thus to make different, often more sophisticated sets of new strokes; the cycle repeats, with the final art-object taking a form that the artist would not have envisioned without the use of the sketch pad as a tool. This is a fascinating process, in its own right. Nevertheless, humans can and do reproduce without sketch pads. Individuals and their sketch pads do not share a common fate. Additionally, it would appear that the standard examples of extended systems – those involving external language, mathematics, external memory storage, and nautical artifacts – share the same fate (pardon the pun) in the dialectic.

It would seem, then, that the individuation criterion, endorsed at least tentatively by leading DST-theorists – the shared-fate criterion – does not yield a HEC-friendly result. Advocates for DST might insist that phenomena of inheritance and selection alone legitimize a more liberal way of characterizing developmental systems. This approach provides no traction for HEC, though, given that inheritance is a biological, not cognitive scientific, concept.

To close this section consider that extended selection requires a stable environmental contribution, yet cognitive science is interested in abilities that can be exercised flexibly across a wide range of environments. This attendant temporal mismatch alone should make one doubt that DST – which is, in the first instance, concerned with phenomena on a large-time scale – has anything to offer HEC-theorists, who so often are concerned with fleeting, soft-assembled, one-off, and otherwise short-lived systems.

III. Responses to Stotz

In a recent paper, Karola Stotz (2010) presents a series of criticisms of one of my earlier discussions (Rupert 2009) of some of the issues covered so far. In this final section, I respond to those criticisms and comment on some further relations between DST and HEC.

*a. Screening-off*

Stotz (2010, 486) takes me to task for my appeal to screening-off. The basis of her criticism seems to be that in biological cases, application of Salmon’s screening-off criterion doesn’t have the effect of bracketing the contribution of nongenetic material.

This alone is no criticism of anything I say, for I did not have DST as my target. The question is, “What kind of relation must exist between a nontrivial causal contributor and instances of intelligent behavior in order for that contributor to count as genuinely cognitive?” To make the question urgent, I appealed to one kind of biological case in which certain causes (the eye color of my grandparents) would seem to be screened off (by the eye color of my parents), but I made no attempt to generalize that example into an attack on DST (in fact, DST had yet to be introduced). My point was that when screening-off occurs, it seems taxonomically relevant: the thing screened-off, even if it is a genuine causal contributor to the phenomenon in question, should not be considered part of a unified system comprising causal contributors that are not screened off. And my target here has ultimately to do with cognition, the screening-off of external contributors by sensory states. More narrowly, my concern was to begin the chapter by giving a counterexample to the simple-minded General Causal Principle, which takes any nontrivially contributing causal factor to be part of a system that produces the phenomenon of interest. At this point in my chapter, DST hasn’t yet been introduced, so Stotz’s response – pointing out that some nongenetic contributors are *not* screened off – is a red herring.

b. *Appeal to shared-fate*

According to Stotz, Rupert “either unknowingly or deliberately conflates [Sterelny and Griffiths’] discussions of group selection with their treatment of extended developmental systems.” I’d say it’s neither of the two; I knowingly co-opted the framework, following (and citing!) Griffiths and Gray (2004). (Furthermore, in an essay on which the 2009 book chapter was based I acknowledged the gap between application of the shared-fate criterion to questions about group selection and application to the more general question of systems-individuation; see Rupert 2008, note 6.) Griffiths and Gray say that they need a principle to individuate developmental systems. Griffiths and Gray introduce the criterion of shared fate to solve this problem, which – even if they get off track in the way suggested in section II above – they take to apply to everyday organisms, for example, the way in which leaves of a tree appear as parts of a whole tree or the way in which cells band together to form a human body (Griffiths and Gray 2004, 423). (Selection on individual trees and individual humans are not normally seen as cases of group selection!) I think where Stotz has gotten confused is in the way Griffiths and Gray borrow the material on trait groups and shared fate – which are, frequently enough, discussed in connection with group selection – to try to solve their more general problem of systems individuation (see their citation to Griffiths and Sterelny 1999 at Griffiths and Gray 2004, 423). Although, as argued in section II, they might have stopped short of full generality, they certainly had something more than group selection in mind. So, I was simply following the lead of Griffiths and Gray, asking whether the conceptual tools of trait groups and shared fate might shed light on the more general problem of systems individuation.

All that being said, I could just as well respond in a simpler fashion. I was trying find something, *anything*, in the DST literature that might help HEC-theorists. Mere causal contribution doesn’t give them what they need. Does the DST-related literature offer any principle that is more discriminating and that might be pressed into service by the proponent of HEC? My point is that regardless of Griffiths and Gray’s or Wilson and Sober’s intentions in proposing their individuative principles, those principles do not help to establish HEC. So, even if I were misinterpreting these authors, the point would be largely irrelevant, given the dialectic: the HEC-theorist still comes up short.

c. *The dynamics of evolution.*

According to Stotz, “Rupert seems to generally misunderstand the process of natural selection” (2010, 487). After quoting something close to the what I say above about the artist and sketchpad, Stotz says, “It is hard to be independent of resources you are affected by” (ibid.).

Stotz’s interpretation of the text strikes me as most uncharitable and somewhat confused. In a footnote to my claims about reproductive independence, I wrote “The form of independence I have in mind is not full-blown probabilistic independence but rather the more intuitive idea of there being a non-negligible probability of one thing’s reproducing successfully when the other does not” (2009, 114). And the careful reader will see that this squares perfectly well with my discussion of trait-groups and shared fate. I explicitly bracketed the use of trait-groups – in which the presence of a trait in some members of the groups changes the probability of reproduction in the others, who may not have the trait – as the basis of systems individuation, objecting that it leads to a profligate metaphysics, *unless* trait groups converge, which is what one finds in the case of common fates. A criterion of systems individuation resting on *mere* probabilistic dependence – in particular, that any two things that are part of the same trait group are part of the same system – would seem to yield a mess of unwieldy and overlapping systems. So, I moved to the shared-fate criterion, and two things *can* fail to share their fate even when the reproduction of one is probabilistically dependent on the presence of the other. Thus, the reminder in my footnote about the form of independence I had in mind. Yes, the average fitness of humans, as a population, would differ, were they lacking some of their external resources, and thus human rates of reproduction are not probabilistically independent of these resources. But humans are still reproductively independent of those resources, in the sense I explicitly identified in the text, the sense that’s relevant to the shared-fate criterion.

I think Stotz sees a contradiction (or at least some confusion) in my views about selection because she has taken her eyes off the dialectic. The question is how we separate mere causal contributors from the genuinely cognitive ones. Many causal contributors make a probabilistic difference. But attention to that property alone doesn’t yield manageable systems (cf. Griffiths and Gray’s worries at 423). The question is, among those external (or internal) resources, including those that make a probabilistic difference (whether at the population or individual level), what makes some of them genuinely cognitive and not others. Shared-fate is a reasonable criterion for the HEC-theorist to invoke, but it doesn’t yield a favorable result, or so I argued. None of this reasoning rests on a mistaken conception of natural selection.

Lastly, consider a pair of positive arguments for HEC that one might divine from Stotz’s work (and that of other DST-theorists). First, there is a frequent emphasis on necessary conditions: the suggestion often seems to be that something is part of the developmental matrix if it is a necessary condition for the appearance of a life-cycle (perhaps limited to certain properties, such as that it resemble a previous life-cycle). In order for this argument to get off the ground, proponents of DST must clarify the sort of necessity they have in mind. I don’t think they have in mind the full-blown modal sort; for there’s surely some possible world in which an organism’s life and phenotypic traits are brought about by the contribution of factors other than the ones active in our world. Of course, if one includes the external resources in question as part of the life cycle, then their presence is trivially necessary for the appearance of the life-cycle. In that case, however, the DST-theorist stands in need of a distinct rationale for including the external resources as part of the set of developmental resources that make up a single evolutionary unit. She may well have in hand such a rationale – to do with facts about inheritance and selection. In which case, appeals to the necessity of a resource’s contribution to the phenotypic outcome seem otiose. Moreover, the HEC-theorist needs her own, cognition-based rationale for such inclusion.

Part of my concern is a general worry about appeals to necessary conditions in discussions of systems individuation and extension (cognitive or biological). Why should anyone think that if A is necessary for outcome B, then on that basis alone, A is part of a theoretically important, integrated unit or a B-producing system? Or, put a bit differently, why should anyone think, on that basis alone, that A shares theoretically important properties (such as ‘being cognitive’) with all other necessary conditions for the appearance of outcome B. Perhaps clarification of the sort of necessity at issue will help to make such principles more plausible, but I have argued otherwise (2009, ch. 2).

Second, I’m leery of the DST-theorist’s emphasis on the inclusion of cognition-related resources in the developmental matrix, at least if this is meant to ground an argument for HEC. So far as I can tell, the inclusion of such resources doesn’t much support HEC. The explanatory project of cognitive science differs from that of evolutionary biology. Perhaps systems the integrity of which is validated by important theoretical machinery in biology include external resources that help to produce intelligent behavior. At the same time, from the standpoint of cognitive science – out to explain intelligent behavior rather than the distribution of phenotypic traits (or kinds of organisms or changes in either of these) – those external resources might not count as part of the cognitive system or as part of the cognitive processing that produces intelligent behavior. One way to see the potential for a gap here is to note that the typical HEC-theorist draws on real-time examples, not historical ones, and when the HEC-theorist does draw on historical examples, she invites the question of why those historical contributors count as parts of the cognitive system or part of cognitive processing. *Of course* my grandparents’ political discussions, overheard when I was a child, causally influenced my current thinking about politics. Why in the world should that make my grandparents part of my cognitive system? Sadly, they are no longer with us. In contrast, biology is interested in long-term cycles of development and selection, and that might make a better case for including cognition-related resources in the developmental matrix, even if those resources are not genuinely cognitive (as determined by cognitive science).

To anyone but the hard-core reductionist, this situation should seem, not only unobjectionable, but unsurprising. If cognitive science and biology are autonomous sciences (in that the laws of cognitive science don’t reduce to the laws of biology and causal-explanatory kinds of cognitive science can’t be identified with causal-explanatory kinds in biology), then one would expect the two sciences to carve up the world in different ways. The unity of the natural world is preserved on this picture because the resources appealed to in cognitive science – the matter and basic properties – are recognized by biology. At the same time, successful explanatory practice in cognitive science might – and I claim, does – draw a central theoretical distinction between sets of resources that, from a biological standpoint, should be treated as unified.