A BOOK CLUB ON CLARK BARRETT'S 'THE SHAPE OF THOUGHT'

The International Cognition and Culture Institute

March 2016

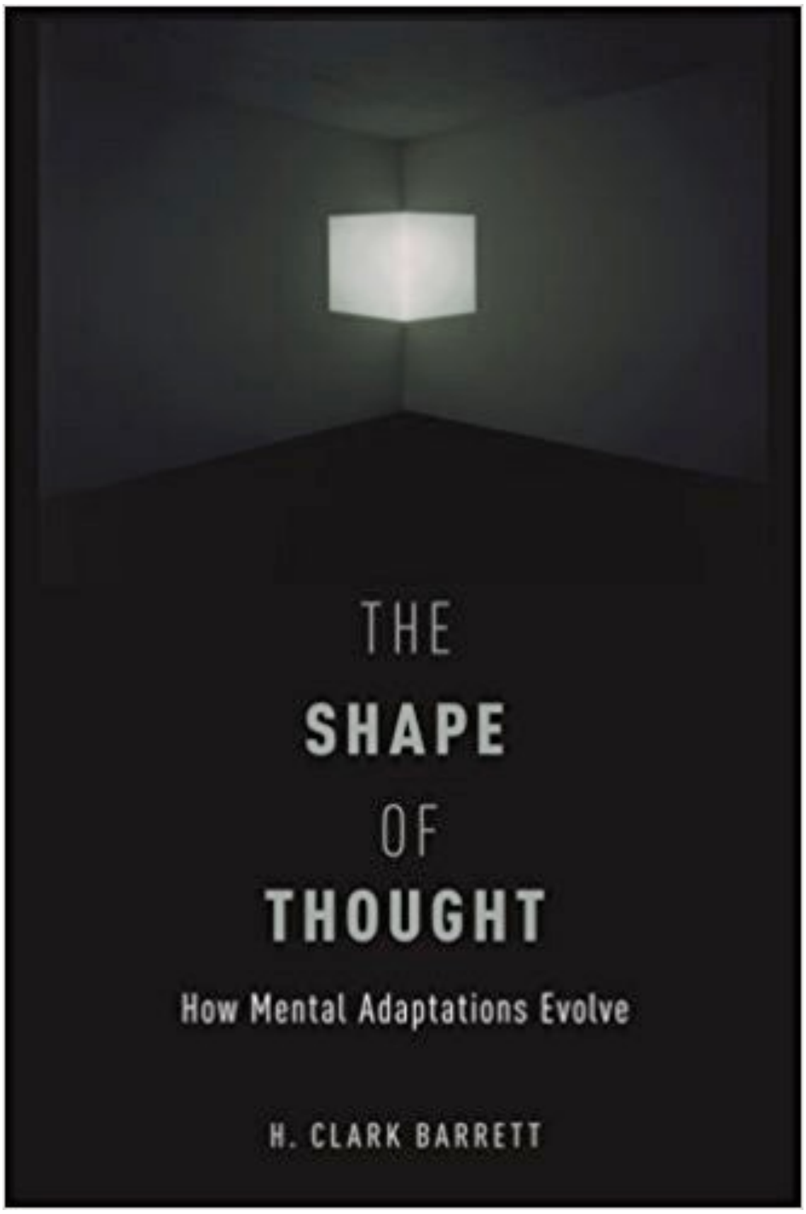
Clark Barrett, evolutionary psychologist and anthropologist, professor in the department of Anthropology at UCLA, published in 2015 a book, *The Shape of Thought: How Mental Adaptations Evolve*, which is both an introduction to evolutionary psychology and a major contribution to its development. This is an important and highly readable book that we strongly recommend to our readers. There was much to learn from it, and also much to discuss. We therefore opened a mini book club about it, which consisted of two reviews of the book, one more critical by Daniel Burnston, the other more favourable by Olivier Morin, and of Clark Barrett's response to these two reviews.

TABLE OF CONTENTS

1

THE EVOLUTION OF EVOLUTIONARY PSYCHOLOGY

By Daniel Burnston



"The Shape of Thought - How Mental Adaptations Evolve" Oxford University Press, 2015 (416 pages).

I am an anti-adaptationist, at least about the mind. I am also someone who has long been convinced by what many take to be the damning criticisms of evolutionary psychology (henceforth, 'EP') — the view that (to put it generally), we must and can only understand the organization of the mind in evolutionary terms. So it was with some trepidation that I agreed to review Barrett's book, and it was with some surprise and delight that I found it so insightful, erudite, and interesting. There is a lot of wonderful material in this book, even for those who don't agree with EP as an approach to the mind. And that's good because, despite the many agreements I have with parts of the book, I remain as steadfastly opposed to EP as I was before. I'll begin by saying what I take the valuable contributions to be, before explaining why I am still unmoved.

The broad project of the book is to update EP in such a way as to unwed it from a range of conceptual associations that have grown up around it. Barrett is against the view that "modular vs. non-modular, lower-level vs. higher-level, unconscious vs. conscious, automatic v. controlled, innate vs. learned, specialized vs. unspecialized, and evolved vs. something else" (p. 266) are all the *same* distinction, at least in terms of how they carve up the mind. Rejecting this view allows Barrett to argue that EP is compatible with big chunks of current evolutionary theory and cognitive science that are standardly taken as problems for, if not outright falsifications of, the program. Here, in no particular order, are some of the important conceptual moves made by Barrett. He convincingly claims that:

- EP is not committed to dual-systems views of the mind.
- "Preformationist" views of innate mental traits are hopeless, and that an appropriate view of EP must take into account both evolution and development.
- The context-sensitivity, plasticity, flexibility, and interactivity of cognitive mechanisms is incompatible neither with construing them as functionally specified nor as modular. Indeed, we should expect most computing mechanisms to exhibit both specialization and certain types of flexibility and context sensitivity.
- Probabilistic and inductive processes are vital to both evolution and cognition, and should be an important part of EP's story.
- Learning and cultural evolution are, in part, dependent on the "enabling" constraints present in the mechanisms that pre-date those processes.

Barrett argues that evolution is importantly *open*. It shapes probabilities, implements beneficial constraints, and grows systems incrementally rather than by incorporating entirely new, dedicated mechanisms for specific outcomes. This openness allows for EP to incorporate the above perspectives whole, without denying a role for evolution in cognitive explanation. On all these points, I am largely agreed (although I'd quibble with many specifics, for instance his way of "saving" modularity; see Burnston & Cohen, in Zeimbekis & Raftopoulos, 2015), and Barrett guides the reader through them with a steady hand.

So, with all this agreement, what's the worry? Unfortunately, these advances, impressive though they are, entirely fail to address what I take to be the biggest problem for EP—the commitment of the view to adaptationist accounts of mental traits.

There are at least three main lines of criticism that have been leveled against EP. The first is biological—EP is supposed to be committed to a type of degree of adaptationism that is unlikely to hold. The second is cognitive-scientific—EP is supposed to be committed to views of cognition/functional specialization that are false. The third is, I suppose, primarily philosophical—EP is supposed to offer explanations that fail to meet the standards for good explanations. One can find variants of these critiques throughout the literature on EP (including in the book length critiques by Fodor, Buller, and Richardson). My primary objection is a version of the third. With regards to the first two, my main problem with the book is that it can be frustratingly myopic when considering alternative theoretical perspectives. On the biological side: there is a significant debate in biology about to what extent phenotypic traits are due to adaptation/natural selection, versus those that are due to drift—random genetic, developmental, and environmental factors or events that, according to some, are capable of pushing traits towards fixation despite a lack of influence from selection (for discussion and elaboration of this view, see Kimura, 1979; Beatty, 1984; Millstein, 2002). Drift is only mentioned a couple of times in the book, and only obliquely: Barrett is willing to countenance drift as helping to produce genetic variation (p. 301) or helping to set up the environmental (including social) circumstances in which adaptation can then occur (p. 220), but drift is never considered as a serious alternative for how mental traits might arise.

On the cognitive science side, Barrett favors a model on which cognitive processes are hierarchical and sequential, saying that it is "generally agreed that processing in most domains is hierarchical, occurring in stages" (p. 107). On these kinds of views, cognition works via a large series of input-output devices (which Barrett sometimes calls "parsers"), each operating on the information passed to it by previous specialized mechanisms. This fits in well with his model of additive evolution, on which new functional parts are evolved to interact with what has come before in a way that is fitness-increasing (for the relationship between hierarchical mechanisms and additive evolution, see pp. 300–308). However, it is simply not "generally agreed" that *all* processing is hierarchical. I will discuss one case, visual processing, in detail below. Here are a few other examples. Barrett cites Treisman's model of attention, on which attention serves to highlight already-processed perceptual information. However, more current models suggest, as Mole (2015, in Zeimbekis & Raftopoulos, 2015) nicely summarizes, that "processes responsible for the allocation of attention [are] inextricable from the processes that are responsible for the perception of the things to which we attend" (p. 225). On these models, attention works concurrently with perception, by biasing competition amongst perceptual processes (see Desimone,1998; Duncan, 1998; Reynolds & Desimone, 1998.) In terms of language processing, for instance in reading a sentence, Barrett seems to favor a view on which individual pheonomic structures are parsed and then recombined, such that there is "a chain of processing necessary to build (extract, infer) … meaning from patterns of ink on a page" (p. 283). There is a long history, however, of "interactive" architectures for linguistic processing (see Samuel, 1997), inspired by phenomena such as phoneme restoration, on which phonemic representations are not universally represented prior to lexical ones. Moreover, Port (2007) has argued that lexical representations in memory are stored primarily as communicatively relevant chunks (phrases, etc.) rather than as discrete phonemes. Spivey (2007, ch. 7) argues that semantic context-effects on word recognition are incompatible with stage-based accounts of lexical processing.

One area in which the hierarchical view may be on better footing is in action-understanding, in which Barrett suggests that we perceive and understand actions by distinctly parsing their elements and subgoals and then combining those subgoals to represent the entire action. There is a huge amount of phenomenological and social evidence that we do in fact do this to some extent—we do tend to *explain* actions in this particular way, although I am skeptical of the claim that action perception works according to these principles (see below). However, it is worth noting that these considerations are often taken to support a hierarchical view of the motor systems that plan and produce actions, and this view has been heavily challenged. Uithol et al. (2012) argue that action-stages are abstractions imposed from outside by our social practices for explaining actions, and that they do not match the neural data (cf. Uithol et al., 2014). Moreover, Graziano (2002) convincingly shows that, in the motor systems of the brain, behavior is represented as concerted suites of motor action and not as isolated discrete elements. He further expands this into a general critique of hierarchical approaches to neural organization (Graziano & Aflalo, 2007).

Of course, this is not to prejudge the outcome of these debates. But the fact is that there *are* debates about the cognitive structure underlying each of these processes. While at times Barrett does attempt more sophisticated *versions* of the hierarchical picture (e.g., ones that combine parallel and serial processing, or ones that employ central "bulletin boards" with widely available information; see pp. 292-294), someone who finds the foregoing perspectives compelling (as I do) might be inclined to question the fundamental view of cognition that Barrett assumes from the outset.

I state these worries because they will come up again. As mentioned, however, my main criticism is about the explanatory role that is supposed to be played by adaptationism on Barrett's account. The standard explanatory critiques of EP are that adaptationist hypotheses fail to live up to appropriate standards of evidential support and testability (Richardson, 2007), and/or that they rule out genuine alternatives without sufficient argument (Lloyd, 1999). While I think these criticisms are relevant (see above), my version is slightly different, and is inspired more by another common type of claim in the book. There are a number of places in which Barrett does cite competing claims about cognitive architecture—Gauthier v. Kanwisher on face perception, debates about whether dynamical systems approaches deny functional specialization, and even Marr v. Gibson on visual processing—and argues that an adaptationist story will be available *however these debates come out*. In my view this is evidence that EP cannot play the primary role in the mind sciences that Barrett wants it to. Here's why.

Barrett commits to a set of claims that shape the argumentative pattern throughout the book. They are as follows:

1. The importance of mechanism: understanding the function of a trait or mechanism depends on understanding its causal roles in the system of which it is a part.

2. The adaptationist axiom: *all* mental features, like all other biological features, must be the result of adaptation by natural selection.

3. The "first law" of adaptationism: "It depends". The way that adaptation has shaped a particular trait or mechanism is entirely dependent on the details of the case.

4. The "centrality" of evolutionary explanation: We can only understand the organization of the mind in evolutionary (adaptationist) terms.

These four claims inspire an argument schema that occurs throughout the book. First, look at a particular mechanistic/information processing account of some trait (action understanding, vision, language, food concepts, theory-of-mind, etc.), as per claim 1. Second, cite the axiom (claim 2) to insist that we need an adaptationist explanation. Third, define an "adaptive problem" that might have been solved by having the mechanism function in the way posited by the mechanistic account (to cover claim 3 for that specific mechanism). Fourth, conclude that the adaptationist explanation is central to the understanding of the system (claim 4). My main concern is that, on Barrett's argument schema, *the adaptationist explanation adds little if any functional understanding to our view of mechanistic organization*.

Consider a mechanism for a psychological trait X, and assume that all are agreed on its current functional role—what it computes and how, and how those computations contribute to mental functioning in general. Suppose, however, that there are differing opinions on the evolutionary story. Perhaps theorists A and B agree that it's an adaptation, but disagree on what adaptive problems it was designed to solve. Theorist C, on the other hand, is convinced that X is entirely due to drift. The point is that, by the argument schema, we've already fulfilled the mechanistic understanding, and therefore the outcome of the evolutionary debate *makes no difference* to our understanding of the mechanism's role in cognitive organization. But, per claim 4, the evolutionary story is supposed to be central to this very understanding—Barrett states repeatedly that we must understand cognitive organization in evolutionary terms. The argument schema fails, on its own terms, to justify the centrality claim.

The reason for this is the combination of claims 2 and 3: the adaptationist axiom and the "it depends" law. Claim 2 says that for every trait there must be an adaptationist story. Claim 3 tells us that there won't be any specific ideas we can draw on in claiming what particular traits will look like and how natural selection shapes them (Barrett, of course, does offer some general characterizations of what natural selection does, but the details depend entirely on the case). Hence, in characterizing the trait we must rely almost entirely on claim 1, the mechanistic description. But because we must already have the mechanistic understanding to embark on the adaptive problem question, the adaptationist explanation in any specific case is entirely parasitic on the mechanistic one. And thus, I claim, it adds nothing substantial to it for understanding mental organization.

To see this clearly, consider just one case in detail—the hierarchical view of visual processing. Barrett argues (and indeed, this is a standard view), that vision "carves up visual processing into different tasks, with some coming before others in the processing and some coming later. Later processes operate on the results of earlier processes" (p. 66). On the hierarchical view, vision works by first having specialized feature-processors operating in parallel, which "tag" stimuli with a bunch of feature values (for edges, motion, color, etc.). Then, object perception tags these assorted features or "cues" as together constituting an object. Then, categorical object perception tags that object as being of a particular type ('face', 'lion', etc.). Similar stories are told for how we perceive animacy and facial expressions. However, the sequential hierarchy implied here has been questioned from a number of angles. Jonathan Cohen and I (Burnston & Cohen, 2013, 2015, drawing on a range of perceptual illusions, have argued that feature representations are not asymmetrically prior (either causally or temporally) to object representations. In general, feature and object processing are mutually constraining, and one only gets definitive object *or* feature representations as a result of the combined evidence for both feature and object categories. At the categorical level, Bar (2004) has argued that category assignment to objects relies on contextual cues present in the general scene, and hence that object categorization is in general a combined top-down and bottom-up process. In terms of visual neuroscience, Aflalo and Graziano (2011) have questioned the traditional idea of an anatomical hierarchy in extrastriate cortex, and I have critically examined the idea of a representational hierarchy (Burnston, 2015). Even some of the original and most influential proponents of the hierarchical view of visual cortex—David Van Essen and Daniel Felleman—have changed their minds. They no longer take the evidence to support a division of vision into serial hierarchical processing levels (Hegdé & Felleman, 2007; Hegdé & Van Essen, 2007).

It is entirely possible that all of these claims are wrong, and the hierarchical view in fact right. That's not the point. The point is that we have here a significant question about cognitive architecture, about which there are different competing accounts, and for which the argument schema provides literally *no resources for analysis*. Barrett's claims about the fitness-increasing benefits of dividing up vision hierarchically—for instance that it subsequently allows us to track object permanence by looking for spatiotemporal and/or property continuity amongst already represented features (pp. 70–72)—is based on the supposedly accepted fact that vision is divided up hierarchically. If the non-hierarchical view turns out to be correct, then perhaps another adaptationist explanation can be found for it (an account of object permanence is certainly still important on the non-hierarchical view, and may even still involve elements of the spatiotemporal continuity and property tracking accounts). However, *we'll already know* how vision is organized—non-hierarchically. The adaptationist theory won't make a difference to how we understand mental organization in this case. None of this is to deny, of course, that there is a question about how particular psychological mechanisms come about. That question may even be important and interesting in evolutionary biology. But not all historical questions about X are relevant to every particular explanandum regarding X (otherwise we'd cite the Big Bang a *lot*). I am suggesting that the evolutionary question is one such for explaining the function and organization of psychological mechanisms.

Good types of explanation get the facts right. But they are also, in an important sense, forward-looking. They articulate hypotheses that shape how we investigate the domain. What the above examples show is that Barrett's view of adaptationist explanation fails this criterion. On the schema, all the work is (and must be) done by mechanistic analysis. The adaptationist explanations are a cavalry that only ever shows up after the battle has been won. They're a bench player that you only put in at mop-up time. They're the confectioner's sugar sprinkled on a pastry—maybe nice to have, but ancillary and, ultimately, dispensable. From this view, the many places in which Barrett offers some (contentious) view of a mechanism, then switches to adaptation-speak with some variant of the claim that the mechanism *must* have evolved by natural selection to work that way, are extremely frustrating. It's not (just) that these claims lack evidence or rule out alternatives without argument. It's that they're simply unnecessary for explaining what we want to explain.

I can imagine a world in which evolutionary theory (maybe even including adaptationism) would make a genuine (if not quite *central*) contribution to understanding mental organization. *If* we had sufficient evidence to constrain evolutionary hypotheses about mental traits, *independently* of mechanistic considerations, and *if* we had a principled way of applying those considerations in the context of subsequent causal/functional analysis, we might hope for a kind of pluralism where evolutionary and mechanistic understandings would mutually further and support each other.

I can imagine such a world, but Barrett doesn't prove that we're in it. In fact, his argument schema seems to preclude our passing its borders. Hence, as far as EP goes, number me still among the unconvinced.

Comments

***

Dan Sperber

Dan Burston begins, "I am an anti-adaptationist, at least about the mind" and proceeds to criticize Barrett's book from this anti-adaptationist perspective. What I miss in his text (beside an actual review of Barrett's book, which I believe he distorts is several ways, but I hope Barrett will address this himself) are arguments against adaptationism that are specific to the mind. Objecting to adaptationism in principle and in general (rather than arguing for some specific way of pursuing the program) amounts to rejecting Darwinian evolutionary theory and all the explanations of biological phenomena it has spawned. I would find it hard to take such a position seriously, and maybe so would Burston himself given his qualification of his anti-adaptationism: "at least about the mind."

Of course, the mind/brain has many distinctive features (as does, for that matter, any organ), more specifically, the pursuit of its function crucially involves modifying itself and the behaviour of the whole organism on the basis of environmental inputs. But, in fact, Barrett's book is finely attuned to the specificity of the brain as an 'organ of plasticity." Where, then, are the arguments to the effect that the mind, or the functions of the brain and of the nervous system generally, are somehow exceptions to what we know of biological organs and their evolution, rendering an evolutionary approach at best useless, at worst radically mistaken?

Olivier Morin

My impression is that Burnston puts under the name "adaptationism" the research strategy that shows a favourable bias toward adaptationist hypotheses at the expense of others (e.g. explanations based on drift, mutation, or migration). Adaptationism tries to come up with answers based on adaptation before trying other options. Adaptationism, in that sense, can be a good research heuristic (I think it is), but one can be a good Darwinian and doubt its fruitfulness. How sound the adaptationist gambit is likely to be will vary from field to field, and it could be argued that psychology is one field where it does not work as well as it does elsewhere. One argument is the brain's mutational target size (an idea put forward by Geoffrey Miller). Arguably our most complex and fragile organ, the brain is at the receiving end of a host of random mutations that, on their own, can't do the organism much harm, but whose accumulated effect on a sensitive organ like the brain can be devastating. This might explain why we have diseases of the mind (like schizophrenia or anorexia) that are lethal, frequent, and genetically determined at the same time (which is usually not the case with diseases of the body). I am an adaptationist, about the mind as about other things, but Burnston's position does not strike me as incoherent or anti-Darwinian. Drift too is a Darwinian process.

Dan Sperber

Olivier writes, in defence of Burnston, "My impression is that Burston puts under the name "adaptationism" the research strategy that shows a favourable bias toward adaptationist hypotheses at the expense of others (e.g. explanations based on drift, mutation, or migration)".

Drift? Actually, Barrett does discuss drift. To quote Burnston, "Barrett is willing to countenance drift as helping to produce genetic variation (p. 301) or helping to set up the environmental (including social) circumstances in which adaptation can then occur (p. 220), but drift is never considered as a serious alternative for how mental traits might arise." The last sentence is equivocal: it all depends what is meant by "arise." Genetic variations, that may affected by drift, cause traits to arise. The adaptationist claim in the matter, is that drift (and, more generally, forces other than selection), do not explain how complex traits making a positive contribution to fitness evolve (excluding then your interesting example of deleterious mental illnesses).

If Burnston had such adaptive traits in mind (rather than simple features that are not by themselves adaptive and that might indeed be an effect of drift), it would have been good for him to give an example of the kind of alternative, drift-based explanations that he thinks Barrett shouldn't ignore.

What Burnston does instead is point out that there are alternative accounts in the psychological literature for some of the mental mechanisms that Barrett discusses, alternatives that he, Burnston prefers, and to the development of which, in some cases, he has contributed. Surely, this in itself isn't an objection to Barrett's own account: in psychology, there always are competing accounts. What would have been more useful would have been to show the kind of non-adaptationist story that would explain the evolution of these mechanisms on Burnston preferred account (or why his account makes evolution irrelevant).

More generally, Barrett, as Burnston recognises, is not a rigid, dogmatic, extremist adaptationist fitting the caricature commonly found in anti-EP literature. He is a true scientist, that is, an epistemic opportunist, willing to exploit any source of genuine insight. He believes, as I do, that an Darwinian evolutionary story, where selection plays a uniquely central role, is a major source of insight. Barrett does indeed, to quote Olivier, "shows a favourable bias toward adaptationist hypotheses at the expense of others (e.g. explanations based on drift, mutation, or migration)," and for very good reasons: We gain much more insight in psychology from adaptationist explanations than from "explanations based on drift, mutation, or migration," which doesn't mean we should ignore these other explanations when they are useful. Burnston does not show, or even suggest, that Barrett has been guilty of ignoring such an an explanation. He doesn't give an example where such an explanation would have been better at explaining some psychological mechanism than an adaptationist one. So, let me ask again, in what sensible sense is he an "anti-adaptationist, at least about the mind"?

Daniel Burnston

I'd like to offer a couple of clarifications, since I worry that the main thrust of my argument has been missed.

– "At least." When I say that I am an anti-adaptationist "at least" about the mind, all I mean is that I wish to remain as agnostic as possible about exactly how far my critique extends into projects other than cognitive architecture. Maybe we'll have perfectly good adaptationist explanations of kneecaps and fingernails. I'm only (officially) skeptical about their import for the question of how the brain is organized.

– "Adaptationism." I take adaptationism to be a thesis primarily about explanation. Given some trait of interest, we should be able to give an adaptationist account of that trait. Moreover, adaptationist explanations are central to understanding functions. We don't understand the function of some trait until we have an understanding of what it evolved to do. I take Barrett to be committed to both of these. Since adaptationism is a thesis about explanation, rejecting it is perfectly compatible with believing that there are a lot of adaptations in the world. One just has to doubt either (i) that we will be able to establish adaptationist explanations at the expense of other hypotheses (e.g., concerning drift), or (ii) that adaptationist explanations in fact play a central role in understanding what we want to understand (or both). Despite the fact that I mention (i), my critique is primarily based on doubting (ii). My argument thus doesn't require actually giving alternative drift explanations.

Finally, and I would like to particularly emphasize this, my argument is not based on the fact that I disagree with Barrett's general view of cognitive architecture. I do, but as I said multiple times in the review, I may be wrong. The point is to ask Barrett: given that there are multiple possible views of the functional organization of given cognitive systems, does adaptationism give us any way to adjudicate between them? I don't see how it can on Barrett's account. Since evolution is so flexible, we can't figure out what it will produce until we know what it has produced. So in order to adjudicate the disputes that really matter in understanding cognitive organization, we need to employ the methods of (e.g.) perceptual psychology, psychophysics, systems neuroscience, etc.—fields that rarely make more than passing reference to evolutionary explanations, and to which adaptationism is certainly not central.

In effect I think Barrett's very compelling moves about evolution (all the things in the book that I honestly said I liked)—loosening up on preformationism, incorporating learning and plasticity, etc.—deprive it of the central explanatory place he posits for it, since it leaves the majority of the explanatory work in the hands of non-evolutionary sciences (by which I mean ones to which evolution is not central, not ones that deny evolution).

On a more conciliatory note, and I wish I had talked about this more clearly in the review, I very much like Dan's focus on the pragmatic role of different styles of explanation, and I think this is the most productive way for discussion to go forward. Here is something I am willing to admit: if we have an understanding of both the functional organization of the mind, and how that organization came about, we understand more than if we only know the answer to one of those questions. But Barrett needs a stronger claim, namely that we must know the answer to the second question in order to know the answer to the former.

Here's another way of putting my point. I think Barrett's argument schema has the order of constraint opposite from what he needs. It seems that views of functional organization constrain adaptationist explanations, not the other way around. If we find out we're wrong about organization (by employing the methods mentioned above), we have to change our adaptationist story. If there are constraints coming from the adaptationist side—which, I think, would require a principled way of settling adaptation questions independently of particular views of organization—I would like to know precisely what they are and how they work. If I could be convinced that they existed, I still wouldn't believe they were central, but I might be less of a hard-liner in my view.

Dan Sperber

In reply to Daniel Burnston.

I must be missing something in Daniel's initial post and now in his reply to our comments. The study of psychology started well before Darwin and is still pursued today with, in most cases, very limited use of evolutionary thinking. Short of claiming that all this work has been and is a total failure, who could deny that progress in psychology can and has been achieved without appeal to evolution let alone to "adaptationism"? Regarding evolutionary psychology now, who would deny that its recent development has taken advantage of all the knowledge and competence accumulated in traditional psychology?

In any case, our knowledge of the functional organization of the mind, on the one hand, of its evolution on the other is at best fragmentary, so who would seriously argue that you need to know – know, not just have tentative hypotheses — how the mind is organized before asking how this organization evolved (or the other way around)? If believing this defines an adaptationist, then there are no adaptationists. But there are, and their argument is quite a different one.

Effective research is best pursued by going back and forth between the two perspectives — the study of proximal causation in mechanism at work and the study of ultimate factors in the evolution of these mechanisms — and mutually adjusting these two perspectives. Pre-evolutionary psychology has failed to do so. Because of this, it has had its progress impaired, or even, on some issues, stalled. This is what evolutionary psychologists claim. Neither Barrett nor any evolutionary psychologist need claim that one "must know the answer to the second question [about evolution] in order to know the answer to the former [about organization]". What we do claim is that traditional psychology, not taking advantage of an evolutionary perspective, has failed to develop as well as it could and is stuck with many unsolved puzzles and many weak and implausible hypotheses.

What makes the evolutionary approach uniquely helpful is indeed the Darwinian understanding of the evolution of adaptations. Selection is of course not the only force of evolution, but it is unique in suggesting plausible and sometime testable explanations of the existence and features of adaptations. This is true for psychological adaptations in the mind/brain as it is true for all biological adaptations.

2

"I CAN'T BELIEVE IT'S EVOLUTIONARY PSYCHOLOGY!"

By Olivier Morin

Back in the days when vegetarianism was less fashionable, you could see ads for tofu where a glowing child, her mouth full, exclaimed: "*I can't believe it's tofu!*" One should hope this book will have the same effect on scholars not used to putting Evolutionary Psychology on the dinner table. Clark Barrett's *The Shape of Thought* makes a novel case for adaptationism in the study of the mind, one that addresses almost all of the usual concerns of its opponents. If you think that natural selection could only shape the human mind by building rigid and innate modules—modules that were optimally adapted in the Pleistocene, but can no longer help us act flexibly—, you will discover a discipline that is richer than you thought. If you already know better about the field, this book is much more than an up-to-date introduction to evolutionary psychology. It is a complete rethink of some of its most fundamental notions (chiefly massive modularity, adaptation, and social cognition). In short, I find very few things not to like with the book, except perhaps the picture on the cover (a cross between the monolith in *2001* and an Ikea lamp).

Read this book if you think you already agree with it; read it if you're unconvinced; buy it for your unconvinced friends. They won't believe it's tofu.

While I loved the book, I believe that Barrett's revamped evolutionary psychology reveals problems for that theory. Because it is more nuanced, sophisticated and prudent than some earlier accounts, it lays bare some areas where evolutionary psychology's capacity to come up with novel and surprising predictions can be doubted. This review will focus on two of them: Barrett's views on the proper domain of mental adaptations, and his model of massive modularity.

Adaptations in mind

The first part of *The Shape Of Thought* is devoted to a substantial *aggiornamento* of evolutionary psychology: replace its neo-Darwinian core with an updated evolutionary biology. The field is often seen as wedded to a narrow-minded form of adaptationism. That adaptationism is the bogeyman Barrett wants to exorcise. The adaptationist bogey man gives little thought to the many ways in which development influences evolution. The generation of variation is a black box. So are learning and development. Mental adaptations are innate. Preformationism rules. Barrett dispels this view and sets the record straight in a way no evolutionary psychologist had done. (Though some of these points had been made before—here for instance.) Barrett also corrects the view that adaptation consists in bringing passive organisms to take the shape of an unchanging environment. Friends of Developmental Systems Theory, Evo-Devo, Complex Systems theory, niche construction, or dialectical biology *à la* Lewontin will no doubt find Barrett's brand of adaptationism much more appealing than earlier versions.

The core idea of evolutionary psychology, Barrett-style, is that cognitive structures develop in a way that reflects the adaptive challenges previously faced by the lineage we come from. As a result, today's minds contain information about yesterday's selection pressures. This does not mean we need to represent that information. Our minds work in ways that reflect our evolutionary past, exactly like plants or bacteria, and for the same reasons. Any organism, for Barrett, embodies a series of "inductive bets" that its ancestors placed against nature. A cactus' ancestors had "bet" that their environment would be arid; the cacti that bet differently did not become ancestors. These are metaphorical bets, of course: the product of (more or less) random variation, not conscious decisions. Some of them have been made relatively recently and only by humans, others have been standing for much longer. Barrett claims these bets can be used to make predictions about human psychology, just like they can be used to study cactus physiology. When doing so, we need to keep in mind that our ancestors did not bequeath us ready-made cognitive structures, but flexible developmental programs. The targets of adaptation by natural selection are processes, not structures, and these processes can handle environmental novelties in a highly open-ended fashion. Adaptations need not be innate (and even "innate" adaptations must develop). Cognition develops in a way that uses the information left in the genome (and elsewhere) from previous inductive bets, in a way that permits us to behave flexibly. Evolutionary psychologists study dynamic systems that spawn thoughts, concepts and modules as the need emerges, that is to say, constantly from birth to death. Cognitive fossils is not what Barrett is after.

How do we use evolutionary theory to make predictions on how the mind works? This is a downside of Barrett's *aggiornamento*: he needs to make the answer to that question a great deal more complicated than it used to be. In standard evolutionary psychology, the way to go would be to identify a selection pressure (avoid predators, for instance); to figure out an adaptive way to meet it (spot animated things and select an immediate flight-or-fight response); to identify what instantiated the selection pressure in the past (snakes and spiders, not cars or drones). That set of adaptively relevant events constitutes the adaptation's environment of evolutionary adaptiveness (EEA), which in Barrett's account is not really an environment at all (even less so a time and a place; certainly not the Pleistocene savannah). Barrett insists, quite sensibly (like his predecessors did before him), that the EEA is something more abstract, "a long smear of events" (p. 168). Environments, in fact, do not exist independently of the animals that inhabit them. They are, in part, shaped by the animals' own behaviours and cognition. In Barrett's framework, the organisms themselves shape the adaptive challenges they have to face. This makes EEAs even more elusive than they usually were.

What are mental adaptations adapted to?

The inductive bets that our ancestors made against nature were not, as we saw, real bets in any sense. They lacked intentionality. They became bets a posteriori, after selection had removed the bets that were less fit. So it could seem weird to ask what these bets were "about," as though they were anticipations of events, like those you might find inside a trader's head. A purist would say that the EEA is just a dead list of causal factors with no bearing on the future. Yet the task of evolutionary psychology requires that we find a way beyond this limit—some method that can allow us to predict whether a new thing or event will resemble the EEA or not. This means finding coherent shapes in the "long smear of event" that was human evolution. This proves to be a tricky task.

The problem of defining an EEA brings us close to a notion developed by Dan Sperber and developed at length in Barrett's book: the proper and actual domains of mental adaptations (p. 27 and passim). Evolutionary psychologists know of two ways to define a cognitive mechanism: by the kind of input that it is capable of processing (its "actual domain"), and by the kind of input that the mechanism got through the filter of natural selection by treating: its "proper domain." The actual domain is easy to investigate in the lab, by toying with cognitive mechanisms and seeing what makes them tick. Not so the proper domain. It belongs in the EEA: it is the set of things and events that the organism reacted adaptively to, thanks to that particular cognitive device. The proper domain is where we can find the "inductive bet" that the particular device we're studying has placed against nature, and has been winning (so far).

Take face perception. According to some, our brains contain specialised areas that become active when we are exposed to human faces, but also to photographs, masks, make-up, cartoon faces, etc. Suppose this is indeed the area's actual domain. What is its proper domain? To answer this, we would need to determine when approximately the cognitive device stopped evolving (before or after the appearance of face paint? of masks?); what adaptive challenges it faced (was it only dealing with humans, or do we include dogs as well? or even other animals? if so, which ones?); what this could tell us about its architecture (does it include a specialised eye-detecting device? Did we need face recognition to be included in face detection, or should the two be separate?). To make a genuinely new contribution to psychology, adaptationist theorists would need to settle these issues with evolutionary theories and data: no looking over our shoulder at lab results (at least at first).

How is this done? Barrett adamantly refuses to provide his readers with any kind of general guidance. His answer, one of the book's leitmotiv, is "It depends". "It depends" is "the First Law of adaptationism", the book's first and last word on how to come up with adaptationist hypotheses. Some readers will no doubt find the First Law liberating; but it will feed others' suspicions.

One such suspicion is that proper domains are simply inferred from the actual domain; that we learn what make mental mechanisms tick by reading the experimental literature, then project this knowledge into the past. Nothing really wrong with this, but we were told that evolutionary psychology would change psychology, and provide it with new, testable hypotheses. This it cannot do if it is merely projecting experimental findings backward in time. Although I disagree with Daniel Burnston's bleak view of the field (see his review here), I share his impression that Barrett often comes close to biting the bullet of unfalsifiability. Face perception is a case in point. There are brain areas that seem to respond selectively to faces. How selectively? This is debated. You might think that evolutionary psychology could help orient the debate: after all, perceiving faces is an evolutionarily relevant task, and mental adaptations should have evolved around it. Surely, we can determine their proper domain, and explain it to other researchers? Well, no; at least not according to Barrett. We must wait for the neuroscientists' answer (p. 118–119). Perhaps they will conclude that there is a specific face perception area; perhaps they will conclude that it is, in fact, much more general. The area's proper domain will be whatever these specialists (who study the mechanism's *actual* domain) decide. "It depends."

One part of the book that somewhat gives this game away is its treatment of culture (chap. 8–10). Barrett claims that a wide range of psychological mechanism imitation (p. 45) to communication (p. 240) to Theory of Mind (p. 211)—are adaptations to culture. They evolved in an EEA that contained culture, and the adaptive problems that they solved had to do with acquiring it; but what is culture? Well, "it depends." In chapters 8 and 9 it consists of socially transmitted information that is "highly variable and constantly changing" to keep pace with an environment that is itself variable (p. 211). In chapter 10 it is made of long-standing traditions that allow good ideas to accumulate. The prototypical cultural obejct is language in chapter 8; in chapter 9 it is fads and fashions (lead by the rich and prestigious, whose deplorable habits we naively copy, p. 230); in chapter 10 it is technology. In fact, forget about those examples. The real answer is much simpler than that:

"Culture acquisition mechanisms evolve because their products—culturally transmitted behaviors—increase fitness. Culture, in this case, is just a word for the knowledge and behaviors these mechanisms acquire and transmit (…)." (p. 226)

Culture, here, is whatever culture-acquisition mechanisms take as input today. There goes the distinction between proper and actual domains.

I do not subscribe to the view that adaptationist hypotheses are all untestable, post hoc or circular speculations (the evolutionary study of human family ties, for instance, shows how fruitful adaptationism can be—Chapais, 2010; Hrdy, 2011). Yet it is true that Barrett does not seem to be aiming at a bold, novel and testable theory of how the mind works. What he builds instead is an elegant, coherent framework that can accommodate work from a wide variety of fields and perspectives (in itself not a small achievement). This becomes evident when the book tackles the "massive modularity" debate, in its much expected last chapters.

In and out of the igloo—and back in again?

As this blog's readers no doubt know, the debate started from Fodor's proposal that some parts of the mind—the modules—were domain-specific processors, rapidly and automatically treating a narrow range of inputs. Some parts only: Fodor takes the mind's core to work in a completely different fashion, integrating a limitless array of information in a way that defied mechanistic understanding. This Barrett calls "the 'igloo' model of the mind: a crunchy outside composed of rigid innate modules and a soft center composed of general-purpose cognition" (p. 265; Barrett's knack for metaphors is a delight for readers). Dual-processes models of the mind (like Keith Stanovich's) are also pinned down as variants of the "igloo model." Apparently any theory where a central processing unit flexibly organises a broad array of information sent by a multiplicity of more specialised units, is a theory that belongs in the igloo (p. 266). Barrett makes his ambition clear: Let's tear down the igloo! Does he succeed?

Massive modularity started out as the contention that Fodor's modular model might apply to the human mind as a whole, not just to some of its parts; but many massive-modularists in fact reject the whole Fodorian deal, including its description of modules. Barrett is one of them. He takes none of the Fodorian premises on board. The distinction between domain-general and domain-specific mechanisms is rejected (every mental activity is necessarily domain-specific, writes Barrett, p. 27—in another passage that avoids circularity by a hair's breadth). So is encapsulation (p. 269–270). Mental adaptations are not particularly fast or automatic (p. 272, 276). As for the impression that mental adaptations must be innate, or that evolutionary constraints must restrict cognitive flexibility, the book does an admirable job of dispelling it.

In fact, saying that modules are not innate would massively understate Barrett's point. The book draws on the lessons of "modularisation" theorists (like Annette Karmiloff-Smith) to argue that our mind constantly spawns new mental adaptations, pretty much as needed (p. 304, p. 318, p. 329): modules for reading, modules to ride a bicycle or to play a videogame. Each of these modules can freely interact with almost any other. Each can "choose" (Barrett's word) what input to process, flexibly, in a way that is sensitive to a broad range of information. How so? Well, the assembly of modules is synchronised, through a "central pool of information, like a bulletin board" (p. 93) where information is shared and tagged. The tags allow other modules to know which information is worth processing at any given time. This central bulletin board is also the place where attention emerges. Attentional processes direct the activity of modules, sending them input to process in priority, and thus helping the organism react flexibly to novel events. This central bulletin board (it is clear from these chapters that there is one big central pool where most of the information is gathered) is implemented by long-distance neural hubs connecting a wide variety of areas, as suggested by Bars' "global workspace" theory.

Here (and in no other place in the book), I couldn't believe it was tofu because I suspected that, in fact, it wasn't. I need to squint hard to tell Barrett's "bulletin board" from the "System-2" of dual-process theories. Here is a central hub "where information from many parallel processes comes together and is integrated" (p. 295); here is where attention originates and sends information to modules, controlling and constraining them to process it (p. 271); here is where consciousness and the sense of will emerge. It is hard to see how the many devices feeding the controlling, information-integrating central hub could avoid being, in comparison, less central, less conscious, less in control, and condemned to process a narrower range of input. Barrett sees his model as the opposite of the igloo model— the oposite of "a two-layer model of mind, with a general-purpose homunculus sitting on top" (p. 286); but where exactly is the difference? As an answer, Barrett accuses his opponents (the denizens of the igloo) of anthropomorphising the central processing unit, and treating the brain as "one undifferentiated blob" (p. 287). Is he being quite fair, though?

Tellingly, the book occasionally slips back into old-school modulespeak. We are warned that mental adaptations should not be considered particularly fast, unconscious, or domain-specific (cognition is domain-specific through and through). Yet, in a discussion of pop-out effects in perception (p. 113), we are told that such fast, unconscious reactions are "a good sign that a specialised detector is at work." That is a small and harmless inconsistency, but one wonders whether we ever came out of the igloo.

Another worry that Barrett's account does not dissipate has to do with the "spawning" of modules (what Dan Sperber calls "teeming modularity"). Barrett's view relaxes the classic definition of modularity to the point that anything, it seems, could become a module. Modules can be formed in response to entirely novel situations, like learning to bike, to write, or to play a video game. Where, then, does modularisation stops? One answer is that it never really does. (I am tempted to attribute this answer to Sperber.) At a limit, one might say that your mind spawns a module when you learn to play *Candy Crush*; another when you learn somebody's name; and so on. Indeed, if you buy into teeming modularity, it is hard to resist the view that there is a module for *this sentence that you are reading now*. There would be, in this view, a module for every thought. Barrett, it seems, won't bite that particular bullet. He might be right: after all, teeming modularity raises difficulties of its own. (For instance, what is the proper domain of the thought that "this software needs updating", as opposed to its actual domain?) The deeper problem is that the book provides no general way of identifying mental adaptations or telling them apart. We are constantly brought back to the "First Law of adaptationism": "It depends."

***

These reservations concern not so much Barrett's excellent book as what it reveals about the current state of evolutionary psychology as a theory. *The Shape of Thought* is the best introduction to it that I know of; certainly the most nuanced and sophisticated; also the only one that might appeal to some of the discipline's usual opponents. Barrett knows to avoid the sore spots, and sidesteps all the traps. This achievement comes at a price: his outlook is prudent to a fault, and he suggests few untrodden research paths where refutable predictions could be made and tested. There are novel insights in the book (the analysis of social ontology as a set of short-cuts for causal reasoning, in chapter 4, is superb), but overall, it reads like a conclusion to evolutionary psychology, more than an introduction.

Full disclosure statement: Olivier Morin's new book, *How Traditions Live and Die*, is also published by Oxford University Press.

3

ANGLES OF APPROACH

By Clark Barrett

First, let me thank Daniel Burnston and Olivier Morin for taking the time to read and review my book. I appreciate both the kind remarks they have made about the book, as well as the challenges they have raised. Debate is exactly what one hopes for when one writes a book such as this. Many of these challenges are well-taken, and I will attempt to address them below. There were also, I believe, some misunderstandings. Both Burnston and Morin seemed to expect, in different ways, more than the book offered; I suspect this is due to different views about how the sciences of the mind should, and in fact do, proceed. We all seem to agree on the object we're trying to approach, just not on how to approach it.

Let me begin with a brief list of things that the book was *not* intended to be. It was not intended to defend a specific theory of mental architecture, nor to provide a complete account of the mind, how it evolves, or how it works. It was not, unlike many books jostling for attention in the market of academic brands, intended to postulate a single "secret sauce" via which some swath of mental phenomena would be deftly explained. In fact, the book was not even intended to advance a central thesis—with the exception, perhaps, of what I called the First Law of Adaptationism: "it depends." As I noted in the book, this law doesn't want to be part of any club that would have it as a member. But, it sums up fairly concisely my view of how evolution works: evolution, like history, is essentially one damned thing after another, with *some* principles, which sometimes explain, in some cases, why B but not C follows from A. Those principles, of course, are highly contingent on the details of the case and disappointingly unlike the neat, grand laws of physics, mathematics, or logic. Hence, Morin's "downside" of the book: it's complicated.

Given this, I realized when I wrote the book that it might come up against some Gricean confusion, because most science books are expected to advance and defend some central thesis—usually something bold and counterintuitive. Readers therefore justifiably look for such a thesis, and try to take it apart. I've had some evidence of Gricean mismatch in responses to talks I've given on the book, which go something like: "Thanks for an interesting talk, but—what is there to disagree with?" To which my reply is, "Absolutely nothing—thanks!" An oddly disappointing exchange, for both parties.

So if the above is a list of things *The Shape of Thought* wasn't, then what is it? Let me quote from the Introduction:

In this book I'd like to build a case for what I think a properly "holistic" evolutionary psychology should look like: an evolutionary psychology that brings all mental phenomena, from brain development to culture to consciousness, under the rubric of evolutionary explanation – at least potentially. But let me emphasize that my intention is not to give a complete account of how the mind works. Nor will I claim that all of the phenomena I'm talking about are completely, or even mostly, understood. Far from it. While evolutionary psychology has made substantial contributions to understanding the mind over the past two decades, it is still in its infancy. There is no question that as we discover more about the brain, aided by accelerated technological advances in areas like genetics and brain mapping, the theories and methods of evolutionary psychology will have to evolve. However, I think we are already in a position to see what kind of framework we'll need for thinking about how minds evolve to fit the world. That is the kind of framework I aim to depict here. (p. 11)

The idea of a "framework," here, is important. A framework is not a thesis. It's not even a theory (depending on what your view of a theory is). It's a set of conceptual tools, models of how fragments of the world's causal fabric work that can be used to theorize, explore, and hopefully in the end help explain things. Morin, in his review, recognizes this: "Barrett does not seem to be aiming at a bold, novel and testable theory of how the mind works. What he builds instead is an elegant, coherent framework that can accommodate work from a wide variety of fields and perspectives."

Note that nowhere in the above description of the book's goals does the word "prediction" appear (in the book the word "predict" and its variants appear 82 times, but almost all of these are about the predictiveness of cues, strategies, or inductive bets that organisms might use). This is not because I think that the framework I'm offering *can't* predict things. I certainly hope, and believe, that it can (more on this below). But I think the idea of "prediction" has been oversold in the natural sciences. It's something one can hope for, and it's nice work if you can get it. But in my view, it's not the only way science proceeds, nor even perhaps the major way. And at its worst, the prediction game as played by many evolutionary and non-evolutionary social scientists is a recipe for collective hallucination and false positives. That might be off-putting to some of my colleagues, but, unfortunately, wishing it were otherwise won't make it so.

I emphasize this (perhaps unorthodox?) view because a central aim of my book was to step back, take a look at the theoretical landscape, and see if we can clear some smoke away and separate what is solid—what aspects of evolutionary, cognitive, and developmental theory we can rely on as biologically sensible—from the historical detritus of our discipline that people only believe in because it was taught to them in graduate school. I confess that I am particularly concerned with psychology as a field that is largely unmoored from sound biology, at least in some sub-disciplines of psychology such as social psychology (not that theories in psychology aren't often "biology-ish;" but that can be precisely the problem). However, the problem of intellectual parochialism is not unique to psychology. There are biological anthropologists, for example, who study the evolution of behavior but don't bother with the mountains of data about the organization of the human mind produced by psychologists and neuroscientists.

A goal of the book then, was—as Joe Henrich said so kindly in his blurb on the back cover—"setting the house back in order." Meaning: opening the windows, letting some air in, throwing away the immediately obvious junk, looking at what's left, and asking what's worth keeping, polishing, salvaging, or repairing, and what's worth replacing with something newer, better, more efficient, or more likely to be true. A major problem, here, is that the relationship between truth and beauty, in biology at least, might not be quite as clear-cut as it is in physics. Social scientists might have some beautiful theories: simple, symmetrical, logical, tidy. But, as I argue in the book, if the choice is between simplicity and reality, then reality's got to win, and our approach strategy has to aim for it, grace be damned. Take, for example, Fodorian modularity: beautiful, perhaps, logically neat, but as likely to be true as the mind is to resemble a cube of quartz. More than one person has asked me whether it wasn't so much easier and cleaner when we defined modularity in terms of cognitive encapsulation, and my reply is: "yes, yes it was."

To be clear, the evolutionary psychology that I present in the book *can* make predictions, because this evolutionary psychology seeks to make use of the full set of theoretical advances and tools from all fields that are relevant to understanding the evolution of the mind—provided that their logic is biologically sound. Typically, in my view, the best predictions will emerge from careful use of formal models, designed to capture relevant empirical facts about what they are attempting to model. You want predictions about human mate choice? Construct a model based on reasonable assumptions about human mating structure, sexual selection, and the decision-making problems individuals face in seeking a mate, and see what the model predicts. You want predictions about the developmental trajectory of language acquisition? Formalize the adaptive problems and tradeoffs inherent in the aspect of language acquisition that interests you, and model the shapes of the reaction norms you'd expect to see under various assumptions. The book attempts to sketch the kinds of ingredients that might go into such models, and provides a way to think about setting them up. It is quite obviously, however, not a modeling textbook; some assembly required.

Burnston's commentary

Let me turn now to some of the specific critiques made by Burnston and Morin. Burnston begins with the striking conversation-opener "I am an anti-adaptationist, at least about the mind." I'm assuming Burnston doesn't mean that he denies a priori the possibility of adaptations in the mind (if so, this might be a short conversation). Instead, his skepticism seems to be grounded in a version of the standard epistemological argument raised against evolutionary psychology, beginning with Gould and Lewontin, which goes something like: sure, there might be adaptations, but we can never really know, so we shouldn't bother.

Before moving to that, let me address Burnston's claim that I focus excessively on adaptation, at the expense of other processes that shape mental traits, such as drift. If I gave the impression that I don't think drift, historical contingency, developmental constraint, homology, and other factors aren't important, that's unfortunate, especially since they appear throughout the book beginning in Chapter 1. But more importantly, there is a reason why the book is subtitled *How Mental Adaptations Evolve*, not *How Mental Traits Evolve*. Why? Because the book is about mental adaptations. And as Burnston himself points out, not all traits are adaptations—indeed, in all likelihood, most aren't. While he's right that there is a debate about the degree to which processes such as drift account for the evolution of traits, there is not a debate about which process accounts for the evolution of adaptations.

Still, I think Burnston overlooks the role that vicissitude and chance events play in the picture of mental evolution presented in *The Shape of Thought*. Here's a passage from the opening of Chapter 10:

Look around you: would the details of what you see be guaranteed to occur in any possible universe? Would it contain Nintendo, breakdancing, neckties, The Real Housewives of Orange County, jeans shorts, handshakes, clowns, and moustache wax? If you rewound the tape of history and let it play again, is this exactly what you'd see every time? And even with the benefit of hindsight, would any theory you can imagine actually be able to predict these things?

In evolution, as for any historical process, prediction is a tall order. You'd be as hard-pressed to predict the things above as you would be to predict the current position of the continents based on the starting state of the earth, or to sit in your armchair in 1,000 A.D. and predict World War I. But that isn't to say that these things aren't caused, nor that there aren't principles behind them. (p. 243)

This passage is about cultural rather than genetic evolution, but of course the same points hold for both. The point I'm making, in fact, echoes one made by the master anti-adaptationist himself, Stephen Jay Gould, in his book *Wonderful Life*: much of evolution involves historical accident (the "rewinding the tape" metaphor comes from him; note, too, the remark about prediction). The question for the study of adaptations is not whether or not chance events occur; it is how natural selection operates within a world of chaotic, willy-nilly occurrences.

Burnston's version of the epistemological argument made by Gould and Lewontin is as follows, from his commentary:

Consider a mechanism for a psychological trait X, and assume that all are agreed on its current functional role—what it computes and how, and how those computations contribute to mental functioning in general. Suppose, however, that there are differing opinions on the evolutionary story. Perhaps theorists A and B agree that it's an adaptation, but disagree on what adaptive problems it was designed to solve. Theorist C, on the other hand, is convinced that X is entirely due to drift. The point is that, by the argument schema, we've already fulfilled the mechanistic understanding, and therefore the outcome of the evolutionary debate makes no difference to our understanding of the mechanism's role in cognitive organization. But, per claim 4, the evolutionary story is supposed to be central to this very understanding—Barrett states repeatedly that we must understand cognitive organization in evolutionary terms. The argument schema fails, on its own terms, to justify the centrality claim.

To paraphrase, the argument seems to be something like the following:

(1) Assume a case where evolutionary thinking makes no difference to our understanding.

(2) In that case, evolutionary thinking makes no difference to our understanding.

(3) So, evolutionary thinking can't play a central role in our understanding of the mind.

It seems to me that (3) doesn't follow from (1) and (2). Logic aside, though, the question arises of how often Burnston's scenario might actually hold. First, I'd like to see the stats on how often psychologists all agree on the "current functional role" of a psychological trait. Provided we could find some unicorns of this kind, we'd then want to know how often it was the case that all evolutionary explanations for the trait in question were equally plausible or implausible—a perfectly flat probability function, so that we could say strictly nothing about the evolution of the trait. As I stress in the book, an evolutionary approach (or any scientific approach, really) is not about homing in on the single true explanation with 100% certainty; it's about weighing between alternatives based on a combination of theory and data. Degrees of confidence count. Sometimes even very basic evolutionary theorizing, coupled with existing evidence, can make some hypotheses quite plausible and others the opposite. To illustrate with two examples from the book:

There is fairly clear evidence for specialized face processing in the human brain, but there is a debate about whether this is due to a specialization *for* face processing, or a byproduct of something else (basically, experience alone, without a history of selection to be good at learning faces). Homology of the mechanism across primates, and the likely massive fitness benefits of recognizing individuals in primate social groups, renders the evolved specialization account quite plausible. In fact, it seems unlikely that there hasn't been selection for face recognition at least since the origin of the primate lineage some 60 million years ago: experimental data show that lemurs, and not just haplorhine primates, can recognize individual faces, adding to the phylogenetic parsimony of the hypothesis (Marechal et al., 2010). Both possibilities, of course, remain on the table, but evolutionary considerations are quite relevant to weighing their relative plausibility.

Contrast this with another example. There is fairly clear evidence for specialized word processing in the brains of literate humans. However, the hypothesis that this reflects an evolved specialization *for* reading has a low a priori probability because of the recent historical origins of writing systems. Thus, while the proximate evidence for face and word specializations in the brain are roughly the same (brain mapping, lesion studies, etc.), the set of plausible evolutionary explanations for each are not the same at all. The point is that, contrary to an "anything goes" caricature of adaptationism, not all evolutionary hypotheses are equally plausible, and ironclad certainty is not necessary for evolutionary thinking to play an important role in our theorizing.

Regarding the "central role" of evolutionary thinking, then, my claim is not that evolutionary thinking can always provide insight in psychology. Often it can't. The claim, instead, is that because the mind is most definitely the product of evolutionary processes, then the ultimate explanation for it and its components must be evolutionary. True, it might be hard to know what the correct explanation is, in many cases. But I find it hard to understand the argument that we shouldn't try, and that we should instead content ourselves with a completely proximate, mechanistic description of a system that we *know* is the product of evolution. That resembles a kind of "don't ask, don't tell" policy, which generally don't turn out well.

One last point on Burnston's commentary: he takes issue with the idea of hierarchical organization presented in the book. I think we are talking past each other to some degree here, because I largely agree with the details of what he's arguing. My point is that most of the brain and its processes are structured hierarchically in the sense that smaller-scale, local structures and processes are nested within larger ones. Poldrack and Yarkoni (2016) describe this as the idea that "lower-level units are repeatedly configured into higher-level circuits" (p. 20.11). Bullmore and Sporns (2009) and Meunier et al. (2010) provide useful reviews of the hierarchically modular organization of the brain's connectivity structure. There is lots to debate here, but since my aim is not primarily to defend a highly specific model of mental architecture, I'll put this aside for now, and will return to the question of "igloo" modularity below.

Morin's commentary

Morin's review of the book was much more positive than Burnston's (so naturally, I liked it more). His introduction captured very well the spirit of the book. I particularly appreciated his remark that the book "is much more than an up-to-date introduction to evolutionary psychology. It is a complete rethink of some of its most fundamental notions." That is certainly what the book was attempting: a kind of frame expansion that allows us to think about mental adaptations using a much broader set of conceptual tools, allowing it to make contact with areas of biology and cognitive science such as evo-devo, culture-gene coevolution, niche construction, epigenetics, dynamical systems, and embodied cognition.

The question Morin asks, if I read him right, is: can this project succeed without turning the concept of adaptation to mush? It's certainly a question worth asking, and I can tell you that many people have voiced similar concerns to me (some are quite alarmed at the prospect of evolutionary psychology even vaguely flirting with ideas such as niche construction or cultural group selection—hide your children!). There are charges of near-circularity, unfalsifiability, and the lament that making things more complicated is a "downside." An initial reply to that might be: Bummer. Stuff is complicated. But I realize that's not very helpful, so let me try to clarify why I think the wide-angle evolutionary psychology I've sketched in the book improves, not lessens, our chances of making real scientific progress in understanding the evolution of the mind.

Here's a metaphor. Let's say you want to catch a fish, or even some unknown number of fish. You don't know how many there are, how big they are, what shape, or even *where* they are. What do you do? You cast a wide net. From there, you draw in the net, see what you've caught, and adjust the size and shape of the mesh as needed. Just because you start with a big net doesn't mean you can't later adjust it or make more, special-purpose nets; but if you start in the opposite direction, a bunch of fish will get away, and you'll never know it.

Starting broad and narrowing in is my approach strategy for several concepts in the book, including modules, culture, and mindreading. When building the basics of a discipline that's meant to capture all mental phenomena, I think it's best to start with concepts defined generally enough to capture *both* the currently existing phenomena that scholars are talking about using that term, and the ones they might not have noticed come under the same conceptual rubric

This is why, for example, I start with the broadest possible definition of mindreading. Mindreading, the book proposes, should be defined as anything that "uses a cue or cues that index another's mental state" (p. 129). This is a far broader sense than most psychologists would be comfortable with, especially since it does not require the *representation* of mental states. But I think it's the best biological place to start. Compare Krebs and Dawkins, from their 1984 paper "Animal Signals: Mind-Reading and Manipulation:"

Animals will come to be sensitive, then, to the fine clues by which other animals' behavior may be predicted. The clues that a mind-reader may employ are varied and numerous, and are much discussed in the ethological literature..." (Krebs & Dawkins, 1984, p. 387)

Some might argue that the definition is so vague as to be useless, and that Krebs and Dawkins provide the reader with no general guidance for how to go about generating and testing hypotheses about these "clues." Other readers might say, and have, that this definition doesn't rule out learning to associate observable cues (e.g. gaze) with behaviors (e.g. attack) without any representation of mental states. Nor does it rule out unlearned behavioral reflexes that couple a cue with a behavior. Indeed. (In the book I use the term "index," very carefully, in the Maynard Smith and Harper sense). My reply is that the broadest possible definition sets boundary conditions on a family of biological phenomena, within which we can later make finer-grained distinctions and functional taxonomies. For example, we might want to distinguish representational from non-representational mindreading, with all the additional problems that operationalizing such a distinction across species might entail. But a virtue of starting broad is that it allows for the comparative method to be applied to the full scope of the varieties of mindreading (and many biologists feel that the comparative method is the *only* way—design logic be damned—to settle questions about adaptation). By looking at varieties of mindreading across species, how they are distributed, and how they correlate with ecological and social problems in different taxa, we can find out much more than simply defining "theory of mind" as something only humans have, as many scholars do.

Morin seems to like my treatment of mindreading and social cognition in the book, so he probably doesn't take issue with the wide-angle strategy here. But he does point to other varieties of this strategy as reasons why the book's ideas verge on circularity and unfalsifiability. For example, he takes issue with my treatment of culture, and of proper and actual domains. He doesn't like the idea that we might make inferences about proper and actual domains from observations. Nor does he like that I define culture as what cultural transmission mechanisms transmit, stating "Culture, here, is whatever culture-acquisition mechanisms take as input today. There goes the distinction between proper and actual domains."

Let's set aside "culture" for a moment and consider a term like "social cognition." Suppose I were to refer to the inference some commentators have made, based on Donald Trump's remarks about 'two Corinthians,' that he doesn't read the Bible very often, as an instance of "social cognition." Despite the fact that this is an instance of social cognition "today," it doesn't imply that the term social cognition is empty, nor does it mean we can't think carefully about proper and actual domains of mechanisms of social cognition in the Trump / Bible example. Trump himself is clearly part of the actual, not proper, domain of social cognition; ditto the Bible. In fact, I can hardly think of a topic that is better elaborated in the book than the distinction between proper and actual domains. It's the reason for all the discussion of open reaction norms, types, tokens, and so on. On my account, Trump is a token (!) of the evolved conceptual category PERSON, which is a conceptual type that is part of the proper domain of social cognition. And there are additional proper / actual, type / token analyses one could do on that example. The point is that broadly defining a domain like "social cognition" or "culture" doesn't immediately throw out the logic of the proper / actual distinction within that domain.

Culture, as I presume Morin agrees, contains *many* proper and actual domains; that's the whole idea behind cultural attractors, which I think is a very useful idea (along with the idea of cultural evolution, which is not... am I going to start an argument here? incompatible). The book discusses multiple examples of domains within the overarching domain of culture (language, tools, moral norms). Saying that cultural transmission mechanisms were selected to transmit cultural information is no more circular than saying that the function of language acquisition mechanisms is to acquire language, that the function of mate choice mechanisms is to choose mates, that the function of predator avoidance mechanisms is to avoid predators, or that the function of perceptual mechanisms is to perceive. These simply refer to phenomena but are not explanations of them.

Let me return, before closing, to two issues related to differences in angle of approach: the roles of a priori prediction and post-hoc explanation, and the concept of modularity. Then I'll close.

Looking forwards, looking backwards, and peeking

Morin's discussion of prediction and falsifiability uses the example of face perception, which I've broached above. I'll quote extensively to capture his argument:

Take face perception. According to some, our brains contain specialised areas that become active when we are exposed to human faces, but also to photographs, masks, make-up, cartoon faces, etc. Suppose this is indeed the area's actual domain. What is its proper domain? To answer this, we would need to determine when approximately the cognitive device stopped evolving (before or after the appearance of face paint? of masks?); what adaptive challenges it faced (was it only dealing with humans, or do we include dogs as well? or even other animals? if so, which ones?); what this could tell us about its architecture (does it include a specialised eye-detecting device? Did we need face recognition to be included in face detection, or should the two be separate?). To make a genuinely new contribution to psychology, adaptationist theorists would need to settle these issues with evolutionary theories and data: no looking over our shoulder at lab results (at least at first).

How is this done? Barrett adamantly refuses to provide his readers with any kind of general guidance. His answer, one of the book's leitmotiv, is "It depends". "It depends" is "the First Law of adaptationism", the book's first and last word on how to come up with adaptationist hypotheses. Some readers will no doubt find the First Law liberating; but it will feed others' suspicions.

One such suspicion is that proper domains are simply inferred from the actual domain; that we learn what make mental mechanisms tick by reading the experimental literature, then project this knowledge into the past. Nothing really wrong with this, but we were told that evolutionary psychology would change psychology, and provide it with new, testable hypotheses. This it cannot do if it is merely projecting experimental findings backward in time. Although I disagree with Daniel Burnston's bleak view of the field, I share his impression that Barrett often comes close to biting the bullet of unfalsifiability. Face perception is a case in point. There are brain areas that seem to respond selectively to faces. How selectively? This is debated. You might think that evolutionary psychology could help orient the debate: after all, perceiving faces is an evolutionarily relevant task, and mental adaptations should have evolved around it. Surely, we can determine their proper domain, and explain it to other researchers? Well, no; at least not according to Barrett. We must wait for the neuroscientists' answer (p. 118–119). Perhaps they will conclude that there is a specific face perception area; perhaps they will conclude that it is, in fact, much more general. The area's proper domain will be whatever these specialists (who study the mechanism's actual domain) decide. "It depends."

Never mind the "stopped evolving" bit, or other aspects of this passage I'd take issue with. Here I think we might have a difference of opinion about how science can, and does, proceed. Consider Morin's statement, "To make a genuinely new contribution to psychology, adaptationist theorists would need to settle these issues with evolutionary theories and data: no looking over our shoulder at lab results (at least at first)." The claim here resembles the "no peeking" rule for data analysis. "Data peeking," as it's sometimes called in the literature, can be bad for a couple of reasons. First, it's bad to "make a prediction," peek at the data, modify one's prediction, and then claim it was the original prediction. Second, it's bad to make a prediction, get some data, look at the data, see if your prediction is confirmed (p < .05??), and if not, keep collecting data or adjusting your protocol until it is. Bad. We all agree.

Morin suggests that the alternative is “reading the experimental literature, and project[ing] this knowledge into the past. Nothing really wrong with this, but we were told that evolutionary psychology would change psychology, and provide it with new, testable hypotheses.” My suspicion, here, is that psychology *has* changed because of evolutionary psychology, including people taking the good parts on board without recognizing them as such—but perhaps nobody’s noticed. That aside, I’m not sure I agree that there are just two starkly delineable alternatives: (1) providing new, testable hypotheses and (2) projecting knowledge into the past. Indeed, I think that’s an anemic view of how real progress in understanding the evolution of the mind—which needs evolutionary theorizing, not just brute-force empirics—occurs.

Consider how we’re learning about the genetic differences between humans and chimpanzees. Important questions in this area are: which changes in the genome are due to selection and which are not? And, what evolutionary theories might explain this? One such theory is Kaplan et al.’s “embodied capital” theory of life history and brain evolution that I describe in the book (Kaplan et al., 2000). Under the “no peeking” rule, this theory is supposed to make predictions independent of the data, which it does. For example, it predicts that brain size, extended juvenile periods, extended brain plasticity, and reliance on hunting will co-evolve, and there is paleoanthropological and archaeological evidence for those parts of the predictions that fossilize (with the caveat that such evidence is always tentative, and new technologies give us better and better data that can adjust the picture; new data on tooth formation, for example, suggests long life histories might have evolved later than originally thought).

But there are a variety of ways the embodied capital theory could play out at the level of genes and gene regulation. Loosely, we’d predict selection in our genus to act differentially on genes influencing brain growth and genes prolonging early development, and that genes involved in brain plasticity would continue to be expressed later in life in humans than in other apes. It turns out that all of those are true, but in order to find out the details we have to peek. For example, in a recent review of 36 genes that show evidence of being uniquely altered under positive selection in the human lineage, O’Bleness et al. (2012) identify 13 that are involved in increasing brain size, higher brain function, or some other change consistent with the Kaplan et al. hypothesis (I haven’t included genes influencing reproduction and developmental rate, but there are some of those too; see also Somel et al. (2013) on extended expression of brain plasticity genes). For most of these genes listed as influencing the brain, however, the exact effect on the phenotype is unknown: 9 are listed with a “plausible” effect, 3 with a “likely” effect, and 1 with a “definite” effect. Consistent with the evolutionary theory? Yes. Predicted? Yes. *Exactly* predicted? Not really—not the specific genes, anyway.

These genetic data were obtained via brute force empiricism, and thus, in the stark contrast made by Morin and Burnston between looking forwards and looking backwards, evolutionary theorizing offers nothing in explaining these genetic changes because they were obtained independently of someone peering into their crystal ball and predicting them. Indeed, nobody predicted these *specific* genetic changes because nobody tried to find out the functions of these genes (for the most part) until it was realized that they were uniquely derived in humans. And yet, I'd argue, one would be foolish to ignore evolutionary theory in attempting to understand these changes. More than that, I'd argue that we'll be unable to explain them without theoretical tools of the kind presented in the book, or something like them. In the case of brain genes and other uniquely derived changes in human brains that have been selected for because of their psychological benefits—which are, therefore, adaptations—I think we will. But there are lots of ways to catch a fish, lots of approach trajectories, all of which we should consider.

Modularity

In his chapter in the seminal volume *Mapping the Mind*, Dan Sperber remarked: "If modularity is a genuine natural property, then what it consists of is a matter of discovery, not stipulation" (Sperber, 1994, p. 42). I not only agree with this, I have long thought it contains a frequently overlooked kernel of wisdom about psychological "constructs" more generally. The argument that I make in the book, and that many biologists who now use the concept of modularity also make, is that modularity is a property we can *measure*. This does, of course, require some measurement criteria, which in turn depend on establishing a biologically sensible and quantifiable concept of modularity. The concept that most biologists and network scientists endorse has to do with the degree to which a system can be decomposed into sub-structures. When one applies this concept, the available evidence points to the mind's modularity as being hierarchical: modules at one level of organization are nested within modules at higher levels of organization. This nesting is not strictly hierarchical, but statistically so.

The Shape of Thought critiques a "two-layer" or "igloo" model of the mind as consisting of a set of "peripheral modules" (perceptual and motor, mostly), surrounded by a non-modular "central" system where, among other things, conceptual processing occurs. In the book, I do not argue that the distinctions made in two-layer models of this kind don't exist. For example, it's clear that some information enters conscious awareness and some doesn't; some processes are effortful and others aren't; some processes use lots of information and some use little; etc. What I argue is that there aren't *two layers*, with one composed of fast, automatic unconscious, evolved modules, and another not.

Instead I'm arguing something more akin to what Marvin Minsky was arguing in his *Society of Mind*: mental architecture is more like a pandemonium of diverse kinds of processes, interacting to produce the whole of cognition, including emergent features that result from these interactions. In the terminology I introduce in the book, at least some of these are likely to be *selectedly* emergent features. One thing about emergence is that to explain the products that emerge at one level of a system from interactions at lower levels of the system, you don't necessarily need to posit extra stuff at the higher levels. To take an often-used example, water has different properties than hydrogen atoms and oxygen atoms do alone, but you don't need to posit anything other than the hydrogen and the oxygen to explain the water.

In my discussion of a bulletin board / pandemonium / enzymatic model of emergent cognition, Morin claims that I'm contradicting myself:

Here (and in no other place in the book), I couldn't believe it was tofu because I suspected that, in fact, it wasn't. I need to squint hard to tell Barrett's "bulletin board" from the "System-2" of dual-process theories... where exactly is the difference? As an answer, Barrett accuses his opponents (the denizens of the igloo) of anthropomorphising the central processing unit, and treating the brain as "one undifferentiated blob" (p. 287). Is he being quite fair, though?

Maybe not; maybe there is more agreement than I thought (callback to introduction: sigh!). All I'm saying is that a pandemonium or enzymatic model has structure all the way up, not an unstructured middle part. Instead, it derives its flexibility and power from interactions. The interactions might be different and diverse throughout the system, but there's no natural dividing line one can draw between two layers; the global features are composed of the local features, as water is composed of hydrogen and oxygen.

When it comes to characterizing neural structure in a formal way there is not necessarily just one correct way to do it, but I think there is great utility in network-theoretic measures such as those employed by neuroscientists such as Bullmore and Sporns (2009). When you impose network-theoretic metrics there is no circularity; modularity has a clear technical definition (actually there are several related ones, but as long as you specify which one, it's clear). Generally, formal definitions of modularity have to do with how much clustering you see in the connections in a network, compared to what you'd see by chance. Importantly, modularity can be hierarchical, since it's measured with respect to particular scales of the network, so you can have modularity within modularity, and that is indeed what emerges from empirical studies of the brain's connectivity (Bullmore & Sporns, 2009; Meunier et al., 2010).

As the resolution of our data about the brain increases with technological advances, what do we see emerging from the fog? Not an igloo. Brain structure is not, as nearly as we can tell, modules around the edges with a non-modular central system in the middle. Instead, it's modules within modules, all the way up. These are fuzzy, statistically defined modules, but that's exactly what you'd expect from complex biological networks (it's also true of networks of gene interactions, networks of biochemical reactions, and other biological systems).

When I say "what you'd expect," here, do I mean, what I (or someone else) *predicted*? Hm. Not me, but maybe somebody did, and if they did, that would be very nice. But I hope it's clear by now that it's the object appearing out of the fog that most interests me, along with the best way to get there.

References

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10, 186-198.

Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, 9(4), 156–185.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (2nd ed., pp. 380–402). Oxford, UK: Blackwell Scientific.

Marechal, L., Genty, E., & Roeder, J. J. (2010). Recognition of faces of known individuals in two lemur species (Eulemur fulvus and E. macaco). *Animal Behaviour*, 79, 1157-1163.

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, 4, 200.

O'Bleness, M., Searles, V. B., Varki, A., Gagneux, P., & Sikela, J. M. (2012). Evolution of genetic and genomic features unique to the human lineage. *Nature Reviews Genetics*, 13, 853-866.

Poldrack, R.A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annual Review of Psychology*, 67, 20.1—20.26.

Somel, M., Liu, X., & Khaitovich, P. (2013). Human brain evolution: transcripts, metabolites and their regulators. *Nature Reviews Neuroscience*, 14, 112-127.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39–67). New York, NY: Cambridge University Press.

Comments

***

Daniel Burnston

Thanks for your interesting response. I would like to raise a couple of quick questions. As to your target in the book, I took you to be both (i) updating evolutionary psychology to get rid of the outdated “swiss-army knife of preformed traits” model, and (ii) to carve out a specific place for evolutionary theory in cognitive science (the “under the rubric of evolutionary explanation” claim). As I mentioned, I think the book was a smashing success with regards to (i). Even after your response, though, I am confused about how (ii) is supposed to work. You’re right that my critique was an attempt to update the standard epistemological worries, but I raised them specifically to cast doubts on (ii). I think it would be useful to more clearly articulate the options for how to read the claim.

It would be good to know both the organization of the mind and its evolutionary history.

It would be good to know both the organization of the mind and its evolutionary history, and hypotheses about each of these constrain the other.

It would be good to know both the organization of the mind and its evolutionary history, and (non-explicitly-evolutionary) cognitive science has erred in its methods by not giving explicitly evolutionary arguments.

I read the book as making something between claims (2) and (3) (what I called the “centrality” claim), and I am worried whether your argument supports either of them. It seems your response, “I fail to see the argument that we shouldn’t try,” is at most a statement of (1). I am not against (1), but it also doesn’t seem to bring anything under the explanatory rubric of anything. It just says “it’s better to know more than less.” So, it seems you need to argue for at least claim (2). My argument was meant to suggest that the constraints are not mutual, but only go in one direction. The worry that your style of evolutionary explanation is too backwards-looking is not just a claim about the value of prediction, it’s a question of whether (2) (much less 3) is sustainable on your view.

The two cases I bring up, of which you only discuss one, are meant to work together to bring this tension into focus. They’re not meant to suggest that we ever do have complete agreement in psychology—the question is whether our disagreements can be mediated by your evolutionary perspective. If we know all of the mechanistic facts, then the evolutionary considerations don’t matter for understanding the organization. Hence, no constraint. But—and this is where the two cases work together—if we don’t know all of the organizational facts, then there are not independently establishable evolutionary facts we can draw on to tell us what they are. I think, here, that Olivier and I were stating some of the same worries.

So, take your case of face perception. There are a number of cognitive scientists who have based a good chunk of their careers on arguing that the FFA simply isn’t specialized for faces. They read the data as suggesting that the expertise framework for describing FFA function is true, and think that the expertise framework is incompatible with the face-specialization claim. What you offer in your response is the purported evolutionary value of face perception. Should this convince anyone who’s a fan of the expertise framework? Well, they often suggest that the intuitive idea of an evolved specialization for faces is simply wrong, since it’s not supported by the data as they read it. Notably, the subsequent debate about specialization in the FFA has focused on data generation—how to get more fine-grained measurements, how to analyze the data, etc.—and not on discussing the evolutionary claim. So, even if we admit that the FFA must have an adaptationist story behind it, it seems there are two possibilities—either the FFA is specialized for faces, and evolved to represent them, or the FFA mediates perceptual expertise, and is evolved to do that—and the answer we arrive at depends entirely on non-evolutionary data. To insist otherwise seems like a return to the old kind of adaptationist reasoning I'd thought was being left behind.

Here is another case: Sandy Mitchell (e.g., her 2002) discusses how, in social insects, everyone used to think that particular social structures must have evolved—that they must have been selected for due to their advantages for hive survival. She and Robert Page, however, along with a number of other modelers, showed that social organization can emerge from much simpler principles, for instance forage-for-work algorithms, threshold-response algorithms, and learning algorithms. She takes these to be views of development of individuals within the hive, and says that they show the standard adaptationist intuition to be highly questionable. That is, if we discovered that the mechanism works in the way described in these models, the intuitive evolutionary claim would just be wrong. Your entirely correct claim in the book (and, I take it, the response) is that the developmental story is not incompatible with your view of evolution. I completely agree. The worry is just that we have to figure out the mechanism first—no amount of intuitive appeal for either evolutionary claim supports it until we've actually determined that the mechanism works in the way it describes. Of course, it would be good to know, whichever version of the mechanism turns out to be right, how the system evolved. But this is just claim (1), not (2) or (3).

So, the current mechanisms constrain evolutionary views, but I'm not sure the opposite is the case. To sum up: just how central, on your view, is evolutionary explanation supposed to be? One thing to do here—the way that I would be inclined to go—is to stress the heuristic usefulness of the adaptationist claims made above. They provide a target, for instance, that future empirical or modeling work can attempt to overturn. But this sounds like a much weaker view than you were pushing in the book. I'm not sure how this all interacts with your "embodied capital" case, since I'm not as familiar with that thesis. Perhaps you can expand upon that point in relation to the foregoing?

A final small point. I support your endorsement of network-based approaches to hierarchy and modularity. However, it is not at all clear that these divisions will map onto processing stages. Sporns and daCosta (2006), for instance, show that V4 is amongst the most centrally connected areas in the brain—and hence at the "top" of the network hierarchy—despite being decidedly "mid-level" in the traditional visual hierarchy. But it seemed to me your view of hierarchical processing was at least partially committed to the traditional visual hierarchy (i.e., first features parsed, then objects, then identiry). So I'm not sure this move is quite as anodyne for you as you suggest.

References

da F. Costa, L., & Sporns, O. (2006). Hierarchical features of large-scale cortical connectivity. The European Physical Journal B, 48(4), 567-573.

Mitchell, S. D. (2002). Integrative pluralism. Biology and Philosophy, 17(1).

Tadeg Quillien

Daniel Burnston raises the interesting question of whether the following claim holds:

"2) It would be good to know both the organization of the mind and its evolutionary history, and hypotheses about each of these constrain the other."

Looking at the psychological literature, it is easy to find plenty of evidence in support of the assertion. Even scientists who are not specialized in evolutionary biology routinely use hypotheses about the mind's evolutionary history to constrain their hypotheses about its organization. An entire book by Stanislas Dehaene (Dehaene 2009), is premised on the simple fact that cognitive mechanisms for reading cannot be an adaptation to reading. This simple (even obvious) premise leads Dehaene and other researchers to make non-trivial predictions, notably on the development of reading abilities (where children first do not make a distinction between symmetrical letters like b and d, as one would expect if reading co-opts more general visual recognition mechanisms), and the constraints that basic facts about visual cognition in ancestral environments impose on the form that writing systems take (Changizi, Zhang, Ye, Shimojo, 2006).

The demise of behaviorism provides another good example. The equipotentiality assumption (the thesis that any two stimuli can be associated equally well, regardless of their nature), at the heart of Skinner's framework, has been most forcefully challenged by researchers who first recognized that it made no sense from an adaptationist point of view, where organisms come equipped with inductive bets about their environment (Garcia and Koelling 1966, Seligman 1970).

And then there is the mountain of research produced by evolutionary psychologists, who generate highly original predictions, such as the fact that coalitional affiliation decreases categorization by race (Kurzban, Cosmides & Tooby 2001), and uncover the structure of domain-specific algorithms such as the one for kin recognition (Lieberman, Tooby & Cosmides 2007).

It is probably not the case that knowledge about evolution will be useful in all of cognitive science's endeavours, but its usefulness in generating novel predictions is by now beyond debate.

References

Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. The American Naturalist, 167(5), E117-E139.

Dehaene, S. (2009). Reading in the brain: The new science of how we read. Penguin.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. Psychonomic Science, 4(1), 123-124.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. Proceedings of the National Academy of Sciences, 98(26), 15387-15392.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. Nature, 445(7129), 727-731.

Seligman, M. E. (1970). On the generality of the laws of learning. Psychological review, 77(5), 406.

PS: Daniel Burnston also writes that "either the FFA is specialized for faces, and evolved to represent them, or the FFA mediates perceptual expertise, and is evolved to do that". However, those are not the only two possibilities. Looking again at Dehaene's work, the Visual Word Form Area is probably specialized for the processing of written language, but probably did not evolve to do that. But it seems very intuitive that there should be a one-to-one mapping between historical function and current specialization, and we need more researchers trained in evolutionary thinking to overcome these intuitions.

Clark Barrett

I am not claiming that evolutionary theorizing should play any more special role in cognitive science than it does in biology more generally. It seems to me that it is common in biology for the observations to come first, which are then used in various ways to generate and adjudicate between different theories about what gave rise to the observations. There are cases in biology where theory predicts things that are then observed, but those are rarer. Using evolutionary theory to explain observations seems to be standard operating procedure in biology, and so it should be in cognitive science.

I'm arguing that biology and cognitive science should be continuous in tools, theories, and standards. I'm not a huge fan of the language of constraint (as discussed at length in Chapter 1 of the book), but, perhaps where we can agree to disagree is regarding your claim that the "constraint" relationship between cognitive science and evolutionary theory is unidirectional. In my view, evolutionary theory "constrains" cognitive science to the degree that the mind is a biological entity and thus subject to principles of how biological things evolve. Cognitive science, in turn, "constrains" evolutionary theory through, among other things, information theory; organisms must receive information through physical means, can't be clairvoyant, etc. I guess I don't agree, then, that the constraints only go in one direction, and that once we have the "mechanistic" explanation, we're done.

Notably, even in cases in biology where we do have all the observations ("mechanistic facts," in your terms), additional theoretical work is (usually) necessary to explain those observations. For example, the persistence of sickle cell anemia in some environments was neither predicted by evolutionary theory (as a specific case), nor explained in the absence of evolutionary theory, which explains the distribution of alleles due to their fitness effects under heterozygote advantage (and is thus, in this case, an adaptationist explanation). An evolutionary model, not just an understanding of the sickling mechanism, is necessary to explain the distribution of genotypes that are observed.

What I'm claiming is that we don't have the complete explanation for biological things without explaining how they evolved, and mental mechanisms are those kinds of things. Where that puts me with regard to the three assertions you list I'm not sure; reading the statements, I think I agree with (1) and (2), but I don't think cognitive science has particularly "erred" (3). It's just that the mind's mechanisms, as the products of evolution, won't be fully explained without explaining how they came to be.

That's not to say we have the evolutionary explanation for most things in biology. We don't. But that's perfectly consistent with the idea that our explanations for most things are incomplete, which is the starting point for my book.