First, let me thank the ICCI for hosting this book club, and 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.