INTERFACE FOCUS

rsfs.royalsocietypublishing.org

Research

Cite this article: McShea DW. 2017 Logic, passion and the problem of convergence. *Interface Focus* **7**: 20160122.
http://dx.doi.org/10.1098/rsfs.2016.0122

One contribution of 12 to a theme issue 'Convergent minds: the evolution of cognitive complexity in nature'.

Subject Areas:
biocomplexity

Keywords:
intelligence, evolution, convergence, reason, emotion, Hume

Author for correspondence:
Daniel W. McShea
e-mail: dmcshea@duke.edu

Logic, passion and the problem of convergence

Daniel W. McShea

Department of Biology, Duke University, Science Drive, Box 90338, Durham 27708, NC, USA

DWM, 0000-0001-9398-0025

Our estimate of the likelihood of convergence on human-style intelligence depends on how we understand our various mental capacities. Here I revive David Hume's theory of motivation and action to argue that the most common understanding of the two conventionally recognized components of intelligence—reason and emotion—is confused. We say things like, 'Reason can overcome emotion', but to make this statement meaningful, we are forced to treat reason as a compound notion, as a forced and unhappy mixture of concepts that are incommensurate. An alternative is to parse intelligence in a different way, into two sets of capacities: (i) non-affective capacities, including logic, calculation and problem-solving; (ii) affective capacities, including wants, preferences and cares, along with the emotions. Thus, the question of convergence becomes two questions, one having to do with affective and one with non-affective capacities. What is the likelihood of convergence of these in non-human lineages, in other ecologies, on other worlds? Given certain assumptions, convergence of the non-affective capacities in thinking species seems likely, I argue, while convergence of the affective capacities seems much less likely.

1. Introduction

Survival is not logical. By that I mean a desire to survive does not follow strictly logically from the facts of a situation that threatens me. There is no principle of logic that connects the fact of a bus bearing down on me with a desire to get out of the way. Strict logic is a serious business. If all Minions are yellow, and Kevin is a Minion, then Kevin is yellow. No further assumptions are required. And while the premises are debatable, the conclusion—given the truth of the premises—is not. The same cannot be said for the bus and my desires. If I do not want to get out of the way, that is unfortunate for me, but—strictly speaking—no principle of logic is violated. The same goes for any set of facts and any desire. It is not logical to prefer to avoid an embarrassing situation. No desire, not even a mere preference, follows strictly logically from any set of facts.

So logic was the wrong word. Maybe we should have invoked intelligence or reason instead. We might say that certain preferences are intelligent, or that they are the product of reason. Reason and intelligence work better because these words connote more than just logic. For one thing, they imply various cognitive capacities like calculation and problem-solving. But this does not help us much, because like logic these too are neutral terms, carrying no component of desire. My laptop can calculate and solve problems but it does not *want* to do these things. Still, cognitive involvement aside, intelligence and reason also carry a component of motivation—of wanting, preferring or caring—a component that is often implicit. We say that it is intelligent to prefer to avoid an embarrassing situation, on the implicit assumption that I desire not to be embarrassed. It is rational for me to want to split the pie evenly, on the implicit assumption that I want to be fair. Or reason dictates that I want to feed and clothe my children, on the implicit assumption that caring for offspring is something I desire. These work because reason and intelligence are hybrid words, compound words, with both non-affective components (logic, calculation, problem-solving) and affective components (wants, preferences, cares) built into them.

THE ROYAL SOCIETY

Problem solved. But now a new problem arises, one with real consequences for our understanding of reason and intelligence generally, and more narrowly for the study of these capacities in other species. When we call crows intelligent or rational, what are we saying they share with us? A capacity for logic, calculation and problem-solving? Or the affective structure that motivates the use of these capacities? Or some combination of the two? The hybrid words leave us uncertain about what the question is. The same goes for the issue of the convergence of life elsewhere on human-style intellectual capabilities. When we ask about the likelihood of intelligent life elsewhere, are we asking about the evolution of the non-affective capacity to deploy logic and solve problems? Or about the evolution of wants, preferences and cares—of affective profiles—that are similar to ours? As I discuss later, our answers about the likelihood of convergence—or at least our first guess at answers—could be very different for affective and non-affective capacities.

Here I develop a case for rethinking the problem of convergent intelligence, for posing the questions we ask in a different way. To begin, I revive and defend David Hume's theory of motivation and action to argue that logic alone motivates nothing, and that all deliberate thought and action necessarily has a non-logical affective component, a component of what Hume called 'passion'. (The same goes for other non-affective capacities, such as calculation and problem-solving, which alone motivate nothing.) We need to keep the affective and non-affective separate, I argue. And compound words like reason and intelligence, which consist of an uncertain mix of the two, tend to muddle things. To promote clarity, I propose a revision of our vocabulary and a re-parsing of concepts, one that separates affective from non-affective capacities. With these properly separated, the question of the likelihood of convergence becomes two questions, one with an obvious answer and the other answerable in principle and requiring students of the subject to think about the problem in a new way.

Two caveats before beginning. First, in what follows, I am forced to use language that is unavoidably imprecise, unavoidably because the various categories of mental processes are not well understood, but also because the only words available to describe them have multiple meanings, usages that differ from one context to the next. One solution would be to impose order by inventing new terms, but that tactic has a poor track record. The other is to use conventional terms, and to try to overcome the imprecision by using a lot of them, using a number of descriptors to point the reader to concepts that lie roughly at the intersection of all of them. Thus, I use a trio of words, wants–preferences–cares, to refer to the calm affective capacities, processes of the mind that are united by their role in gently motivating thought and action. Importantly, in using these three, I do not mean to say that there are not other processes that motivate or other descriptors that overlap strongly with these. Further, I use logic–calculation–problem-solving to point the reader toward a subset of the non-affective capacities. No one of these identify exactly what I mean. The hope is that together they come close. And again I do not suppose the list is complete. Indeed, it is clear that there are many other non-affective capacities, including those involved in perception and memory (as well as an obscure storytelling capacity that I shall say something about later).

Second, my arguments are aimed at conventional or colloquial uses of certain words, not at the technical language that has evolved in philosophy, neuroscience, or comparative psychology. Indeed, words like reason and intelligence are typically avoided in technical treatments, with the result that the confusions addressed here do not arise. On the other hand, my arguments do extend to nontechnical treatments by thinkers in these fields, especially popular books in which these words are typically used freely and are prone to mislead. (See below.)

2. Hume's theory of motivation and action

Hume distinguished two kinds of logic, *a priori* and causal. (I lean here on Shaw's [1] understanding of Hume's theory of motivation and action, as well as Hume's *Treatise* [2] itself.) *A priori* logic is concerned with relationships among abstract ideas, and causal logic has to do with relationships among ideas about the physical world [1]. Actually, Hume uses the word 'reason', not logic, but as will be seen in the next section, it is clear from his argument that he understands reason more narrowly than we do today, and that our usage of logic better fits his meaning.

A priori logic is what allows us to conclude that if all A's are B's, and X is an A, then X is also a B. *A priori* logic allows us to conclude, after a number of steps in a geometric proof, that the sum of the interior angles of a triangle is 180°. *A priori* logic makes connections among abstract ideas, connections that are timeless, universal, and independent of any of the physical properties of the world. In contrast, causal logic enables an engineer to calculate the stress on a bridge support, based on his knowledge of physics and the properties of materials. Given certain materials, and certain principles of physics, the load limit for the bridge follows logically. Causal logic allows me to conclude, based on my knowledge of human nature that if I make a loud noise in a quiet library, people will look at me. Or rather, if I make a loud noise, many people in the library will *probably* look at me, since my knowledge of human nature is probabilistic. For present purposes, what matters is that both kinds of logic apply only to ideas, that is, to representations of some kind. The logician, the mathematician and the geometer are concerned with representations of abstract concepts, while the engineer and I are concerned with representations of physical objects and processes. Further, the relationship between premises and conclusion in both cases is one of implication or entailment, and therefore the output of a logical operation is in both cases also a representation. The engineer's representations of material properties and his representations of the principles of physics entail another representation, a number representing the stress on a bridge support.

Can either *a priori* or causal logic entail a want, preference, or care? No, because logic applies only to representations, the output of logic can only be more representations, and wants, preferences, and cares are not representations [1]. They are not ideas, they are not propositions, and they make no claim about the world. In modern terms, they are events, things that happen in a brain. And they contain—in Hume's words—no representative quality. So how, Hume's theory asks rhetorically, can a set of representations entail an event with no representative quality? It cannot, any more than a weather map can wash out a picnic.

Of course we have other motivationally neutral thinking capacities besides logic. Consider a capacity that might be called problem-solving. I have a problem: the neighbour's dog barks at me when I come home. Thinking about the

problem, I consult my experience, make some inferences based on it, and conclude that if I use the back door of my house instead of the front, the dog will not bark. And this discovered solution evokes in me a preference for the back door. But of course, again, evocation is not entailment. Problem-solving is a species of causal logic, a manipulation of representations of things in the world, and the logical consequence of these manipulations must also be a representation. It can never be a want, preference or care. It cannot be something with no representative quality. In other words, there is no necessary relationship between the solution to the problem of the barking dog—the idea that the back door would solve the dog problem—and my desire to use the back door. There is in fact a connection, but it has to do with the contingent structure of my brain, the way my problem-solving processes connect to my affective processes. The same goes for other non-affective thinking capacities, such as computation, which connotes the manipulation of mathematical symbols, another species of Humean *a priori* logic. And the same goes for calculation, which walks the line between *a priori* and causal logic.

Some concerns will arise. First, one might argue, ideas and representations are also brain events, so it should be possible for a representation-as-brain-event to cause a want, preference or care. The idea of the bus bearing down on me corresponds to some brain state or process, and that state or process physically causes fear in me, in other words, a desire to get out of the way. That is true, but causation is not the same as logical entailment. The sight or thought of the bus causes fear but does not entail fear. The bus causes fear physically, but not logically, just as lightning causes thunder physically but not logically. Now it is true that the idea of the bus, together with certain facts about human brain structure, might entail the idea of fear, enabling me to predict that I will be afraid. Likewise, the idea of lightning could, together with the laws of physics and certain physical conditions, entail the idea of thunder. But the idea that I will be afraid is not fear, and the idea of thunder is not thunder. Among other things, the idea of thunder cannot shake my house. Thus, the counterargument fails. Wants, preferences and cares can be caused. They can be *evoked*, under the right conditions, in a brain. But they cannot be logically entailed.

Second, the Humean claim that wants–preferences–cares have no representative quality would be controversial today. The cognitive school of thought in psychology (see [3]) might argue that all affective states—especially the emotions but also wants, preferences and cares—have a representative component in that they are intentional, that is, they are *about* something. I do not just care; I care about my cat. I am not just afraid; I am afraid of heights. And my cat and heights are present in my mind as representations, closely associated with the care or fear that I experience at the same time. The point is an interesting one, worthy of a longer discussion on some other occasion. Here I will just note that this view need not worry Hume's point. We can acknowledge a close association between an affective state and a representation without conceding that the affective state is identical to—or even partly comprised by—the representation. The relationship could instead be regular and causal, with the cause being the brain state associated with the idea of heights and the effect being fear of heights. That is, fear of heights might be *about* heights in that fear is regularly elicited or caused by the idea-of-heights brain state [4]. It might even be a special kind of fear, a mental fear-of-heights event that is distinguishable from other fear-of events. But still the relationship could be causal. And in that case, once again, there is no logical entailment. The regular elicitation of the fear by the idea, and the tight association between the two, would still be neurological, not logical.

3. Hume's reason and calm passion

As discussed, Hume's 'reason' encompassed *a priori* reasoning and causal reasoning, both of which in modern terms we would call non-affective. For affect, he used the word 'passion', which for him included what we think of as the emotions, especially the strong emotions, such as anger, fear, disgust, and so on. But he also identified a subcategory, what he called the 'calm passions' [2] described them as 'aversions' and 'propensities'. Shaw [1] gives a wonderful example:

> The most obvious application of this argument [for the calm passions] would be to cases of cool, seemingly passionless deliberation: e.g. in the course of planning the family budget I set aside funds for the children's education. A cursory inspection (or recollection) of the contents of my conscious mind at the time of acting would reveal to me some purely rational considerations that I had in mind at the time. I had some thoughts whose contents are (roughly) expressible in such words as 'If I don't put aside so much per month over so many years we will never able to see them through university.' But cursory introspection or recollection would not detect any separately identifiable desire-factor present to my consciousness at the time. Considered philosophical examination, Hume assures us, will reveal one, a 'calm' one—in this case, kindness to children [1, p. 58].

This pro-children sentiment, this propensity to favour them, manifest in this case as a desire to support their education, is one of Hume's calm passions. Also involved here, of course, are some of what Shaw calls purely 'rational' processes or what I am calling logical, computational, problem-solving, or more generally, non-affective processes. For example, there is the computation of the consequences of not saving money for their education. But these non-affective processes alone do nothing at all to motivate the act of saving. An affective component, a calm desire to support the children, is needed. Indeed, some such soft sentiment is needed to motivate the computation of consequences in the first place.

There is some debate about exactly how we are to understand the calm passions [1], but two things about them are clear. They are affective, in contrast to logic, calculation, problem-solving, etc. (i.e. in contrast to reason in Hume's sense), and they motivate thought and action. And that is sufficient for my purposes. The calm passions play the same role in driving thought and action that is played by the capacities I am calling wants, preferences, and cares. For example, it is my preference for not being embarrassed that motivates me to calculate a way to escape a mildly difficult social situation, and then to actually take steps to escape. Without a preference of some kind, no calculation of escape routes takes place, no action is motivated, and no action occurs. Like the strong passions, preferences motivate. But unlike fear and anger, a preference may not noticeably raise my heart rate. Indeed there may be no detectable physiological signs at all. A preference is affective, like emotion, but it is calm affect.

4. A failed ontology of mind

There is a habit of speaking, common in everyday discourse, that pits reason against emotion. I say that overcharging by

the merchant caused me to become angry, but I overcame that anger using reason. Or I say that the sight of the snake frightened me, although rationally there was no need to be afraid, because it was harmless. And even where reason and emotion are not opposed, they are conceived as separate brain functions. We find this separation captured imaginatively in science fiction, in beings with no emotion, supposedly guided only by reason (e.g. Mr Spock of the Star Trek series, and many others.)

This ontology, this notion that the mind contains these two sometimes conflicting capacities, reason and emotion, is also embedded in academic thought. It was made famous in Plato's metaphor of a charioteer—representing human reason—struggling to guide and control the chariot's two horses, one representing the virtuous passions and the other the appetitive and lustful passions. And it pervades modern thought, in philosophy, psychology, economics, and many other fields. The psychologist Jonathan Haidt [5] offers a modern version of this opposition between reason and emotion. Like Plato, he uses an animal metaphor, with a rider representing the self, guided by reason, and the animal representing the emotions:

> The image I came up with for myself . . . was that I was a rider on the back of an elephant. I'm holding the reins in my hands, and by pulling one way or the other I can tell the elephant to turn, to stop, or to go. I can direct things but only when the elephant doesn't have desires of his own. When the elephant really wants to do something, I'm no match for him [5, p. 4].

I think the current predominant ontology of mind—reason and emotion as independent capacities, sometimes in opposition—lines up pretty well with this story. But sadly, despite the story's superficial coherence, it does not work. There are two ways to understand reason versus emotion, both problematic.

1. On the one hand, we could understand reason as logic, calculation, problem-solving or some other non-affective capacity. But as Hume showed, non-affective capacities cannot motivate. The charioteer, or the elephant rider, cannot even *want* to influence the horses (in Plato's story), or the elephant (in Haidt's), nor could she have any preferences for what they do or do not do. The purely logical being of science fiction has no motivations and would not even get out of bed in the morning. With no affect, there is no force, no oomph, not even a leaning toward this behaviour rather than that. My laptop is a purely logical being. And it wants nothing.

Thus, emotion on the one hand and reason in the sense of logic, calculation, problem-solving on the other are not only unable to oppose each other but they are incommensurate. And claims about their opposition are not false, they are non sequiturs.

I grant that this will not be an easy case to make, so deeply embedded is this way of talking in ordinary discourse. And some who understand and accept the foregoing reasoning, will still be inclined to leave some wiggle room, some room for a slight, or even just occasional, opposition between emotion and reason in the sense of logic, calculation, and problem-solving. Consider the following seemingly innocuous remark, also from Haidt, regarding the opposition of reason and 'intuition', which he understands in this context to be an emotional response.

> People may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition. In such cases reasoning truly is causal and cannot be said to be the 'slave of the passions.' However, such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high [6].

Haidt elsewhere seems to accept Hume's well-known argument that 'reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them' [2]. But here he tries to open at least a small window for some very mild opposition by reason, opposition that can just occasionally lead to victory. This move might sound completely reasonable. Even slaves have a will of their own, a will that must occasionally allow them to stand against their masters, no? Emphatically no. Hume would have none of it. And perhaps he would have done well to use a different metaphor. Reason in the sense of logic, calculation, and problem-solving is not exactly a slave of the emotions, or more broadly, of affect. Rather, it is a tool of affect. It is a screwdriver. And no screwdriver has ever objected in the slightest to the use to which it is put.

2. On the other hand, we could understand reason more broadly, as including a soft affective component, Hume's calm passions. And then the conventional opposition between reason and emotion becomes a conflict between different kinds and intensities of affect, between the calmly passionate component of reason and the fiery passion of emotion. This opposition creates no conceptual trouble, and in fact this is the opposition that Hume argues occurs when we make difficult decisions. The urges of fiery passions are not blocked by logic or calculation, but by the softer, persistent urgings of calm passions, or more powerfully yet, by coalitions of calm passions. I feel a powerful urge to light into the merchant who overcharges me, but this urge is overcome by a coalition of lesser motivations: the embarrassment that would accompany making a scene in public, the desire to set an example for others watching me (possibly including friends and family), the distasteful-to-me diminishment of my sense of myself as a civilized person that would follow an outburst on my part, etc. The battle is between one kind of passion and another, between calm passion and strong passion, or in more modern terms, between soft emotion and strong emotion, nothing else.

Notice that the logical, calculating, problem-solving component of reason has a role in this scenario. In the moments after I experience an insult, these non-affective capacities go to work computing the consequences of various alternative actions. They tell stories or paint pictures for me of the likely outcomes if I, say, raise my voice, posture aggressively, or make threats. Drawing on past experience, they produce images of the people around me and their reactions to my behaviour. They generate a hypothetical future self, a future 'me' reflecting on my own behaviour. So far this process can be thought of as affectively neutral, but the next step is not. I respond affectively to these alternative scenarios, to the imagined responses of those around me and of my future self, in this case responding with aversion, disapproval. It is this last event, the affective response to the output of my non-affective scenario-generating processes, that competes with my anger to decide what I will do next.

So the broader understanding of reason would be fine, except that the word now becomes a complicated beast. Reason must now be recognized as a compound concept, combining a logical, calculating, problem-solving, storytelling component with a component of calm passion. We use

compound concepts all the time, for example, the idea of 'happiness' to describe some barely effable mix of health, access to resources, and achievement of our strongest and most enduring goals in life. But reason in the compound sense is different in that the components are incommensurate with each other. Logic and calm passion are not even apples and oranges, they are apples and gravity. (An appropriate parallel, perhaps, in that a falling apple is powered by gravity just as logic is powered by calm passion.)

5. An alternative and a proposed reparsing

There is an alternative to reason versus emotion. It is a parsing of the mind that recognizes the shared affective quality of what I am calling wants, preferences, and cares, a parsing that recognizes their status as affective processes akin to emotion, and therefore their essential role of affect in motivating all deliberate behaviour. This view denies that decision making is ever an opposition between affective processes and non-affective processes like logic, calculation, and problem-solving (whose role is instead to neutrally inform those affective processes). And it insists that all decision-making is a conflict among affective processes of some kind.

Encouragingly, despite the pervasiveness of the reason-versus-emotion view, the above view is not wholly foreign to conventional thinking. For example, when we praise a person for their logical, calculating, and problem-solving capabilities, it is understood that we are not necessarily praising their motivations. Implicitly, we are recognizing the valuing (affective) and value-neutral (non-affective) components of mind as independent.

My sense is that the field of psychology would find most of this agreeable as well, especially the notion that emotions, wants, preferences, and cares are all affective processes. But the attention of the field has mostly been elsewhere, on the study of *strongly* affective processes, often limited to the so-called basic emotions like fear, anger, joy, and so on. The questions asked have mostly to do with the expression of emotion [7,8], the physical correlates of emotion [4,9], emotions as adaptations [10,11], the localizability of emotion in the brain [12,13], and the relationship between cognition and emotion [14,15], including the role of emotion in 'biasing' cognition. And the strong connection defended here between emotions and wants–preferences–cares is rarely acknowledged explicitly. Damasio [16] comes close to acknowledging it, invoking the same conceptual scheme for calm decision making as for actions driven by strong emotion. Still he sees affect as biasing decision-making, rather than as constitutive of it. And the use of the word bias seems to allow that pure logic, calculation, or problem-solving, could motivate under certain circumstance, i.e. where biases are absent, something the Humean argument denies. On the whole, in the material in psychology and neuroscience I have surveyed, the theoretical position that emotions and wants–preferences–cares are the same sort of beast, that they are similar and closely related phenomena, is often implied, never denied, but only occasionally explicit, usually in passing.

I propose two things, one of them sadly impractical. First, we should eliminate the word reason, at least in its colloquial usage. The mix of logic, calculation and calm passion that it claims to capture does not align with any sensible parsing of concepts. Indeed, the point of the word reason in its modern sense is mainly to hide stances that are embarrassingly difficult (or just plain embarrassing) to explain inside a logical narrative that appears—like a mathematical calculation—to be self-explanatory. One hears things like: I wanted to punch the guy in the bar, but reason dictated that I restrain myself, so I turned and left. Or this: reason clearly indicated that more people can be housed if the parcel of land is developed. Or the rise of civilization was a triumph of reason. The implication of all of these claims, applying reason across a huge range of scales, is that human action can be motivated by cold calculation of some kind, by some value-neutral process requiring no further justification. And in each case, what it hides is the ensemble of (mostly) calm passions that in fact do all of the motivating, and that massively require justification while at the same time being complicated to justify. What makes not punching the guy in the bar a good thing? Why should we be so sure that more people need to be housed? Was the rise of civilization a good thing? The word reason not only does not say, does not explain, but it tries to head off, to short circuit, the critical evaluative questions, instead demanding of us a humble bow of obedience and acceptance. How dare we question reason! Obviously, this proposal—eliminating the word reason—is the impractical one.

My second proposal is that in technical discussions of the capacities of mind, if we must use the word reason, we use it in a technical sense to refer only to non-affective capacities. This is Hume's usage, so perhaps it would be apt to call it 'Humean reason' to distinguish it from colloquial reason (which in turn would be, if not abolished, at least frowned upon in such contexts). Consistent with this, I propose a scheme in which the conceptual division is the one Hume assumed, between affective processes (including wanting, preferring and caring, as well as the strong emotions) and non-affective processes (Humean reason, which includes logic, calculation, problem-solving and so on). My proposal is shown schematically in figure 1.

To be clear, this merging of calm and violent passions—of wants–preferences–cares and emotions—under the heading of affective capacities has a limited purpose: to reveal and formalize the separation between them and the non-affective capacities. For other purposes, we will certainly want to maintain distinctions among the various types of affect. Indeed, the emotions seem to be distinctive in a number of ways—for example in their somatic effects—and nothing in the argument here denies that.

Before proceeding, we do need to raise for consideration the possibility that the entire folk psychology of mind and all of its component capacities, from logic to wants to emotions, is false. These may not be natural categories at all. That possibility acknowledged, there is very little to be said about it. At this moment, given how little is known in psychology and neuroscience about the structure of the mind, there seem to be no alternatives.

6. Conclusion: convergence and the in-principle independence of affective and non-affective capacities

There are many ways to formulate a question about convergent intelligence. We could ask about the likelihood that high intelligence would evolve on other planets. Or the

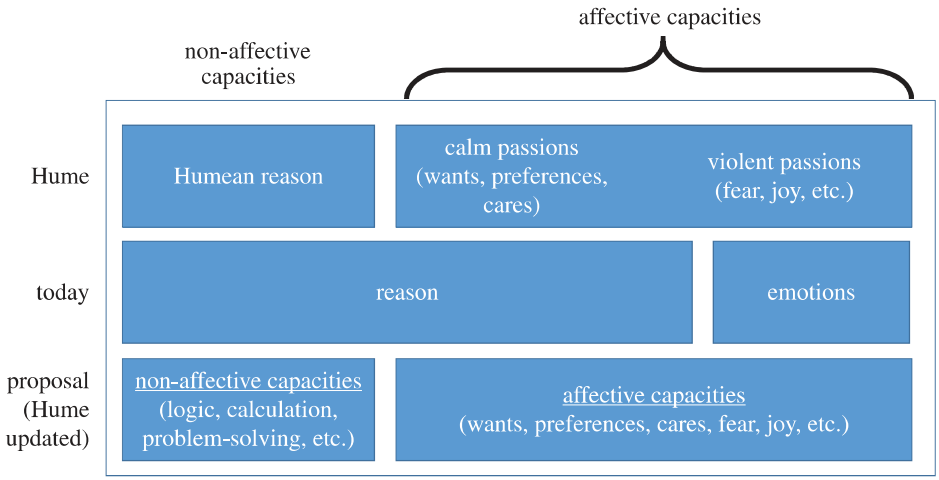

Figure 1. A proposed reparsing of mental capacities. See text. (Online version in colour.)

same question about high intelligence in any given earthly lineage. We could ask what level of intelligence—relative to human intelligence—has been achieved by certain specific lineages, say, chimps, dolphins, or octopuses. Under any formulation, regardless of whether the focus is elsewhere or here, whether it is on this species or that, it is easy to take for granted that we know what intelligence is. Intelligence is what humans excel at, isn't it? It is one of the traits we know best. It is a capacity we think we deploy often, every day, and one whose workings seem directly accessible to our always active minds.

The argument of this paper does not answer the question of what intelligence is, but it does disturb certain conventional complacencies about it and does so in a particular way. Regardless of which of the many possible understandings of intelligence we adopt, the question of convergence is really two questions, or rather, two broad sets of questions. There is a set of questions having to do with convergence on non-affective capacities: logic, calculation, problem solving and so on. And there is a set having to do with convergence on affective processes: wants, preferences, cares, as well as emotions. The point here has been that the capacities targeted by these two sets of questions are in principle independent of each other, and therefore the answers are independent of each other. Convergence on non-affective capacities is in principle independent of convergence on affective capacities. If it turns out that dolphins and dogs have convergently similar affective profiles—similar wants, preferences, cares—it does not follow that they have comparable capacities for solving problems in logic or geometry. For example, dolphins might have a three-dimensional problem-solving capacity that dogs lack. More interesting though is the converse point: crows and chimps could in principle have arrived convergently at the same facility with logic and geometry, but they might still have very different affective profiles. Crows might not even *want* to solve many of the problems that chimps want to solve.

7. Speculations on the likelihood of convergence

It is hard to resist the temptation to offer some tentative estimates of the likelihood of convergence for affective and non-affective capacities. I frame my speculations as answers to questions about the probability that other species that we are moved to call 'intelligent', here or elsewhere, will have capacities that are broadly similar to ours. (I put the word in quotes in recognition of the fact that its standard usage is problematic, as discussed.)

First, consider non-affective processes. It seems to me that logic is logic. Calculation is calculation. And no matter how many alternative neurological pathways there are to get from premises to conclusion, or from the set-up of a mathematical problem to its solution, the conclusion or the solution will be the same. The shortest distance to a food source for a creature of a given scale, with a given method of locomotion, etc., ought to be the same regardless of how it is computed, regardless of the software, hardware, or wetware employed. (Now I have heard that there are alternative logics, and even alternative mathematics, such as the various non-Euclidean geometries, but consideration of these is beyond my capability. I will say that they present no problem for my argument here—that the probability of convergence is high—if they all have the same inexorability, the same affect-independence, as the standard frameworks.)

So logic is logic, but that said, I must back off a bit by acknowledging that some of our most powerful non-affective capacities are much more open ended, much less predictable than logic or calculation. As discussed earlier, we seem to have the capacity to tell ourselves stories, a kind of search strategy for exploring the consequences of our thoughts and actions. Like all non-affective capacities, storytelling must be motivated by affect. And once motivated, it delivers alternative scenarios to which affect then reacts. Earlier I used the example of a merchant overcharging me. Here I will use a different example, one that involves no overt behaviour but instead plays out entirely in my own head. Suppose I am caught in a potentially embarrassing social situation. Perhaps a disparaging remark about me is imminent. I will find myself highly motivated to avoid the embarrassment before it becomes real. Some non-affective capacity, or combination of such capacities, races through alternative scenarios. I imagine myself walking away, or devising a way to spin the impending remark that defuses it, or devising some clever rejoinder that changes the subject. In the space of seconds, my imagination paints a picture of several possible futures, each a likely result of each course of action. And in the same breath, I react affectively to each, approving of those that seem most effective. The process is bracketed by affect, motivated by it beforehand and

evoking it at the conclusion, but it is itself non-affective. It neutrally tells a story, paints a picture, projects a future. What is interesting is that this process seems to underlie most, if not all, of our conscious decision-making. And we use it on all timescales, not just to make decisions in the heat of the moment, but to plan a day, to plan our lives. Damasio has aptly described it as: ' "minding" the future, predicting it, anticipating it in simulated form …' [16, p. 146]. I think of this storytelling capacity as the heart of human-style judgement, of human-style 'intelligence', and therefore as something it would be useful to know more about. At present, however, given the current state of our knowledge about it, it is hard to say anything about the likelihood of its converging.

For affective processes, the situation is clearer. That is, it is clear that no *a priori* answer on probability of convergence is possible. It might seem that certain affective commonalities are inevitable among species that we are disposed to call 'intelligent', that they are expected to share a common affective profile. One might argue that all such species are likely to have wants, preferences, and cares that dispose them to be social, to offer extended care to their young, to cooperate, to discourage certain kinds of non-cooperation, and so on. Still, all of these could vary enormously, in the precise ways they are satisfied, in intensity, in the context in which they are evoked. More generally, consistent with what we know about evolution, our default expectation for any feature that does not depend on universal truths (i.e. those of logic and mathematics) should always be variation, difference. That is, affective processes should differ enormously, even among species with well-developed non-affective capacities. The passion to survive may have exceptions in such a species that has a life history like black widow spiders. A preference for fairness in a honeybee-like species may be differentially weighted toward sisters and against brothers. A preference for avoiding embarrassment would likely be absent in any asocial species. It is questionable whether octopi embarrass. Tigers certainly don't. The point is that every species with advanced non-affective capacities is going to have a unique adaptive history guiding the evolution of its affective capacities, not to mention the divergent tendencies arising by chance. And therefore, a reasonable first guess would be that every species is going to have a unique affective profile.

Finally, convergence aside, the argument here raises a problem for the already fraught exercise of making comparisons across species lines. When we set out to compare non-affective intellectual capacities—logic, calculation and problem-solving ability—we must bear in mind that these capacities work always and only in the service of the affective capacities—wants, preferences and cares. And therefore intellectual capability is always relative to a species affective profile. To put it colloquially, thinking is useful to a species only to the degree that thinking gets it what it wants. So how are we to compare the capabilities of, say, dusky titi monkeys with those of baboons? Dusky titis are smart about getting what they want, say, about the nuances of maintaining a pair bond. Baboons are also quite smart, but about different things, like navigating dominance hierarchies. Since the two species want such different things, since they are motivated to apply their non-affective capacities for such different purposes, one wonders whether it is even meaningful to ask which is smarter.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

Acknowledgements. Thanks to Russell Powell, Irina Mikhalevich and the rest of the participants in the Convergent Minds Conference at Boston University (November 2015). Also thanks to two anonymous reviewers for helpful comments.

References

1. Shaw DJ. 1998 *Reason and feeling in Hume's action theory and moral philosophy*. New York, NY: The Edwin Mellen Press.
2. Hume D. 1978 [1740] *A treatise of human nature* (ed. LA Selby-Bigge), 2nd edn. Oxford, UK: Oxford University Press.
3. Lazarus RS. 1991 *Emotion and adaptation*. New York, NY: Oxford University Press.
4. Prinz J. 2004 Embodied emotions. In *Feeling and thinking* (ed. RC Solomon), pp. 44–58. Oxford, UK: Oxford University Press.
5. Haidt J. 2006 *The happiness hypothesis*. New York, NY: Basic Books.
6. Haidt J. 2001 The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814–834. (doi:10.1037/0033-295X.108.4.814)
7. Darwin C. 1998 [1872] *The expression of the emotions in man and animals*. New York, NY: Oxford University Press.
8. Ekman P. 1998 *Introduction to the expression of the emotions in man and animals*. New York, NY: Oxford University Press.
9. James W. 1884 What is an emotion? *Mind* **34**, 188–205. (doi:10.1093/mind/os-IX.34.188)
10. Plutchik R. 1994 *The psychology and biology of emotion*. New York, NY: Harper Collins.
11. Cosmides L, Tooby J. 2000 Evolutionary psychology and the emotions. In *Handbook of emotions*, 2nd edn (eds M Lewis, JM Haviland-Jones), pp. 91–115. New York, NY: The Guilford Press.
12. LeDoux J. 1996 *The emotional brain*. New York, NY: Simon and Schuste.
13. Barrett LF. 2015 Ten common misconceptions about psychological construction theories of emotion. In *The psychological construction of emotion* (eds LF Barrett, JA Russell), pp. 45–79. New York, NY: The Guilford Press.
14. Arnold MB. 1960 *Emotion and personality*. New York, NY: Columbia University Press.
15. Zajonc R. 1980 Feeling and thinking: preferences need no inferences. *Am. Psychol.* **35**, 151–175. (doi:10.1037/0003-066X.35.2.151)
16. Damasio A. 2003 *Looking for Spinoza*. Orlando, FL: Harcourt Books.