Forthcoming in B. Gaut and M. Kieran (eds.), *Creativity and Philosophy*, Routledge.

Talking about more than Heads: the Embodied, Embedded and Extended Creative Mind

Michael Wheeler

Introduction: the Path of Creation

In the opening paragraph of his fascinating book, *How Music Works*, the vocalist, guitarist, songwriter and former Talking Heads frontman, David Byrne,[1] describes what he calls the 'romantic notion' of creativity, in the following words:

> [C]reation emerges out of some interior emotion, from an upwelling of passion or feeling…. the creative urge will brook no accommodation…. it simply must find an outlet to be heard, read, or seen. The accepted narrative suggests that a classical composer gets a strange look in his or her eye and begins furiously scribbling a fully realized composition that couldn't exist in any other form. Or that the rock-and-roll singer is driven by desire and demons, and out bursts this amazing, perfectly shaped song that had to be three minutes and twelve seconds – nothing more, nothing less. (Byrne 2012: 13)

What Byrne calls the 'romantic notion' of creativity is, I suggest, the conventional wisdom about creativity: the essential properties of the created product are determined ('fully realized') both by and within certain internal psychological (emotional and cognitive) states and processes, and are then implemented in (they 'burst out into') the external world. In other words, on the romantic/conventional picture, creativity has an *inside-to-outside* logic of explanation. Of course, the details will occasionally be messy. Thus a songwriter might work out the final steps in a particular melodic progression by actually playing the guitar or the piano. But the creativity-romantic will treat any such goings-on either as exercises in fine-tuning, or as a matter of revealing the minutiae of a basically preformed inner creation.[2]

Byrne opposes the romantic/conventional picture. He says: 'I think the path of creation is almost 180° from [that] model. I believe that we unconsciously and instinctively make work to fit preexisting formats' (*ibid.*: 13). According to Byrne's alternative model, then, the internal psychological processes that contribute to acts of creation mould and adaptively fit their products to 'preexisting formats', where the term 'preexisting' signals, among other things, that the formats in question are external to the psychological processes concerned. (More on exactly what the term 'formats' means soon.) So the essential properties of the created product (e.g. the artwork) are determined by certain external structures that are somehow captured by the internal psychological processes concerned. In other words, for Byrne, creativity has an *outside-to-inside* logic of explanation. Having advocated this way of thinking, Byrne follows it to its natural endpoint: the most impressive examples of creativity – instances of true genius – are those in which the created product is best suited to a particular format. As he puts it: '[g]enius – the emergence of a truly remarkable and memorable work – seems to appear when a thing is perfectly suited to its context' (*ibid.*: 29).

In what follows, I shall argue that Byrne's official model of creativity fails to do justice to a complex and intertwined relationship that often obtains, between the internal and the external, in the path of creation. So, whereas Byrne resists the romantic/conventional account of

creativity by, in effect, reversing the direction of its explanatory logic – that is, by recommending that we shift from an inside-to-outside logic to an outside-to-inside one – I shall use a combination of theoretical articulation and illuminating examples to suggest that both these models are ultimately inadequate, because creativity routinely has an entangled, inside-*and*-outside logic. Just what this means will become clearer as we go along, but the tagline is that *the creative mind is embodied, embedded and extended*.[3]

Niche Products

To bring Byrne's account of creativity into better focus, it will be useful to approach it by way of some claims made by Margaret Boden (1990, 2004, 2010, this volume). Boden characterizes creativity as 'the ability to come up with ideas or artefacts that are *new, surprising and valuable*' (Boden 2004: 1). For now, let's focus on newness and surprise. (We'll be coming back to value later.) Boden unpacks newness in terms of a distinction between psychological (or individual) creativity (newness to the person who produces the idea or artefact) and historical creativity (the special case of psychological creativity in which the newness is also newness in human history). Given our interests here, it is psychological creativity that we really care about. Boden then proceeds to identify three kinds of surprise that correspond to three forms of psychological creativity. Roughly: (i) an unfamiliar combination of familiar ideas; (ii) an unexpected idea revealed within an already-available conceptual space (where a conceptual space is, roughly, a style of thinking; more on this below); and (iii) an apparently impossible idea resulting from a transformation of a conceptual space. Given this tripartite framework, one might be tempted to think of Byrne's favoured account of creativity as being tantamount to, and thus as being limited to, Boden's second form of creativity, that is, as a process of exploring already-available conceptual spaces. Indeed, if Byrne-style pre-existing formats were equivalent to, or versions of, Boden-style conceptual spaces, there would seem to be some mileage in this idea. Boden even appeals to the notion of 'fit' with an existing conceptual space when she introduces her second kind of creativity (*ibid.*: 3). Despite such lures, however, the invitation to think of Byrne's account in this sort of register should be resisted.

To explain: Byrne develops his account of creativity via a series of historical examples, and it soon becomes clear that the term 'format' should be given a very broad interpretation, because the set of external elements to which created products are claimed to be moulded and fitted comes to encompass a wide range of physical, social and technological structures. For example:

> *Contexts* (Byrne 2012: 29; see also 252-62): consider the music scene that grew up in and around the club CBGB in New York in the mid- to late-nineteen seventies, a scene which was pivotal in the careers of Patti Smith, Television and Talking Heads, among others. According to Byrne, this scene depended for its development not only on the talents of the various artists concerned, but also, and essentially, on a bundle of external factors such as, for example: appropriately sized and located venues that allowed artists to play their own music; the club policy at CBGB of allowing band members free entry on their nights off; a general social atmosphere of alienation from the prevailing music scene; and low-cost rents in the neighbourhood around the club so that artists could live in the area where the scene emerged.
>
> *Physical Spaces*: consider again CBGB, whose size, lack of reverberation and noisy bar-like atmosphere shaped the kind of music that Talking Heads came to write (*ibid.*: 14-15).

> The first two properties meant that the details in the music could be heard, allowing a certain intricacy in musical structure (witness also the complex dual guitar interplay of Television), while the third meant that the artists had to play loud.
>
> *Social Expectations*: Scott Joplin observed that jazz solos and improvisations were born in dance joints, because the written music would run out before the energy and enthusiasm of the dancing crowd did (*ibid.*: 21).
>
> *Technological Innovations*: the limited data storage available on early vinyl records meant not only that most pop songs in that era were written so as to be less than four minutes long (*ibid.*: p.92), but also that some classical music compositions started to incorporate decrescendos at the point where the record would need to be turned over, and crescendos at the beginning of the next side, in order to ensure a smooth transition in the listening experience (*ibid.*: p.93).

It is hard to see how the kinds of structures that Byrne identifies here as examples of pre-existing formats could qualify as Boden-style conceptual spaces. After all, Boden (2004: 4) glosses conceptual spaces as 'structured styles of thought' and as 'disciplined way[s] of thinking,' and her flagship examples include genres of music, dance, visual art and architecture, as well as scientific theories and styles of cooking. By contrast, Byrne-style pre-existing formats are a motley crew of physical, technological and socially distributed external factors that are perhaps most illuminatingly understood as the *environmental niches* in which the three forms of creativity highlighted by Boden unfold, and in which, therefore, conceptual spaces are explored and transformed.

This way of interpreting Byrne-style pre-existing formats allows us to couch the outside-to-inside logic of Byrne's account of creativity in a recognizably Darwinian language. Thus the path of creation, as Byrne understands it, may be described as a process of adaptive fit to some environmental niche (pre-existing format). More specifically, created products become adapted to their environmental niches through a process in which human psychological activity shapes the created products under development to those niches. Importantly, this Darwinian gloss on creativity is not a foreign imposition on Byrne's own way of thinking: he himself interprets the creative process in such terms. For example, he describes percussive African music as 'a living thing [that] evolved to fit the available niche' (*ibid.*: 16). Once we view Byrne's account in light of such remarks, it soon becomes clear that a relevant process of adaptive fit may take us beyond the 'mere' exploration of conceptual spaces, and be a factor in the transformation of those spaces. Indeed, when discussing the CBGB-centred New York music scene, Byrne characterizes its elements (see above) as 'guidelines that can steer you away from what might at first seem like obvious or logical moves' (*ibid.:* 266). Steering of this sort may lead one to unexplored territory in an existing conceptual space (such as, arguably, Television's use of technically proficient dual guitar interplay within a musical sub-genre born of garage rock), but it may also lead one out of one's established conceptual spaces altogether (such as, arguably, Patti Smith's experimental fusion of punk rock and poetry). In Darwinian language, one might think of the latter, transformative kind of episode as a speciation event, in which the adaptation of some created product to the environmental niche provided by a certain pre-existing format results in various traits of an existing creative form being modified in such a way that we witness the emergence of a new creative form.

It might seem that the Darwinian terminology with which we have just been working is meant to be heard in a wholly metaphorical register. Is traditional percussive African music

really to be categorized as a living organism? However, although this particular aspect of Byrne's striking image is (presumably) not to be taken literally, there is good reason to treat the claim that such music is under selective pressure to adapt to its environmental niche as a full-blooded truth. After all, as an abstract logic of how change in populations occurs, Darwinian selection has been applied to non-living entities such as, but not limited to, scientific theories (e.g. van Fraassen 1980), technological knowledge (e.g. Ziman 2000) and configurations in industry (e.g. Nelson 2002). Moreover, in another area where a dominant inside-to-outside explanatory logic has been challenged, namely the study of the evolution of language, the idea that a human-produced structure is rightly conceptualized as an entity that is under selective pressure to adapt to its environmental niche has already been advanced as a literal truth – and to great scientific advantage.

To explain: Natural languages have the property of compositionality. That is, they feature atomic elements (words) that may be combined using structure-sensitive rules (grammars), so as to yield complex molecular structures (sentences) whose meanings are a systematic function of the meanings of the atomic elements concerned plus the structure-sensitive rules by which those elements are combined. Now, according to a certain popular psycholinguistic picture, the linguistic-haves and the linguistic-have-nots in our world possess fundamentally different brains, in the specific sense that the brains of the former, but not those of the latter, are thought to contain some sort of innate domain-specific language processing system, one whose fundamental structure is itself compositional in form (e.g. a Chomskyan language acquisition device; see e.g. Chomsky 1986). On this view, the structural properties of public language come from the innate domain-specific language processing system. That's about as inside-to-outside as an explanatory logic gets. However, Simon Kirby and his colleagues have used computer simulations to show that if (a) one places language in its correct learning context, that is, one conceptualizes it as being passed on from one generation to the next by cultural, rather than genetic, transmission, (b) one begins, as seems evolutionarily likely, with a holistic language (one in which there is no systematic mapping from the structure of the symbols used to the structure of the meanings conveyed), (c) there is, as many have suggested, a transmission bottleneck in the language learning process (such that learners are exposed only to some impoverished subset of the language), and (d) language learners have a rudimentary domain-independent generalization capacity, then the language in question will evolve compositionality.[4] This result comes about because (roughly) compositional languages are generalizable languages, and generalizable languages can be recreated in each generation without exposure to the whole language. This makes compositional languages more evolutionarily stable, so that once a generalizing learner, by chance, stumbles across compositional structure, that structure will spread throughout the population.

Whether the explanatory logic of the model of language evolution advocated by Kirby and colleagues is strictly outside-to-inside, in a way that is equivalent to the explanatory logic of Byrne's model of creativity, is not what matters here. The key point is that, on both accounts, a human-produced structure is conceptualized as evolving to fit its environmental niche: natural language evolves under selection pressures established principally by the properties of its evolutionary conduit, a distributed system of human brains communicating via information bottlenecks, while created products evolve under selection pressures established by a range of pre-existing formats (contexts, physical spaces, social expectations, technological developments). Given that the key claims made by the former model are literal in character, and given that the former model is in good scientific standing, we have some reason to treat the pivotal claims made by the latter model as being more than mere metaphor.[5]

Cultural transmission gives us a powerful way of thinking about a *population-level* process by which the adaptive fitting of created products may happen. Such products evolve to fit environmental niches in virtue of the selection of variations and recombinations in culturally transmitted structures. Thus the intricate and layered rhythms of traditional African percussive music, which are passed on from one cultural generation of drummers to the next, are musical forms that don't become obscured in the open air as they would in many indoor spaces. In this way, they are adapted to their sonic environment. On Byrne's account of creativity, however, adaptive fitting is not only a population-level process. As we have heard already, individual creative agents 'unconsciously and instinctively make work to fit preexisting formats' (Byrne 2012: 13). It is here, at the level of individuals and their psychological states and processes, that the adaptive fitting model, as Byrne characterizes it, seems, to me, to overlook something interesting and important about the path of creation. More precisely, it fails to do justice to the complex and subtle relations between 'inside' and 'outside' that are exhibited, in many examples of creativity, by organic human brains in dynamic causal entanglement with the very kinds of external elements that Byrne highlights, elements such as contexts, physical spaces, social expectations and technology. So, when we think about individuals and their psychological states and processes, we need to shift not to an outside-to-inside logic, but to an inside-*and*-outside logic. According to the latter logic, creativity is not so much a matter of the inner fitting of its products and processing to the structure of the outer (a direction of explanatory travel which, in a sense, privileges the contribution of the outer), as it is a kind of interactive dialogue or partnership (in which inner and outer make interlocking and complementary contributions). In order to bring this more entangled path of creation into view, I shall briefly investigate three ways in which it is realized, in the embodiment, the embeddedness, and ultimately the extension of the creative mind.

Embodied Creativity

According to the embodied cognition perspective, psychological states and processes are routinely shaped, in fundamental ways, by non-neural bodily factors. One's first reaction to this proposal might be that it is unclear what it amounts to, until someone tells us what the terms 'shaped' and 'fundamental' mean. That much is certainly true, but here I shall not attempt to provide the kind of theoretical discussion that would ultimately be needed. In part, this is simply a matter of available space. The embodied cognition 'movement' is home to a miscellaneous assortment of projects. Thus embodiment is said to shape, in fundamental ways, things like the nature of concepts (e.g. Lakoff and Johnson 1980), the nature of perceptual experience (e.g. Noë 2004) and the nature of the processing mechanisms that enable action (in the philosophical literature, see e.g. Clark 1997, 2008; Wheeler 2005, 2011). This list is not exhaustive. Given such a diversity of topics, it seems unlikely that the terms in question will reward concise or general definitions. Fortunately, however, we can make do with a rough-and-ready negative characterization of the key idea, in the sense of a general picture of what it is to believe that psychological states and processes are *not* routinely shaped, in fundamental ways, by non-neural bodily factors.

In its orthodox form, contemporary cognitive theory has a neuro-centric bias. This bias isn't irrational. For example, cognitive science has taken great strides forward precisely by working on the assumption that the brain is where the action is. Nevertheless, from an embodied cognition perspective, neuro-centrism remains a bias, and a misleading one at that. Of course, even the most hardened fan of the brain believes that the non-neural body plays a

cognition-enabling role, in helping to harvest perceptual data for the brain through movement and orientation, and in executing the motion instructions that are generated by the brain in action. Moreover, on the orthodox model, the non-neural body can certainly affect what we think and do by (in a way that will soon be illustrated by an example) placing constraints on the representational and instructional states formed by the neurally located mind. However, from the perspective of orthodox cognitive theory, these contributions remain essentially no more than input-output bit-parts in the great psychological drama. At root, the relationship between the cognitive system and the non-neural body is tantamount to the relationship between a computer and its keyboard-and-monitor peripherals. As one might put it, the non-neural body provides the keyboards and the monitors for the neurally located computational mind. The embodied cognition perspective says that non-neural bodily factors are far more important to our psychology than *that.*

One's second reaction to the embodied cognition perspective, as stated, might be that, in the context of the present discussion, it is beside the point, since it does not imply causal entanglement with external elements. In fact, to the extent that the embodiment in question is grounded in bodily acts, such as, say, the physical manipulations of instruments or tools, embodiment naturally encompasses a rich mode of environmental interaction (which is to say that there is a natural route from the embodied mind to the embedded mind – see next section). Moreover, the notion of 'internal' is itself slippery, since the boundary between internal and external is, in some contexts, fixed by the skin (in which case gross bodily forms count as internal), while in others it is fixed by the limits of the brain or central nervous system (in which case gross bodily forms count as external). For the purposes of the present section, I shall exploit this ambiguity and treat the non-neural body as an external factor, on the grounds that Byrne's characterization of the romantic view of creativity clearly takes the essential properties of created products to be determined by psychological states and processes that contemporary cognitive theory ordinarily takes to be located in the brain.

So much for embodied cognition in general. Let's turn specifically to embodied *creative* cognition. As just noted, the orthodox, neuro-centric approach to mind dovetails neatly with the inside-to-outside model of creativity in which the essential properties of created products are determined in the brain and then simply implemented by, or via, the body, perhaps with some feedback-driven fine-tuning. But you may ask yourself whether this model is adequate to explain certain high-grade examples of creativity, such as playing (and, by extension, writing) the delta blues like Robert Johnson. To see why neuro-centrism comes up short, consider a science-fiction scenario in which Robert Johnson's brain is transplanted into my non-neural body. In spite of the fact that the pre-operation me has some guitar playing ability, there is, I submit, no guarantee *at all* that this operation would result in a being (whoever it is – personal identity is not the issue here) who is capable of generating country blues songs based on majestic guitar patterns such as those that drive 'Preaching Blues' or 'If I had Possession over Judgement Day.' To revisit one of Byrne's images, whatever desires and demons our newly constructed hybrid artist might harbour in his brain, they are unlikely to burst out as perfectly shaped, Johnson-style country blues songs. The reason for this is that non-neural bodily factors such as the muscular adaptations in Johnson's arms and hands are partial, but nevertheless essential, determinants of Johnson's distinctive guitar playing style and thus of his ground-breaking outputs.

In response to this claim, the neuro-centrist might endeavour to quarantine the identified shortfall as 'no more than' a deficit in the realization or implementation of the created structures, rather than as evidence of an inadequacy in the neuro-centric model of creativity

itself. But this strategy is doomed to fail, as becomes clear, I think, if we borrow some analytical machinery from John Haugeland (1998). The neuro-centric model of intelligent (including creative) action works like this. The brain (the seat of the mind) decides what needs to be done. It then sends a set of instructions to the body, in order to generate appropriate movements. At the point of sending, the semantic content of the instruction – 'play a D7 chord,' say – is settled. The job of the arms and hands is to decode that semantic content, and thus the transmitted instruction. But, as Haugeland observes, this image of decoded instructions underestimates the intimacy of the coupling between (i) the signals sent out from the brain down neural pathways to the muscle fibres in my arms and hands, plus the tactile and proprioceptive feedback signals produced by the subsequent movements, and (ii) non-neural bodily factors such as the lengths of my fingers, and the strengths and response profiles of my muscles and joints. If this is correct, then the neural signals highlighted in (i), including crucially those signals sent out by the brain that are candidates for being instructions, will have to be understood as *entirely specific to a particular embodied individual*, meaning that 'there need be no way – even in principle, and with God's own microsurgery – to reconnect my neurons to anyone else's fingers, such that I could reliably type or tie my shoes with them' (*ibid.*: 225). The same goes for our example of playing the blues like Robert Johnson.

At this juncture, it might be suggested that Robert Johnson's newly transplanted brain could, after some stretch of time, succeed in the generating Johnson-like delta blues songs using my body, by training my muscles to play in a specifically Johnson-esque style. There will, of course, be limits to this adaptive process (Johnson's brain presumably couldn't stimulate the growth of longer fingers, for example), but that's not the most telling response to the objection. Rather, one needs to focus on the very form of the suggestion. Why does the critic reach straight for the thought that my muscles could be adapted to the playing style of Johnson's brain, rather than the thought that Johnson's brain could be adapted to the playing style of my muscles? The answer, I submit, is that the objection simply assumes that the phenomenon of playing style resides in the brain. But of course that's a presumption that the advocate of embodied creativity will want to deny as a neuro-centric prejudice. Indeed, if creativity is properly embodied, then style is something that is neither solely in the brain nor solely in certain non-neural bodily factors, but instead is distributed over both. So although it might be true, in any particular case, that a transplant victim will end up producing artworks in a style that is associated with the formerly existing embodied individual where the brain in question was previously housed, there need be no guarantee – even in principle, and with God's own training regime – that this will be possible.

Taking a different approach, the neuro-centrist might complain that the thought experiment under consideration does not, in fact, undermine the instructionist model, but shows only that one should relativize the content of certain neural signals to a particular body. On this view, the signal from Robert Johnson's brain counts as an instruction to 'play a D7 chord' only in the context of Robert Johnson's non-neural body. One might give further substance to (I'm tempted to say 'flesh out') this idea by claiming that being in a particular body will give rise to idiosyncratic psychological schemata or heuristics, and thus to instructions that will most likely fail to be implemented correctly in a different body. However, if we respond to the intimate character of the relation between the neural and the non-neural bodily contributions here without assuming a neuro-centric perspective at the outset, there seems to be no good reason to think of the implicated muscles and joints as merely decoding messages sent by the brain, as opposed to being integral parts of the psychological processing concerned (*ibid.*: 226). If that's right, then one can no longer think of tracks such as 'Preaching Blues' or 'If I

had Possession over Judgement Day' as being pre-formed (in all their essential aspects) in the brain, and simply poised to burst out into the world via the non-neural body. The non-neural body is a proper part of the creative psychological machinery in play.

Embedded Creativity

The embedded cognition perspective seeks to register the important, and sometimes necessary, causal contributions made by environmental elements (paradigmatically, external technology) to many cognitive outcomes. More precisely, according to the embedded view, the distinctive adaptive richness and/or flexibility of intelligent thought and action is regularly, and perhaps sometimes necessarily, causally dependent on the bodily exploitation of certain environmental props or scaffolds. Despite the fact of this dependence, however, the embedded theorist continues to hold that the actual thinking going on in such cases remains a resolutely skin-side phenomenon, being either brain-bound or (on a less common, more radical iteration of the view) distributed through the brain and the non-neural body. In short, the embedded theorist seeks to understand the ways in which brain-bound, or perhaps body-bound, thought is routinely given a performance boost by its external (technological) ecology.[6]

One's first reaction here might be that it is simply obvious that creative cognition involves environmental scaffolding and so will qualify as embedded. After all, part of the aesthetic power of Gustave Courbet's oil paintings may surely be traced to the tactile surfaces he produced by the thick layering of paint; and The Byrds' reworking of Dylan's Mr. Tambourine Man might have sounded significantly less interesting if Roger McGuinn had played it on anything other than a Rickenbacker 12 string guitar. But this would be to set the bar for embeddedness too low. For there is in truth a genuine, although no doubt fuzzy, difference between the kinds of external causal contribution that will attract the attention of the embedded theorist and those that won't. To reveal this difference, we can begin with a distinction made by Boden and Edmonds (2009). In what is sometimes called 'computer-assisted art', the computer functions 'merely as a tool that remains under the close direction of the artist, rather like an extra paintbrush or a sharper chisel' (*ibid*.: 137). Such cases are to be contrasted with certain other human-computer creative couplings, such as those operative in computer-based generative art, a genre in which the visual artwork is produced, in part, by sets of abstract rules that are implemented, by a computer, in a manner which is not under the artist's direct control. In these more interesting cases, the computer is illuminatingly described as being 'partly responsible for coming up with the idea itself' (*ibid.*: p.138). Both computer-assisted art and computer-based generative art manifestly involve causal contributions from an external technological element, namely the computer. However, in computer-assisted art, the computer is naturally conceived in line with the orthodox neuro-centric model, that is, as no more than a fancy tool (a sharper chisel) for implementing instructions formed in the artist's head, instructions that specify the essential features of the artwork. In computer-based generative art, however, the computer makes a more fundamental causal contribution, by accounting for the some of the distinctively creative aspects of the work. This is the creativity-specific analogue of accounting for the adaptive richness and flexibility of intelligent thought and action, as mentioned in the generic definition of embedded cognition given earlier. So, unlike computer-assisted art, computer-based generative art is an example of embedded creativity.

Now, however, we need to make sure that the bar for embeddedness hasn't been set too high. Lurking in Boden and Edmonds' phrase 'partly responsible for coming up with the idea

itself' is an implication that, in computer-based generative art, the computer is, in some way, a locus of its own agency or intelligence. But an external technological element may account for the some of the distinctively creative aspects of an artwork, and thus figure in an event of embedded creativity, without exhibiting its own agency or intelligence. Thus consider Byrne's autobiographical account of how he wrote the lyrics for albums such as Talking Heads' *Remain in Light*. Starting with pre-existing instrumental tracks and fragments of melody, Byrne churned out pages and pages of semantically arbitrary (often nonsense) words and phrases whose syllables (he judged) 'worked with' the music. Then he scanned the pages for groups of words that hinted at a subject matter, and he finished off the song from there. As Byrne (2012: 199) describes this process: 'The lyrics may have begun as gibberish, but often, though not always, a "story" in the broadest sense emerges. Emergent storytelling, one might say.' In harmony with an embedded inside-and-outside logic, the final form of the created product here (the lyrics, with their full 'storytelling' aspect) emerges from a dynamic interactive relationship between an external resource (the pre-existing instrumental music) and the artist's internal psychological processing (Byrne's inner linguistic capacities), a relationship in which the external resource genuinely accounts for the some of the distinctively creative aspects of the artwork, and thus is revealed to be, in a deflated sense at least, a partner or participant in the creative process.

A more telling example of embedded creativity occurs during Byrne's brief history of Western classical music as a succession of adaptations to different environmental niches (*ibid.*: 16-21). This is the moment where, by Byrne's own admission, his official model of creativity struggles to cope with the data. However, although the case in question escapes Byrne's own official model, it is, I shall argue, revealingly understood using a key concept from the embedded cognition literature.

According to Byrne, the fact that Western music in the middle ages was standardly performed in huge stone-walled gothic cathedrals with long reverberation times explains why such music evolved to feature modal structures characterized by very long notes: such music minimizes the risk of overlaps and clashes that is present in that particular sonic environment. By the late 1700s, things had changed. Mozart was performing his compositions in grand, but not gigantic, rooms populated by plenty of people in extravagant dress. This different sonic environment deadened the sound and allowed elaborate details in the music to be heard. Mozart composed his music accordingly. He also made his orchestra larger, so that the sound produced could be heard over the noise of the people dancing and talking. And so on. Byrne's official model of creativity as adaptive fit to pre-existing formats copes comfortably with historical example after historical example. But then along comes Wagner, and the model falters (*ibid.*: 19-20). In the 1870s, Wagner had the Bayreuth Festival Theatre built to support a music that he had only hitherto imagined, a more bombastic, dramatic music. His architectural innovation was to introduce a bigger orchestra area that could accommodate more musicians (especially in the bass section) and larger brass instruments. In other words, as Byrne depicts this historical event, with the design of the Bayreuth Festival Theatre, Wagner did not create musical artworks to fit a pre-existing format. Rather, he created a format with a certain kind of music in mind, and then produced the music, most notably *Parsifal*, to fit it.[7] This is, I suggest, a creativity-oriented version of a process that embedded cognition theorists know and love, one that Andy Clark has dubbed *cognitive niche construction* (e.g. Clark 2008; see also Wheeler and Clark 2008).

Niche construction (e.g. Laland et al. 2000) is a recognized biological phenomenon in which animals act on their environments and in so doing modify the selection pressures that shape

the subsequent path of biological evolution. Thus, in a widely cited example, beavers collectively build dams that alter the ways in which rivers flow. These created structures become part of the selective environment in which beaver populations evolve, since future generations of beavers inherit (non-genetically, cf. cultural transmission) both the dams and the altered river flows that they produce. *Cognitive* niche construction is the specifically psychological, and sometimes intra-lifetime, version of this generic biological phenomenon, a version in which human beings build external structures that, often in combination with culturally transmitted practices, transform problem spaces in ways that promote (or sometimes obstruct) thinking and reasoning (Clark 2008; see also Wheeler and Clark 2008).

To bring cognitive niche construction into view, I shall re-use a compelling example that Clark sources from Beach (1988). Consider the way in which a skilled bartender may achieve the successful delivery of a large and complex order of drinks. Fulfilling such an order can strike the casual observer as a relatively daunting task, especially if the memory-relevant resources available to the bartender are thought to be restricted to inner storage and recall. However, it is a fortuitous fact that different kinds of drink often come in differently shaped glasses. (Think of cocktails.) The ecology of the bar is thus characterized by some relatively persistent physical structures (the differently shaped glasses), plus some culturally established norms (the specificity of kind of drink to shape of glass). What novice bartenders learn to do, under the tutelage of their experienced colleagues, is to retrieve the correct glass for each drink as it is requested, and to arrange the differently shaped glasses in a spatial sequence that tracks the temporal sequence of the drinks order. From the perspective of the purely organic cognitive resources available, what the bartender has learnt to do, in replicating this culturally transmitted practice, is to exploit her physical environment in order to outsource complexity and so reduce the burden on inner processing. In effect, she transforms what might have been a highly challenging memory task into a simpler (roughly) perception and association task.

The bartender's practice is a case of cognitive niche construction, and a compelling example of the kind of coupled partnership between inner neural resources, embodied actions and designed environmental structures on which advocates of the embedded cognition perspective get to dine out. And while Wagner's theatre building adventure may indeed escape Byrne's official model of creativity, it is, I submit, another powerful illustration of embedded cognition and, more specifically, of cognitive niche construction. The constructed cognitive niche of the Bayreuth Festival Theatre enabled a certain kind of creative musical thinking, and thus a distinctive created product, in this case a new, surprising and history-changing species of opera.

Extended Creativity

Advocates of extended cognition hold that the physical machinery of mind sometimes extends beyond the skull and skin. More precisely, according to the hypothesis of extended cognition, there are actual (in this world) cases of intelligent thought and action, in which the material vehicles that realize the thinking and thoughts concerned are spatially distributed over brain, body and world, in such a way that certain external (here meaning 'beyond-the-skull-*and-skin*') factors are rightly accorded fundamentally the same cognitive status as would ordinarily be accorded to a subset of your neurons. So, to be clear, 'extension' here has the sense of *spatial* (environment-encompassing) extension, not performance enhancement, although, in some cases of (spatially) extended cognition, psychological performance will indeed be enhanced. Eye-catching examples of external elements that advocates of extended cognition often take to have such cognitive status include smartphones, tablets and at least

some instances of wearable computing, but, in the end, less attention-grabbing items such as notebooks (the old-fashioned kind), tally sticks and abacuses would, under the right circumstances, do just as well.[8]

Given that the embedded perspective already presents itself as doing justice to the significant contribution made by environmental elements to thought and action, one might wonder how the transition from embedded cognition to extended cognition is supposed to happen. This is a complicated and contested issue, a proper treatment of which would take us too far afield, so we will have to make do with a (very) brief statement of the general shape of my own favoured approach (for the arguments and details, see e.g. Wheeler 2011, 2013). What is needed, it seems, is an account of precisely which causal contributions to thought and action count as cognitive contributions and which don't. For my own part, I see no way of meeting this demand, except by appeal to what Adams and Aizawa (2008), two thinkers who are in fact front-line opponents of the extended cognition view, have dubbed a *mark of the cognitive*.

A mark of the cognitive is a scientifically informed account of what it is to be a proper part of a cognitive system that, so as not to beg any crucial questions, is independent of where any candidate element happens to be spatially located. Once a candidate mark of the cognitive has been placed on the table, further philosophical and empirical leg-work will be required to find out (i) whether that account is independently plausible, and (ii) just where cognition (so conceived) falls – in the brain (as the neuro-centrist and some embedded theorists think), in the brain and the non-neural body (as the embodied cognition theorist and some other embedded theorists think), or, as the fan of extended cognition predicts will sometimes be the case, in a system that is spatially distributed across brain, body and world. In principle, then, one might identify an external causal contribution to psychological performance – a generative art program, say – that, by one mark of the cognitive, counts as cognitive, and thus as part of an extended mind, but which, by an alternative mark of the cognitive, fails to count as cognitive, and so forms part of an embedded story. We would then need to decide which mark of the cognitive is better.

So, what is the mark of a *creative* cognitive process? Clearly, this question could be the spark for a wide-ranging discussion in the philosophy and psychology of creativity. In what follows, however, I shall simply select an independently plausible candidate for such a mark and ask whether the psychological process thereby picked out is an extended one. This will be sufficient to illuminate the general shape of the debate that concerns us here.

Matthew Elton (1994, 1995) has argued that a genuinely creative process may be distinguished (conceptually at least) into two distinct phases: *generation* (subsuming the preparation and the incubation of the idea) and *evaluation*.[9] Regarding the latter, one might gloss Elton's account as incorporating, within the creative process itself, a sensitivity to something like Boden's notion of the *value* of a product (see above). Importantly, for Elton, while the generation phase of the creative cognitive process may involve the kind of structured rule-governed routines that characterize something like computer-based generative art, it need not. Elton's evidence here comes partly from the work of the experimental musician and composer Brian Eno. Eno regularly introduces chance and unpredictability into the generation phase of creation, including (a) the deliberate use of a model of synthesizer (the Yamaha DX7) known to have an erratic oscillator, and (b) the random selection of playing cards featuring so-called 'oblique strategies' (cryptic pieces of advice that need to be interpreted in ways that apply to the present situation), at moments when other creative

processes stall (Elton 1994: 213). Crucially, though, Eno's assessment phase is highly controlled. Indeed, he claims to produce a hundred times the amount of music he actually releases (*ibid.* 213).

Something like this Eno-style asymmetry between generation and evaluation also characterizes evolutionary computer art, a genre in which computational algorithms inspired by Darwinian evolution are used to produce artworks (for discussion and examples, see Boden and Edmonds 2009: 143-4). Very roughly speaking, the evolutionary computer artist proceeds as follows. First, she sets up a way of encoding artworks as genotypes. Then, she generates a random population of genotypes that are subsequently decoded to produce what is, in effect, a random population of artworks. An evaluation phase then results in some of these artworks being selected as 'parents' for the next generation. Genetic operators analogous to recombination and mutation in natural reproduction are applied to the parental genotypes to produce offspring artworks. The resulting new population is then evaluated and a new batch of parents is selected. And so on. Over successive generations, artworks are discovered that come to reflect whatever aesthetic criteria are at work (often in an implicit or intuitive way) during the evaluation phase.

Now notice that the two preceding creative scenarios, in which there is an asymmetry between generation and evaluation, are also characterized by a distinctive division of labour between the human beings and the technology involved: Eno always evaluates the products generated by his unreliable oscillators and his randomly selected oblique strategy cards; and, *typically*, the evaluative selection of parent artworks in evolutionary computer art is carried out by the artist, or by other human beings such as the visitors to a gallery. I say 'typically' because there are examples of fully automated evolutionary computer art in which the evaluation is carried out algorithmically using a 'fitness function.' However, the scope of such algorithmic aesthetic judgments may be limited. As remarked above, aesthetic evaluation involves aesthetic values and, as Boden (2004: 10) notes, these are difficult to recognize, put into words or state clearly. Moreover, they are culturally variable, contested (even within cultures) and change over time. Striking a similar note, Elton (1994, 1995) describes aesthetic evaluation in terms of a culturally conditioned, changing, basic aesthetic that can't be articulated explicitly. So, you may ask yourself, could full-blooded, human-like aesthetic evaluation ever be realized in external technology?

This question is one that Boden and Elton each raise in relation to artificial creativity, that is, in relation to the debate over whether any 'mere' machine, such as a computer, could ever be creative, or at least appear to be creative. However, it is relevant to extended creativity too, because it promises to provide part of an argument by which someone who is nervous of the metaphysical consequences of extended cognition might try to relegate, to the status of 'mere' scaffolding, even those external factors that make significant causal contributions to the generation of created products. If this conclusion holds, then of course creativity is 'at best' an embedded, rather than an extended, psychological phenomenon. Here's how such an argument might go.

Against the hypothesis of extended cognition in general, Butler (1998: 205) argues that the phenomena of control and choosing provide what he calls the 'mark of a truly cognitive system.' Furthermore, according to Butler, this sort of executive control of thought and action – the mark of the truly cognitive – happens only in brains, which should make us resistant to counting any of the external causal contributions involved as genuinely cognitive. It seems that a Butler-style argument might be constructed in the vicinity of creativity, as we are

conceiving of it at present, that is, in terms of generation and evaluation. First, recall Eno: it is at least plausible that we treat Eno himself as creative, despite the fact that he out-sources certain generative processes to external technology, precisely because he continues to be the site of the evaluative decision-making. Perhaps this suggests that evaluation is the mark of a truly creative cognitive system. Now plug in the reflections on evaluation by Boden and Elton. They provide some support for the claim that, *in the case of artistic creativity at least*, evaluation (the executive control of creation, as we might now say) cannot be realized in external technological elements. But if truly creative cognition happens just where evaluation happens, and if aesthetic evaluation is brain-bound, then maybe, in the spirit of Butler, we should be resistant to the idea of extended creative cognition, at least in the case of art. (Note: Henceforth I shall suppress the step of limiting this objection to the domain of artistic creativity. Although, given what I've said, this step is strictly correct, it strikes me as plausible that the considerations which, according to Boden and Elton, prevent the algorithmic specification of aesthetic values – e.g. resistance to explicit articulation, cultural variability – will apply to many other kinds of creation-related evaluation.)

Whatever initial attractiveness the Butler-style argument against extended creative cognition may possess, it has a seriously unpalatable consequence and so should be rejected. In shrinking the truly creative part of cognition down to just evaluation, the argument is committed to dismissing, as non-creative, any other processes that are implicated in the generation of the forms that are subsequently evaluated, whatever those processes may be (rule-governed manipulations, pattern completion, perceptual sensitivity, imagery, analogical reasoning, and so on) and wherever they are located (in the brain or in external technology). This move surely lacks independent plausibility, since where such processes occur in the brain during a creative event, we would standardly count them as part of the creative cognition that's going on. To withdraw that judgment solely on grounds of externality would be to beg the question against the extended view.[10] But without the shrinking move in place, the Butler-style critique provides no barrier to the claim that creative cognition is extended. After all, there is nothing on the table to suggest that at least some of the generative processes in the frame couldn't be externally realized within a distributed path of creation, and we have seen positive evidence that some of them (e.g. rule-governed manipulations) could be.

Notice that it would do no good for our Butler-style critic to revise her position by claiming (a) that it is evaluation *plus the choosing of the generation mechanisms* that co-constitute the executive control of creation, and (b) that evaluation and generative-mechanism-choosing are both brain-bound. Any such adjustment will succumb to a response with the same structure as the one I offered previously. For the revised objection to be an argument against extended creativity, the truly creative part of cognition must be shrunk down to just evaluation plus selection. But that means that all the other processes implicated in the generation of the forms that are subsequently evaluated will be rendered non-creative, whatever those processes may be and wherever they occur. As we have seen, this position lacks independent plausibility. Creative cognition, it seems, may sometimes be extended, and not 'merely' embedded.

Conclusion: the Path of Creation (Again)

On a track called 'Life is Long' (written, as it happens, with Brian Eno), David Byrne sings the line 'I can barely see cos my head's in the way.' That's (almost) the message of this chapter. We can barely see the cognitive mechanisms underpinning creativity because our heads are in the way. Once this cranial obstruction is fully removed, the path of creation is revealed to be routinely constituted by dynamic arrays of body-involving and environment-

involving processing loops. In other words, the creative mind is embodied, embedded and extended.

Acknowledgments

For useful comments on an earlier version of this paper, thanks to Matthew Elton, Berys Gaut and Matthew Kieran. For critical feedback on the ideas presented in talk form, thanks to audiences in Cambridge, San Antonio and Stirling. Finally, special thanks to Matthew Elton for buying me a copy of *How Music Works.*

References

Adams, F. and Aizawa, K. (2008) *The Bounds of Cognition*, Malden, MA and Oxford: Blackwell.

Beach, K. (1988) "The Role of External Mnemonic Symbols in Acquiring an Occupation," in M. M. Gruneberg and R. N. Sykes (eds.) *Practical Aspects of Memory*, New York: Wiley.

Boden, M. A. (1990) *The Creative Mind: Myths and Mechanisms* (first edition), London: Weidenfeld and Nicolson.

Boden, M. A. (2004) *The Creative Mind: Myths and Mechanisms* (second edition), London and New York: Routledge.

Boden, M. A. (2010) *Creativity and Art: Three Roads to Surprise*, Oxford: Oxford University Press.

Boden, M. A. and Edmonds, E. A. (2009) "What is Generative Art?" Digital Creativity vol. 20, nos. 1-2, pp.21-46. Reprinted as chapter 7 of (Boden 2010), from which page numbers are taken.

Boden, M. A. (this volume) "Creativity and Biology".

Butler, K. (1998) *Internal Affairs: A Critique of Externalism in the Philosophy of Mind*, Dordrecht: Kluwer.

Byrne, D. (2012) *How Music Works*, Edinburgh: Canongate.

Chomsky, N. A. (1986) *Knowledge of Language*, New York: Praeger.

Clark, A. (1997) *Being There: Putting Brain, Body, and World Together Again*, Cambridge, Mass.: MIT Press.

Clark, A. (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.

Clark, A. and Chalmers, D. (1998) "The Extended Mind," *Analysis* vol. 58, no. 1, pp.7-19

Collingwood, R. G. (1938) *The Principles of Art*. London: Oxford University Press.

Croce, B. (1909/1922) Aesthetic: *As Science of Expression and General Linguistic*, translated by D. Ainslie, New York: Noonday.

Elton, M. (1994) "Towards Artificial Creativity," *Languages of Design* vol. 2, pp.207-22.

Elton, M. (1995) "Artificial Creativity: Enculturing Computers," *Leonardo* vol. 28, no.3, pp.207-13.

Haugeland, J. (1998). "Mind Embodied and Embedded," in *Having Thought: Essays in the Metaphysics of Mind.* Cambridge, Mass.: Harvard University Press.

Kirby, S. (2002) "Learning, Bottlenecks and the Evolution of Recursive Syntax," in E. Briscoe (ed.) *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, Cambridge: Cambridge University Press.

Kirby, S., Tamariz, M., Cornish, H. and Smith, K. (2015) "Compression and Communication in the Cultural Evolution of Linguistic Structure," Cognition vol. 141, pp.87-102.

Lakoff, G and Johnson, M (1980) *Metaphors We Live By*, Chicago: University of Chicago Press.

Laland, K. N., Odling-Smee, J. and Feldman, M. W. (2000) "Niche Construction, Biological Evolution and Cultural Change," *Behavioral and Brain Sciences* vol. 23, no. 1, pp.31-146.

Menary, R. (ed.) (2010) *The Extended Mind*, Cambridge, Mass.: MIT Press,

Nelson, R. (2002) "Evolutionary Theorising in Economics," in M. Wheeler, J. Ziman and M. A. Boden (eds.), *The Evolution of Cultural Entities*, Oxford: Oxford University Press.

Noë, A (2004) *Action in Perception*, Cambridge, Mass.: MIT Press.

Smith, K., Kirby, S. and Brighton, H. (2003) "Iterated Learning: a Framework for the Emergence of Language," *Artificial Life* vol. 9, no. 4, pp.371-386.

van Fraassen, B. (1980) *The Scientific Image*, Oxford: Oxford University Press.

Ward, T. B., Smith, S.M., and Finke, R. A. (1999). "Creative Cognition," In R. J. Sternberg (ed.), *Handbook of Creativity*, Cambridge: Cambridge University Press.

Wheeler, M. (2005) *Reconstructing The Cognitive World: The Next Step*, Cambridge, Mass.: MIT Press.

Wheeler, M. (2011) "Embodied Cognition and the Extended Mind," in J. Garvey (ed.), *The Continuum Companion to Philosophy of Mind*, London: Continuum.

Wheeler, M. (2013) "Is Cognition Embedded or Extended? The Case of Gestures," in Z. Radman (ed.), *The Hand, an Organ of the Mind: What the Manual tells the Mental*, Cambridge, Mass.: MIT Press.

Wheeler, M and Clark, A. (2008) "Culture, Embodiment and Genes: Unravelling the Triple Helix," *Philosophical Transactions of the Royal Society Series B*, vol. 363, pp.3563-75.

Ziman, J. (ed.) (2000) *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press.

---

[1] For the uninitiated, Talking Heads were a New-York-based band active between 1975 and 1991. Their music might be described as located somewhere between new wave and art-pop, although that hardly begins to do justice to the eclectic palette of influences that shaped their body of work. Byrne has since pursued a solo musical career and has produced artworks in other media such as film, photography and fiction. He has collaborated with a wide range of musicians and artists, including, perhaps most notably, Brian Eno. Although, formally speaking, Byrne is not a card-carrying academic theorist of creativity, *How Music Works* rewards being read in precisely that spirit.

[2] The romantic/conventional understanding of creativity chimes with certain prominent accounts of creativity – and especially of artistic creativity – in the history of philosophy. Consider, for example, Collingwood's (1938) notion of artworks as imaginary things. According to Collingwood, in the case of art, the physically realized created product is related only contingently to the artwork proper, with the latter existing in the mind of its creator, even though the weakness of the human imagination means that sensory loops through the environment may sometimes provide assistance during the creative process. A similar theoretical structure may be found in Croce's (1909) account of artworks in terms of internal intuitions that, as a matter of only contingent fact, sometimes require physical realizations in order to be remembered or communicated.

[3] Eagle-eyed readers will be itching to point out that Byrne describes the path of creation (as he understands it) as being *almost* 180° from the romantic model, which is not quite a reversal. It is hard to know what significance one should attach to this qualification, but I am inclined to hear it as a recognition that the notion of 'creation as fit' can't be quite right, or at least that it isn't the whole story. Later in this chapter we will see that Byrne explicitly highlights an important shortcoming of his own model.

[4] For some of the seminal research, see e.g. Kirby 2002; Smith et al. 2003. For a summary of more recent work that extends the research beyond computer simulations to include laboratory experiments with real people, see Kirby et al. 2015.

[5] Kirby and colleagues appeal to cultural evolution to provide a scientifically credible, non-metaphorical account of how, over historical time, human natural languages come to have the compositional structure that they do. However, if, as I suggest in the main text, this opens the door to the proposition that the Byrne-style cultural-evolutionary model of creativity is similarly scientifically credible and non-metaphorical, then, since creativity happens not only in visual art and music, but also in language, there is another language-related upshot, namely that one would be able to generate a cultural-evolutionary account of linguistic creativity. Thanks to Berys Gaut for alerting me to this consequence.

[6] The case for embedded cognition has been made repeatedly. An influential philosophical treatment may be found in Clark 1997. For my own analysis, see e.g. Wheeler 2005, 2013.

[7] 'Wagner first conceived [*Parsifal*] in April 1857 but did not finish it until twenty-five years later. It was Wagner's last completed opera and in composing it he took advantage of the particular acoustics of his Bayreuth Festspielhaus. Parsifal was first produced at the second

Bayreuth Festival in 1882. The Bayreuth Festival maintained a monopoly on Parsifal productions until 1903, when the opera was performed at the Metropolitan Opera in New York.' https://en.wikipedia.org/wiki/Parsifal, last accessed 15/02/2017.

[8] The case for extended cognition was originally made by Andy Clark and David Chalmers 1998; see also Clark 2008. For a more recent collection that places the original Clark and Chalmers paper alongside a range of developments, criticisms and defences of the view, see Menary 2010.

[9] There are similarities between Elton's model and other accounts of creative cognition in the literature. For example, according to the Geneplore model (e.g. Ward et al. 1999), the creative psychological process may be divided into the two phases of generation and exploration, where the former corresponds closely to Elton's notion of generation and the latter overlaps with his notion of evaluation.

[10] Clark (2008: 160) pursues a similar kind of objection to Butler, as part of the general debate over extended cognition.