THE JOURNAL OF
ANALYTICAL PSYCHOLOGY

Journal of Analytical Psychology, 2010, 55, 502–521

Approaching archetypes: reconsidering innateness

Erik Goodwyn, *Minot, North Dakota, USA*

Abstract: The question of innateness has hounded Jungian psychology since Jung originally postulated the archetype as an *a priori* structure within the psyche. During his life and after his death he was continually accused of Lamarckianism and criticized for his theory that the archetypes existed as prior structures. More recently, with the advent of genetic research and the human genome project, the idea that psychological structures can be innate has come under even harsher criticism even within Jungian thought. There appears to be a growing consensus that Jung's idea of innate psychological structures was misguided, and that perhaps the archetype-as-such should be abandoned for more developmental and 'emergent' theories of the psyche. The purpose of this essay is to question this conclusion, and introduce some literature on psychological innateness that appears relevant to this discussion.

Key words: Jung, archetype, domain-specific algorithm, evolutionary psychology, mental module, nativism, innate

Introduction

Jung defined the collective unconscious as a type of innate human nature:

> In addition to our immediate personal conscious . . . there exists a second psychic system of a collective, universal and impersonal nature which is identical in all individuals and is inherited. It consists of pre-existent forms, the archetypes.
> (Jung 1959, para. 90)

Jung therefore posited a strong evolutionary basis for behaviour:

> If the unconscious is anything at all, it must consist of earlier evolutionary stages of our conscious psyche . . . however, the archaic conception [of the blank slate] holds on tenaciously: the psyche has no antecedents, is a *tabula rasa*, arises anew at birth, and is only what it imagines itself to be . . . . [However,] just as the body has an anatomical prehistory of millions of years, so also does the psychic system . . .
> (Jung 1961, p. 348)

These and many other statements make it easy to classify Jung as a 'nativist', as he clearly felt there were significant innate underpinnings to the mind that were autonomous and independent of individual experience.

0021-8774/2010/5504/502
Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

Jung vs. The blank slate and domain general learning mechanisms

Jung's statements argued against the idea that the mind consists of a few generic learning mechanisms that imprint 'experience' upon a *tabula rasa*, explaining all behaviour and subjective experience in terms of learning and culture. Instead, Jung felt that his archetypes, or instincts, were a fundamental part of human psychological existence, and he maintained this opinion up to his death:

> The psyche of the child in its preconscious state is anything but a *tabula rasa*; it is already preformed in a recognizably individual way, and is moreover equipped with all specifically human instincts, as well as with the a priori foundations of the higher functions.
>
> (Jung 1961, p. 348)

Jung held his views in the face of a great deal of opposition from the dominant behaviourist and constructionist positions of the era. At the time, behaviourist dogma asserted that the brain was a generic learning machine that operated via various simple rules of association, and had no innate predispositions. At the same time, developmental psychology was dominated by the similar constructionist theories of Piaget (1929) that proposed an infant with no innate structures that learned via domain-general learning mechanisms. Since the early 1970s, however, a large body of knowledge has formed that has challenged this 'blank slate' position (Buss 2005; Stevens 2002; see also Pinker 1997; Simpson et al 2005; Tooby & Cosmides 2005). In fact recent independent research into affective and cognitive neuroscience, cultural anthropology, evolutionary psychology, psycholinguistics, and neurobiology have essentially refuted the idea of the blank slate completely (Pinker 2002), so much that anthropologist and pioneer in the subject of human universals Donald Brown went so far as to state that 'Behaviorism and the *tabula rasa* view of the mind are dead in the water' (Brown 1991, p. 144). Constructionist viewpoints are still popular, however, and one important line of thought regarding archetypes has emerged that integrates constructionist concepts (Knox 2003). This line of reasoning argues that there are too few genes in the genome to encode any non-trivial innate content to the mind, and therefore we must rethink our concept of archetype in more developmental and 'emergent' terms (Saunders & Skar 2001). In particular Knox argues that meaningful symbolic content, even if only abstractly pre-specified, cannot be inherited.

Contrasting this position is the classical biological approach championed by Stevens (2002). Stevens' viewpoint, that the archetype-as-such is strongly associated with instinctive or 'genetic' neurobiological processes, however, has its critics within analytical psychology:

> [Stevens' viewpoint] has substantial face validity but the appeal to biology has led to certain questions about the innate, *a priori* aspects to archetypal theory . . . . This is because the findings of current neuroscience are calling into question the very thing to which Stevens is appealing—innatism . . .
>
> (Merchant 2009)

It is the purpose of the present work to question this very criticism; I intend to show that neuroscience does not necessarily call innateness into question. Furthermore I am laying the groundwork for future work showing evidence of a more classical viewpoint with respect to archetypes.

Innate content and evolutionary psychology (EP)

Evolutionary psychologists (EPists), affective neuroscientists and ethologists, from different perspectives, have demonstrated that organisms utilize many types of learning that is *domain-specific* rather than domain general—in other words, there are evolutionary reasons why we learn some things, like social exchanges and predator/prey inference more easily than logic puzzles or vector calculus. As we will see, this has a profound impact on the debate over preexisting archetypes because domain specificity must by necessity be pre-existent—in other words, though organisms certainly learn, the existence of domain specific algorithms (DSAs) requires the mind to have a preexisting programme of some kind to 'tell' the organism *what* to learn *about* (Tooby et al 2005).

EPist Steven Pinker, in his exploration of the computational model of cognition concluded that innate, domain specific mental structure (though flexible within certain constraints) is unquestionably selected for (1997, p. 177). The alternative to domain specific learning is domain general learning—in the absence of significant domain specificity, the mind would be considered a 'blank slate'—the constructionist position originated by Piaget is a prime example of this kind of learning, and hence the developmental/emergent theories of archetypes fall into this category because they assume no non-trivial domain specificity. Knox (2003, p. 9), for example, proposes that the archetypes-as-such are the developmental structures known as image schemas because neuroscience appears to argue that symbolically meaningful content cannot be inherited:

> we have to discard the view that [the archetypes-as-such] are genetically inherited and consider them to be reliably repeated early developmental achievements.

But in order to address the question of inherited symbolically meaningful content, data from these fields need to be examined more closely. Jaak Panskepp, a pioneer in affective neuroscience, for example, has pointed out that 'experience is more influential in changing the *quantitative* expression of neural systems rather than their *essential nature*' (Panksepp 1998, p. 17; emphasis added). Panksepp takes issue with EP in several ways which I review below, but he agrees with EPists that the mind was shaped to a large extent by evolution and definitely *not* a blank slate. Echoing the sentiment against domain-general mechanisms, another prominent affective neuroscientist, Joseph LeDoux, observes that 'Evolution tends to act on the individual modules [of emotional processing] and their functions rather than the brain as a whole.... Most likely attempts to find an all-purpose emotion system have failed *because such a system does*

not exist' (1996, pp. 105–06; emphasis added)[1]. Such an all-purpose emotion system that LeDoux is criticizing appears to be behind the idea that archetypes emerge from early developmental experiences, rather than as autonomous preexisting structures that *shape* early experience; Knox, for example, argues that symbolically meaningful content can only emerge from early attachments.

Further data comes from ethology. Gallistel (1995) has shown that throughout the animal kingdom, learning proceeds according to a multiplicity of specialized learning mechanisms rather than operant conditioning. Like evolutionary psychologists, affective neuroscientists, and Jung before them, ethologists feel that all higher animals have a constellation of specialized learning mechanisms. The increasing support for the existence of these 'problem-specific learning mechanisms' (Marler 1991) which are called 'domain-specific algorithms' by human EPists lead cognitive neuroscientists to conclude that:

> As we ascend into the human brain, we can see from an evolutionary perspective how humans must possess special devices for learning. Our human brains are larger because they have more devices for solving problems, and the devices are *shared by all members of the species* . . . Complex capacities like language and social behavior are not constructs that arise out of our brain simply because it is bigger than a chimpanzee's brain. *No, these capacities reflect specialized devices that natural selection built into our brains* . . .
>
> (Gazzaniga et al 2002, pp. 606–08; emphasis added)

Note that I am *not* referring to 'mental modules' as strictly defined by philosopher Jerry Fodor (1983), but something less restrictive and much better supported by the literature (see below). The above sources agree, then, that the blank slate is no longer a supportable model of the mind; rather, they feel that the mind has *some kind* of non-trivial innate structure, though they do not all agree on what it is exactly. In any case, Jung's nativist assumptions have, at first glance, not been falsified *at least on the assumption of innate content*. This is obviously a far cry from the momentous consequences of archetypal theory, but it is a crucial aspect of his formulation, particularly considering, as others have observed, that if research shows that no significant innate content is possible, it fundamentally changes the way we must view archetypes (cf. Knox 2003; Merchant 2009; Saunders & Skar 2001). What remains, then, is an appreciation of how this relates to emotional meaning.

EP and analytical psychology

Some of Jung's nativist sentiments appear to have been arrived at independently by EPists (Stevens 1995, 2002; Stevens & Price 2000; Walters 1994) and

[1] LeDoux's use of the word 'module' is misleading here—he is not referring to the strictly defined 'mental module' of Fodor (1983), but something more akin to the evolutionary psychology concepts of a 'domain specific algorithm', which I review in other sections of this paper.

contrast the emergent theory of archetypes. But in order to understand this position we should examine EP more closely. Tooby and Cosmides (2005, p. 15) outline the fundamental findings of EP. Most relevant here, they state that:

> Research in animal behavior, linguistics, and neuropsychology [has shown] that the mind is not a blank slate, passively recording the world. Organisms come 'factory equipped' with knowledge about the world, which allows them to learn some relationships easily and others only with great effort, if at all. Skinner's hypothesis—that there is one simple learning process governed by reward and punishment—was wrong.

They continue to state:

> The mind is not like a video camera, passively recording the world but imparting no content of its own. Domain-specific programs organize our experiences, create our inferences, inject certain recurrent concepts and motivations into our mental life, give us our passions, and provide cross-culturally universal frames of meaning that allow us to understand the actions and intentions of others. They invite us to think certain kinds of thoughts; they make certain ideas, feelings, and reactions seem reasonable, interesting, and memorable . . . that is, they play a crucial role in shaping human culture.
>
> (Tooby & Cosmides 2005, p. 18)

These are significant statements *against* an entirely domain general learning structure that is assumed by the emergent theory. EPist Hagan (2005, p. 146) elaborates:

> The universal architecture of the body is genetically specified . . . Because psychological adaptations such as vision are no different from other adaptations in this regard, they, too, are genetically specified human universals.

Note the term 'genetically specified' does not mean highly detailed one-to-one genetic correspondence—more on this below. Rather than assume a few generic learning mechanisms such as that proposed by behaviourist or constructionist theories, Tooby and Cosmides (2005, p. 46) argue that:

> Evolutionary psychologists expect a mind packed with domain-specific, content-rich programs specialized for solving ancestral problems . . . human cognitive architecture contains many information processing mechanisms that are domain-specific, content-dependent, and specialized for solving particular adaptive problems.

Cognitive neuroscientists agree, stating that 'human emotions are underpinned by specific but universal psychobiological mechanisms' (Stein 2006, p. 766). Furthermore,

> This fundamental point is at the heart of the evolutionary perspective, and concurs with a vast amount of neuropsychological research.
>
> (Gazzaniga et al 2002, p. 596)

Thus, despite the variations in details, many researchers agree that the mind has a significant amount of *a priori* structure that directs learning and behaviour in specific evolutionarily significant ways. They differ on exactly what this innate structure is, but not that it exists. Thus, Jung's hypothesis that the psyche contains *a priori* archetypes-as-such is perhaps not so implausible, provided we can understand how this may relate to inherited *meaning*. But before I explore this, we need to understand some of the criticisms of nativism and EP.

Critiques of nativism and EP

As mentioned before, data acquired in the last few decades in neuroscience, developmental psychology and other sciences have caused some to consider Jung's intuitions about innateness wrong despite the findings of EP and other sciences. Many of these criticisms have come from within analytical psychology itself. As mentioned, Knox (2003), building on the work of constructionist Karmiloff-Smith (1992), a vocal opponent of EP, argues that there are too few genes in the genome to encode symbolic imagery (more below on this argument). Saunders and Skar (2001) and McDowell (2001) make mathematical arguments that archetypes must be emergent structures for the same reason, though they differ in their approach to innateness, McDowell arguing for a more *a priori* mathematical principle of organization. Merchant (2009) argues that if we assume the emergent theory of archetypes, then the archetypes are likely not autonomous structures unrelated to personal experience as Jung proposed, but are rather constructed without any significant innate direction and subject only to individual experience; this leads him to conclude that perhaps the term 'archetype' is no longer useful or necessary. Therefore this question of innateness is paramount to the question of the nature of archetypes.

Panksepp, the affective neuroscientist whom I quoted previously, has criticized a 'modular' view of the mind (Panksepp & Panksepp 2000). The authors here state that despite their agreement about the significant effects of evolution on the mind, they warn against excessive 'adaptationism'; i.e., explaining every capacity and thought process as the result of some kind of human adaptation. Furthermore they warn EPists not to formulate psychological adaptations without the assistance of the large database of neuroethological data. I heartily agree with this aspect of their analysis, and feel these two fields can benefit greatly from the exchange, but this is hardly enough to throw out the insights of EP as it is developing; furthermore Panksepp (1998, p. 112) still posits a significant innate affective preexisting content to the mind:

> the human brain is a structure consisting of distinct evolutionary layers, with many more homologies existing in the lower strata of the brain than in the higher cortico-cognitive layers . . . [these lower strata are] the archetypal emotional-motivational processes that all mammals share.

This statement is highly reminiscent of Jung's speculation in MDR that the mind consists of 'earlier evolutionary stages of our conscious psyche' (p. 348),

which concerns us here; nowhere does Panksepp suggest that motivational emotional processes emerge during development—rather they are innate and even transcend species.

Elsewhere in his masterful review *Affective Neuroscience* (1999), Panksepp clearly accepts evolutionary preexisting bases for behaviour, particularly at the affective dimension. This approach contrasts the strictly cognitive way emotion has been understood in the neurosciences (LeDoux 1996) that underlies Knox's argument about inherited meaning. Panksepp worked with EPists Gardner and Wilson (2004) in his *Textbook of Biological Psychiatry*, and warns in the introduction that:

> Unfortunately, there has been a widespread tendency in biological psychiatry to neglect evolutionary and emotional systems in considering how the brain/mind is organized . . . .
>
> (p. 27)

Returning to the subject of EP, elsewhere neuroscientists Panksepp et al (2002, p. 106) sum up their position:

> One accomplishment of the EP approach is that it has energized a variety of social psychological experiments that have revealed consistent aspects of the mind . . . . Moreover, a second success of the EP approach is that most scholars now concede the existence of a core human psyche that is largely a product of biological evolution . . . . Disagreement, however, centers on the degree to which these accomplishments are related.

These authors therefore do not discount EP as invalid but rather seek a dialogue between the disciplines of EP, neuroscience and neuroethology in order to further expand EP and refine concepts, *especially with regard to affect*. Moreover, the 'disagreement' cited above is of less importance to archetypal theory than to rigorous neuroevolutionary models of the mind and the details of cause/effect, because the primary question presently is innateness, which I explore more fully below.

In 2006, Panksepp argued further that:

> Most scholars are beginning to concede the existence of a core human psyche . . . . Evolutionary psychiatrists are beginning to agree that much of human mental activity is driven by the ancient affective emotional and motivational brain systems shared with other animals.
>
> (Panksepp 2006, p. 790; see also Jones & Blackshaw 2000)

What Panksepp keeps calling the 'core human psyche' is essentially a collection of principles of emotion, perception, judgment, behaviour and learning that are ubiquitous, preexistent and innate—this formulation goes far beyond constructionist explanations about how emotional meaning originates and positions it firmly within the realm of the innate. While not specifically proposing DSAs as the EPists do, as his approach more advocates neuroethological and neurophysiological comparative studies, he is clearly on board with expressing a

significant level of innate emotional content to the mind—our present concern. Specifically, in his consideration of the neurobiology of dreams, Panksepp cites Stevens (1995) and even goes so far as to argue for 'the potential credibility of Jungian ideas regarding the archetypal images that emerge during dreaming' (1998, p. 366)—expressing a feeling that Jung's classical ideas of innateness are not so different from his own understanding of innateness.

From another venue, Jerry Fodor, in his *The Mind Doesn't Work That Way* (2001) criticizes EP on several grounds. The title of his book refers to Pinker's computational EP treatise *How the Mind Works* (1997). Here, Fodor argues that the computational theory of mind is flawed because the mind does not operate like a Turing machine. Second, he argues that his own concept of 'massive modularity' (Fodor 1983) cannot be accounted for by evolution, and third he argues that evolution is irrelevant to the study of the mind, as much as, say, botany is irrelevant to the study of astronomy. Pinker (2005) counters that none of these criticisms are valid; specifically, he argues that the computational model of the mind does not assume Turing machine equivalence, and hence that criticism is irrelevant. Second, that DSAs are not equivalent to encapsulated modules (a sentiment shared by most EPists, see Carruthers et al 2005), thus the limitations of modules are also irrelevant to EP. Finally, he argues that dismissing evolution as irrelevant to psychology is unsound because the biological evolution of the brain is much more closely related to the study of the mind than the examples Fodor uses in analogy, such as botany and astronomy. Of note, Panksepp and Panksepp (2000) also warn of the dangers of stretching the computational metaphor of the mind while de-emphasizing neuroscientific affective perspectives of the brain—on this I agree, but this point is far removed from the question of innateness, and Panksepp's position certainly would *not* agree that evolution is irrelevant to psychology.

Richardson, in his *Evolutionary Psychology as Maladaptive Psychology* (2007), further criticizes EP on philosophical grounds, arguing that since we do not have a perfect definition of a biological adaptation, we cannot apply evolutionary principles to psychology. Furthermore, he argues that in order to say something is an adaptation, we must recognize what it is an adaptation *for*. Until then, he argues, we must remain agnostic on the issue and therefore cannot make any inferences of the influence of natural selection on cognition. Language, for example, is not an adaptation until we can clearly identify what it is an adaptation *for*. But this self-imposed definition, which he criticizes EPists for not following, is not the definition used by EPists (see Cosmides & Tooby 2005, for examples), so the relevance of this criticism is questionable. In any case, arguing that some capacity is an adaptation is not the same as saying it is innate—which is the issue of concern presently with respect to archetypes.

As it turns out, EP has generated numerous predictions that have subsequently been verified over dozens of countries in the last few decades; predictions in habitat preferences, spatial location memory, disgust adaptations, predator-prey inferences, phobias, resource and mate protection and selection, sibling

rivalry, and many others (for a recent review, see Buss 2009), though the proximate neurophysiological mechanisms for these innate processes are still being investigated. What is not being debated, however, is their innateness, rather what is under question is how these capacities emerge so robustly. EP has actually gained steadily in momentum over the past few years; coverage of EP in psychology textbooks has expanded over the past two decades (Cornwell et al 2005).

Many criticisms of EP are actually criticisms of 'mental modules' (Fodor 1983). But DSAs are not modules in the Fodorian sense, which must be 'encapsulated' by definition (Fodor 1983)—that is, monomodal and inaccessible to other modules. Indeed, multimodality is the rule rather than the exception in the brain (Panksepp 1998), but falsifying pure Fodorian modules is irrelevant to the question of DSAs, as DSAs are not required to be encapsulated (among other differences, see Tooby et al 2005).

Pinker (2005, p. 4) shows how this multimodality works with respect to DSAs according to EP:

> The concept of an artifact, for example—an object fashioned by an intelligent agent to bring about a goal—combines the concept of an object from intuitive physics with the concept of a goal from intuitive psychology. The psychology of sibling relations embraces the emotion of affection (also directed toward mates and friends), an extra feeling of solidarity triggered by perceived kinship, and a version of disgust pinned to the thought of having sexual relations with the sibling . . . [sibling relations are] inexplicable by a theory of social psychology that doesn't distinguish among kinds of human relationships but appeals only to global drives like 'positive affect'.

Other examples come from EPist David Buss (2009, p. 141):

> Although most modern humans no longer struggle with dangerous species, the adaptations that evolved from those former struggles continue to be expressed in the modern world. People do not seek psychological treatment for car phobias, since they rarely occur, despite the fact that modern deaths by car accidents greatly exceed deaths by snake bite and falls from heights . . . treatments of snake and flying phobias though remain booming business today.

These phobias are difficult to account for without invoking evolution and a certain degree of innateness, regardless of whether they are pure adaptations or merely organized exaptations or spandrels.

And so the debate continues, but, regardless of the controversy, for our purposes here, it appears there is a large body of empirical research that supports the notion that there is quite a bit of innate structure to the mind that directs behaviour, affect, perception, judgment, motivation, and cognition—enough to pause before throwing out Jung's idea that archetypes could have an innate, a priori origin. Furthermore, the very existence of this structure can explain why some of Jung's proposals might *seem* Lamarckian when they are not. The question remains, then, as to what nativists mean when they say 'innate', and how might we apply the term when we consider archetypal theory.

Nativism and the term 'innate'

Unsurprisingly, there are a variety of definitions of innateness in the cognitive sciences, as well as misrepresentations (Simpson et al 2005). In 'Toward a reasonable nativism', Simpson (2005, p. 123) reviews the conceptualizations of innateness from both mainstream nativists and their critics. Most importantly, he defines three supposedly nativist concepts that draw fire from critics which are not actually representative of nativist proposals. Drawing from a leading critic, Karmiloff-Smith (1998), the charges are that anything construed by nativists to be innate must:

1. Have highly detailed and one-to-one genetic pre-specification.
2. Require minimal environmental input.
3. Be present at birth.

Karmiloff-Smith correctly observes that these characteristics are *not* plausible given the available literature. Simpson argues, however, that none of these are part of *any* mainstream nativist conceptualization, and that this is a straw-man argument. Rather, nativists simply argue that many psychological strategies, mechanisms, and capacities have been shown empirically to arise robustly despite environmental variation, and when they do change in the course of development they appear *not* to be via psychological mechanisms; meaning they probably change owing to either non-psychological environmental changes (such as brain trauma, drugs or malnutrition) or biological changes (such as normal maturation of the cortex). These capacities, therefore, are defined as '*innate*'—regardless of developmental mechanism. Thus innate structures or capacities must have some kind of genetic pre-specification, but they are not required to be 'highly detailed' or one-to-one. This is important, in that this means that 'innate' does *not* have to mean 'genetic'. Epigenicity is certainly the rule rather than the exception in development—but if a particular faculty develops reliably in everyone regardless of large variation in psychological environment, usually verified via wide cross-cultural studies, then nativists consider this as sufficient criteria to designate it as innate. This definition is therefore both intuitive and practical; moreover it accounts for the highly complex interaction between nature and nurture that is uncontroversial.

Investigators concerned with the exact source of the capacity, whether it be an adaptation, spandrel, genetic drift or whatever, therefore ask a different question than whether or not it is innate. As EPist Atran (2005) has pointed out, saying something is an *adaptation* requires stricter criteria than saying it is a *DSA*, which is itself stricter than saying it is *innate*[2]. The conflation of these terms with respect to archetypal theory adds to a great deal of confusion, as the archetype-as-such may be best classified as merely innate, and not necessarily

[2] There are further designations of adaptations as 'strong' or 'weak' adaptations; for details see cited work.

an adaptation; but even here this still means genetic pre-specification and not an entirely emergent origin of the archetype. In these cases, the genetic pre-specification is in terms of initial constraints and/or guiding principles, and *not* 'highly detailed'—nevertheless, if it turns out that everyone develops the same way along these lines, it's still innate.

In response to point 2 above, neither is environmental input discounted; rather, mainstream nativists define *innate* psychological capacities to be responsive only to *non-psychological* environmental input. Nor is presence at birth (point 3) required any more than teeth or pubic hair to be considered innate qualities of the organism. It is in this sense (i.e., the mainstream nativist definition) of the word *innate* that I will be using throughout this article.

Ultimately, the DSA hypothesizes that the brain, owing to its specific evolutionary history, acquires and codifies some information more efficiently than other information. Thus DSAs are given to account for the fact that humans learn to speak a language easily and precociously, but must expend years of effort learning how to read and write one. Other DSAs proposed that have received considerable empirical support include conspecific recognition, predator-prey inference, numerosity, mate selection, folk biology, theory of mind, mate-retention tactics, social competition tactics, and many others. These DSAs have been observed to be present in dozens of world cultures (Buss 2005; Buss 2009; Carruthers et al 2005) and thus appear highly resistant to psychological environment, which suggests they are innate.

The quintessential DSA: the cheater detection algorithm

In their comprehensive article 'Neurocognitive adaptations designed for social exchange', Cosmides and Tooby (2005) explain the DSA concept more thoroughly and give a prime example in the so-called 'cheater detection algorithm'. EPists have shown through a variety of rigorous experiments that humans appear to have an evolved 'cheater detection' algorithm designed to pick out non-reciprocating conspecifics in the environment (Cosmides & Tooby 2005). This ability is an unconscious skill in tasks that test detection of conspecifics that appear to be 'cheating' in social exchanges. The skill appears to be universal and humans show superiority in it compared to tests that are logically equivalent—this would not happen if the brain used domain general learning mechanisms alone. The tests give the same results regardless of whether it is tested in residents of the United States, United Kingdom, Germany, Italy, France, Hong Kong, Japan, Quito, Ecuador, or among Shiwiar hunter-horticulturalists in the Amazon (Cosmides & Tooby 2005, p. 596) and therefore represents a classic example of a DSA that is reliably available to children by age 3 (cf. Harris et al 2001). This ability furthermore appears to be linked to amygdala activity (Damasio 1999), in that lesions of this brain region result in overly trusting behaviour that leaves one highly vulnerable to being taken advantage of.

In their article, Cosmides and Tooby show that alternative explanations for this effect such as familiarity, social contract, permission schema, 'clear thinking', domain general logic, relevance theory, rational choice or statistical learning do not successfully account for this capacity, thereby showing excellent empirical evidence of an innate structure in the mind.

Knox's emergent formulation: comment

In the excellent *Archetype, Attachment, Analysis* (2003), Knox shows the importance of attachment theory to Jungian analysis. It is not my intent to critique this aspect of the analysis, as Knox convincingly argues that attachments are closely linked to therapeutic considerations in analytical psychology. However, Knox also argues strongly against *a priori* components to archetypes or attachments, and postulates that the infant, working with the proposed archetype-as-such known as the abstract CONTAINMENT image schema, interacts with the actual mother and, only through metonymic expansion, creates an emergent structure she (re-)defines as the mother archetypal image (and complex). This argues that all qualities of the mother image are essentially derived from the actual mother save a few abstract image schema qualities—which refutes Jung's original (1959) conception of the mother archetype-as-such as an innate, content-rich entity that organizes later experience in a highly structured way, is affectively charged, universal, less dependent upon the actual mother, resistant to experience and providing the core for the mother complex.

Furthermore, Knox (2003, p. 47) argues against the possibility of there being any innate symbolic content:

> The genetic infrastructure (30,000 genes) is too small by far to encode the infinite range of symbols that the human mind can produce . . . . Symbolic meaning cannot be inherited . . . .

This is essentially the same 'gene count' argument against innate principles as found in Saunders and Skar (2001), and others. But gene count arguments are irrelevant to debates about innateness for several reasons. First, we have very little information on precisely how genes interact with each other to produce phenotypes (Pinker 2002). Indeed, just how many genes does it take to create a complex symbolic representation? 50? 5000? The fact is we simply do not know, so we cannot assume it is impossible until we have more data.

Second, geneticists argue that 50–70% of the now estimated 20,000 genes of the genome are involved in the construction of the CNS (Keller & Miller 2006). That means that there are 'merely' 10,000 to 14,000 genes to pre-specify the brain in all its complexities. But that also means there are also 'only' up to 20,000 genes involved in pre-specifying the rest of the body in

all *its* uncontroversial innate complexity[3]. Yet, somehow no more than 20,000 genes appear to be sufficient to generate the incredibly complex structure of the various organ systems. This level of intricacy and complexity is phenomenal and universal in spite of large environmental variations; moreover it is inarguably innate. This lends credibility to the idea that at least as much innate complexity in the brain *can* exist, despite the gene count—furthermore, no matter how small the gene count estimate gets, *this same comparison will always apply*.

Third, **epigeneticity does not preclude innateness.** Most theorists seem to assume that if any aspect of any trait is acquired epigenetically it cannot be innate. Nativists only require a skill or capacity to be reliably acquired across the species in a way that is resistant to *psychological* environmental variation for it to be considered innate (Tooby et al 2005). Furthermore, some capacities, such as language acquisition[4], appear to be highly resistant to psychological variation, such that only in the most extreme cases of maltreatment will a child *not* acquire a language of some kind. Considering attachment, *secure* attachments may or may not be formed and are highly subject to psychological environment and hence not innate, but whether or not a child will attempt to form attachments at all is *not* subject to psychological variation. Simply by virtue of being a mammal, all children, barring a neurodevelopmental disorder like autism (which would be *non-psychological environmental variation* again), will attempt to form attachments with their caretakers, and this activity is likely related to highly conserved and universal oxytocin and arginine vasopressin circuitry in the deep brain (Panksepp 1998), and therefore can be considered innately pre-specified.

Finally, Knox (2003, p. 47) asks whether abstract innate imagery can be inherited:

> If an algorithm for caregiving is triggered by the sight of one's own infant, does this algorithm include information about concepts such as helplessness, dependence, attachment? If so, how have symbolic semantic concepts become stored in a package of genetic instructions?

This excellent question leads to the fourth reason gene count arguments are irrelevant. To answer the second part of it, symbolic concepts reflect the *subjective* aspects of the mind (feelings and attachments) as they correspond to the *objective* aspects of the brain (genes and neurons); in other words, this question actually impinges on nothing less than the mind-body problem itself. Indeed, how does consciousness or the mind emerge from genes and

[3] The above percentage of genes involved in the CNS development does not preclude their use in body development; most genes probably 'multitask' for the sake of efficiency. Nevertheless, that still limits the maximum number to 20,000, and the number may be less than this.

[4] Language acquisition is a good example of a capacity that is arguably not strictly an adaptation but is nevertheless innate (see Atran 2005).

neurons? How does *anything* subjective emerge from genes and neurons? Theories abound, but whatever our opinions are, *it still happens*. Thus this question is much deeper than it seems, and as such I cannot field the subject here, but will engage the issue at length in future work. Suffice it to say that having intense subjective *feelings* for one's infant cannot be translated into objective genes and neurons because of the difference between subjective and objective aspects of the mind-brain—like waves and particles, one cannot be translated into another, but in the end, they *are* the same thing; i.e., it is still a monistic phenomenon. By contrast, however, motivation and value can be examined.

However, throughout, Knox (ibid., p. 37) argues that symbolic and affectively meaningful content—meaning archetype-as-such as preexisting affective cores of meaning, value, and motivation—cannot be innate because they are features which

> Darwinian theory defines as non-inheritable because they are always formed from learnt or acquired experience.

EP directly engages this very question—what is *meaning?* Meaning is value—that is, placing some kind of value on an experience. This question therefore requires a deeper look at what evolution has given us in terms of innate value systems, if any. In Tooby, Cosmides and Barrett's (2005) article 'Resolving the debate on innate ideas' this very issue is explored in depth. The problem starts thus: if we are to assume that we only acquire value systems via image schemas or some other experientially based system, we arrive at a serious problem—as Hume (1740) argued centuries ago, we must surmount the barrier that produces an *ought* from an *is*. Furthermore:

> Hume's argument generalizes to any psychological phenomenon that requires valuation to operate. From the point of view of the valuer, value is not a physical property, or a set of patterned relationships among entities in the external world . . . . Accordingly, mental representations of the value of a behavioural outcome cannot, even in principle, be learned through the operation of any content-independent procedures, . . . That is, regardless of what environmental features they are designed to take as inputs during development, motivational machinery and the core concepts they require must be assembled by specialized developmental programs designed by natural selection for that function . . . No stimulus intrinsically mandates any response, or any value hierarchy of responses . . . the infant who is the object of caring attention by one organism is the object of predatory ambition by another, an ectoparasitic home to a third, and a barrier requiring effortful trajectory change to a fourth. It is the brains of these organisms that introduce behaviour-regulatory valuation into the causal stream.
> (Tooby, Cosmides & Barrett 2005, p. 315)

Elsewhere, they elaborate:

> The proprietary content introduced by the architecture constitutes a form of *knowledge*: the architecture must know (in some sense) that living children are better than dead children, social approval is better than disapproval, salt and sweet are better than acrid or putrefying, sex with your mother or father is to be avoided, helping

> siblings is (within certain tradeoffs) better than helping fungi, your mate copulating with your sexual rival is worse than his or her fidelity, spiders on your cheek are worse than in the garden, understanding is better than confusion, skill mastery is better than inept performance, and so on.
>
> (p. 317; emphasis added)

They agree that experience expands these basic motivational programmes and enriches them, however:

> there must be an irreducible core set of initial, evolved, architecture-derived content-specific valuation assignment procedures, or the system could not get started.... Altogether, there has not been very much progress over the last century toward constructing such an inventory, because we have been shrugging off the issue of motivational innateness through the shell game of implying that any given motivation is secondarily acquired, without obliging ourselves to computationally specify how and from what. The field needs to settle on a well-validated, irreducible set of motivational first movers. In our experience, a serious analysis of [this set] *is surprisingly large*....
>
> (pp. 317–18; emphasis added)

Finally, they conclude (p. 337) that these irreducible motivational programmes are numerous, functionally specialized, content sensitive, domain specific and *content generative*, and that this

> architecture operates jointly on values and representations of states of affairs within a given computational system so that knowledge-representing cognitive processes often cannot be intelligibly separated from motivational processes. More generally, the claim is that successful performance on value-related adaptive problems poses an insurmountable *ought from is* learnability barrier that cannot be crossed, even in principle, by content-independent learning architectures, whatever their implementation. Given data about which valuation problems humans solve, this is a method not only for demonstrating the general case for innate ideas but also for identifying specific sets of such computational elements.

Thus, according to these authors, evolutionary theory does *not* preclude the existence of innate cores of meaning but rather quite the opposite.

Conclusion: what does this have to do with archetypes?

Panksepp and Panksepp (2000) argue that many of our innate faculties may be due to interactions of the ancient highly conserved affective centres of the brain and the much newer, highly 'plastic' neocortex (reviewed in Panksepp 1998). But this does not preclude innateness, particularly when we consider that even though the neocortex is highly flexible it is still under tight control of the more ancient affective systems throughout development that are environmentally 'closed' (I will explore this in a future article). Innate processes like DSAs may be due to evolutionary adaptations, or they may be due to spandrels, exaptations, or other robust and environmentally resistant mechanisms of interaction and emergence between ancient and recent brain regions, but whatever their origin, they still exist and provide ample evidence of innateness.

However, let me point out that a DSA is *not* the same thing as an archetype, nor are irreducible cores of meaning. Rather, they are merely innate aspects of the mind that attend and process some aspect of experience, tag it with some kind of affective significance and influence learning, judgment, cognition, motivation and behaviour in some way related to the domain in question. Attachment may be a type of DSA, but it is not an archetype, rather (as I will show in future articles) the mind uses DSAs among other things to *create* archetypal symbols—it is in this way that archetypal symbols are innate and (abstractly) pre-specified. The archetypal symbol, however, is of course much more, and contains aspects of imagery, affect, meaning, and mystery.

In this article I tried to focus on only one tiny aspect of Jung's grand formulation of the archetype-as-such: that of its innateness, in an attempt to answer the question 'is it even possible to have archetypes with significant *a priori* structure?' by asking a similar (much more answerable) question: 'how much innate motivational structure does the mind have?' In so doing I reviewed literature in several fields for and against the idea that the mind has significant innate meaningful content, and found that it is not only possible that we have non-trivial innate structure, it is probable, particularly if we avoid unrealistic and artificial nature/nurture dualisms.

Moreover, there appears to be *a lot* of innate structure—far more than I have been able to review here (for full reviews, see Buss 2005; Buss 2009; Carruthers et al 2005; Tooby et al 2005). Therefore the recent emergentist trend seems overhasty, as the debate is far from over. This is because the answer to the above question is not 'hardly any' but actually 'quite a bit'. In fact, there seems to be abundant evidence that the mind is crammed with innate predispositions, perceptual biases, recognition mechanisms, emotional and expressive subroutines, behavioural urges, and more. As more cross-disciplinary work on evolutionary neuroethology continues, these will come into sharper focus. The increasing level of innate structure that is being recognized, whatever its origin may be, suggests that Jung's *a priori* archetype-as-such was perhaps not so misguided after all.

In future essays I will continue to approach the classical view of archetypes from other angles, building a model from neuroanatomy and affective neuroscience, mental imagery, dream science, and metaphor theory, and each subject will show the degree of innate mental structure from a different perspective. Finally the series will conclude with a discussion of how classical archetypal symbols can be coherently formulated firmly within the framework of the vast scientific literature that has amassed since Jung's death in 1961.

Translations of Abstract

La question de l'innéité a talonné la psychologie jungienne dès le moment où Jung postula que l'archétype était une structure *a priori* contenue dans la psyché. Tout au

long de sa vie et après sa mort, il fut constamment accusé de lamarckisme et critiqué pour sa théorie de l'existence des archétypes comme structures *a priori*. Plus récemment, avec l'avènement des recherches en génétique et le projet du génome humain, l'idée que les structures psychologiques pourraient être innées a suscité les foudres de la critique, y compris au sein même de la pensée jungienne. Il semblerait qu'il y ait un consensus croissant quant au caractère erroné de la théorie de Jung de l'existence de structures psychologiques innées, ainsi qu'une tendance générale à considérer que l'archétype-en-soi devrait être abandonné au profit de théories psychiques plus développementales et «émergentes». Le but de cet essai est d'interroger une telle conclusion et d'introduire un peu de littérature sur l'innéité psychologique afin d'alimenter la discussion.

Die Frage nach dem Angeborenen hat die jungianische Psychologie umgetrieben seit Jung den Archetyp als eine *a priori* in der Psyche vorhandene Struktur postulierte. Während seines Lebens und nach seinem Tode wurde er fortwährend des Lamarckismusses beschuldigt und für seine Theorie kritisiert, nach der die Archetypen als a-priori-Strukturen vorhanden sein sollen. Unlängst, mit dem Aufblühen der genetischen Forschung und dem Projekt der Entschlüsselung des menschlichen Genoms, erntete die Idee, daß psychische Strukturen angeboren sein könnten, selbst im jungianischen Denken noch harschere Kritik. Es scheint einen wachsenden Konsens darüber zu geben, daß Jungs Idee der angeborenen psychischen Strukturen irreführend sei und daß vielleicht der Archetyp-an-sich aufgegeben werden sollte zugunsten von entwicklungspsychologischen und 'emergenten' Theorien des Seelischen. Der Zweck dieses Essays ist es, diesen Schluß in Frage zu stellen und einige Literatur zum psychisch Angeborenen vorzustellen, die für diese Diskussion relevant zu sein scheint.

La questione dell' innato ha perseguitatola psicologia junghiana da quando originariamente Jung considerò l'archetipo come una struttura a priori all'interno della psiche. Durante la vita e anche dopo la morte venne continuamente accusato di Lamarchismo e criticato per la sua teoria degli archetipi come struttura a priori. Più recentemente, con l'avvento della ricerca genetica e del progetto del genoma umano, l'idea che strutture psicologiche possano essere innate è stata criticata ancora più duramente anche all'interno del pensiero junghiano. Sembra esserci un crescente consenso al fatto che l'idea junghiana di strutture psicologiche innate sia sviante e che forse l'archetipo in quanto tale debba essere abbandonato in favore di teorie della psiche più evolutive e 'emergenti'. Lo scopo di questo lavoro è di mettere in discussione tale conclusione, e porta della letteratura su aspetti psicologici innati rilevante per questa discussione.

Вопрос о врожденности преследует юнгианскую психологию с того момента, как Юнг постулировал существование архетипов как априорной структуры психики. На протяжении его жизни и после его смерти его постоянно обвиняли в ламаркианизме и критиковали за теорию о том, что архетипы существуют как некие предшествующие структуры. В более позднее время с появлением генетических исследований и проектов о человеческом геноме идея о том, что психологические структуры могут быть врожденными, столкнулась с еще

более острой критикой даже внутри юнгианской мысли. Похоже, растущий консенсус говорит о том, что Юнгианская идея о врожденных психологических структурах вводит в заблуждение, и что, возможно, с архетипом-как-таковым лучше расстаться в пользу больше ориентированных на развитие и «вновь возникающих» теорий психики. Цель этого эссе – усомниться в подобном умозаключении и познакомить с литературой по психологической врожденности, важной для подобной дискуссии.

La pregunta sobre lo innato ha acosado a la psicología Jungiana desde que Jung postuló originalmente el arquetipo como una estructura *a priori* dentro de la psique. Durante su vida y después su muerte fue acusado continuamente de Lamarckianismo y criticado para su teoría de que los arquetipos existieron como *a priori*. Más recientemente, con el advenimiento de investigación genética y el proyecto genoma humano, la idea que las estructuras psicológicas puedan ser innatas, ha surgido una crítica más dura aún dentro del pensamiento Jungiano. Allí parece haber un consenso creciente en relación a que la idea de Jung de estructuras psicológicas innatas fue confusa, y que quizás la idea del arquetipo *como-tal,* debe ser abandonada y dar cabida a teorías más desarrolladas y 'actualizadas' de la psique. El propósito de este ensayo es de cuestionar esta conclusión, e introducir alguna literatura sobre lo innato psicológico, la cual pereciera pertinente en esta discusión.

References

Atran, S. C. (2005). 'Strong vs. Weak adaptationism in cognition and language'. In *The Innate Mind*, eds. P. Carruthers, S. Laurence & S. Stich. Oxford: Oxford University Press.

Boyer, P. & Barrett, H.C. (2005). Domain Specificity and Intuitive Ontology. In *The Handbook of Evolutionary Psychology*, ed. D. Buss. New York/Edinburgh: John Wiley & Sons.

Buss, D.M. (2005). 'Introduction: The Emergence of Evolutionary Psychology'. In *The Handbook of Evolutionary Psychology*, ed. D. Buss. New York/Edinburgh: John Wiley & Sons.

—— (2009). 'The great struggles of life: Darwin and the emergence of evolutionary Psychology'. *American Psychologist,* 64, 2, 140–48.

Buss, D.M. (Ed.) (2005). *The Handbook of Evolutionary Psychology*. New York/Edinburgh: John Wiley & Sons.

Brown, D.E. (1991). *Human Universals*. Maidenhead, Berkshire: McGraw Hill.

Carruthers, S., Laurence, S. & Stich, S. (Eds.) (2005). *The Innate Mind*. Oxford: Oxford University Press.

Cornwell, R.E., Palmer C, Guinther P.M. & Davis, H.P. (2005). 'Introductory psychology texts as a view of sociobiology/evolutionary psychology's role in psychology'. *Evolutionary Psychology,* 3, 355–74.

Cosmides, L. & Tooby, J. (2005). 'Neurocognitive adaptations designed for social exchange'. In *The Handbook of Evolutionary Psychology,* ed. D. Buss. New York/Edinburgh: John Wiley & Sons.

Damasio, A. (1999). *The Feeling of What Happens*. San Diego, CA/New York/London: Harcourt.

Fodor J. (1983). *The Modularity of Mind*. Cambridge, MA: The MIT Press.

—— (2001). *The Mind Doesn't Work That Way*. Cambridge, MA: The MIT Press.

Gallistel, C.R. (1995). 'The replacement of general-purpose theories with adaptive specializations'. In *The Cognitive Neurosciences,* ed. M.S. Gazzaniga. Cambridge, MA: The MIT Press.
Gardner, R. & Wilson, D.R. (2004). 'Sociophysiology and evolutionary aspects of psychiatry'. In *Textbook of Biological Psychiatry,* ed. J. Panksepp. London: Wiley.
Gazzaniga, M.S., Ivry, R.B. & Mangun, G.R. (Eds.) (2002). *Cognitive Neuroscience: The Biology of the Mind, Second Edition*. London: W.W. Norton.
Hagan, E.H. (2005). 'Controversial issues in evolutionary psychology'. In *The Handbook of Evolutionary Psychology,* ed. D. Buss. New York/Edinburgh: John Wiley & Sons.
Harris, P., Núňez, M. & Brett, C. (2001). 'Let's swap: early understanding of social exchange by British and Nepali children'. *Memory and Cognition,* 29, 757–64.
Hume, D. (1740) [1978]. *A Treatise of Human Nature*. Book 3. Oxford: Oxford University Press.
Jones, I. & Blackshaw, J.K. (2000). 'An evolutionary approach to psychiatry'. *Australian and New Zealand Journal of Psychiatry,* 34, 8–13.
Jung, C. G. (1919). *The Structure and Dynamics of the Psyche*. *CW* 8.
—— (1959) [2006]. *The Archetypes and the Collective Unconscious*. *CW* 9i.
—— (1961). *Memories, Dreams, Reflections*. New York: Random House.
Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: The MIT Press.
—— (1998). 'Development itself is the key to understanding developmental disorders'. *Trends in Cognitive Sciences,* 2, 10.
Keller, M.C. & Miller, G. (2006). 'Resolving the paradox of common mental disorders'. *Behavior and Brain Sciences,* 29, 385–452.
Knox, J.M. (2003). *Archetype, Attachment, Analysis*. London: Brunner-Routledge.
LeDoux, J. (1996). *The Emotional Brain: the Mysterious Underpinnings of Emotional Life*. London: Simon and Schuster.
Marler, P. (1991). 'Song-learning behavior: the interface with neuroethology'. *Trends in Neuroscience,* 14, 199–206.
McDowell, M.J. (2001). 'Principle of organization: a dynamic-systems view of the archetype-as-such'. *Journal of Analytical Psychology,* 46, 4, 637–54.
Merchant, J. (2009). 'A reappraisal of classical archetype theory and its implications for theory and practice'. *Journal of Analytical Psychology,* 54, 3, 339–58.
Panksepp, J. (1998). *Affective Neuroscience: the Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
—— (2004). 'Biological psychiatry sketched—past, present, and future'. In *Textbook of Biological Psychiatry,* ed. J. Panksepp. London: Wiley.
—— (2006). 'Emotional endophenotypes in evolutionary psychiatry'. *Progress in Neuropsychopharmacology and Biological Psychiatry,* 30, 5, 774–84.
Panksepp, J., Moskal, J.R., Panksepp, J.B.N. & Kroes, R.A. (2002). 'Comparative approaches in evolutionary psychology: molecular neuroscience meets the mind'. *Neuroendocrinology Letters,* Special Issue, Supplement No. 4, 23, 105–15.
Panksepp, J. & Panksepp, J.B.N. (2000). 'The seven sins of evolutionary psychology'. *Evolution and Cognition,* 6, 2, 108–30.
Piaget, J. (1929). *The Child's Conception of the World*. London: Routledge and Kegan Paul.
Pinker, S. (1997). *How the Mind Works*. London: W.W. Norton.
—— (2002). *The Blank Slate: the Modern Denial of Human Nature*. New York: Viking.
Richardson, R. (2007). *Evolutionary Psychology as Maladapted Psychology*. Cambridge, MA: The MIT press.

Saunders, P. & Skar, P. (2001). 'Archetypes, complexes and self-organization'. *Journal of Analytical Psychology,* 46, 305–23.
Simpson, T. (2005). 'Toward a reasonable nativism'. In *The Innate Mind,* eds. P. Carruthers, S. Laurence & S. Stich. Oxford: Oxford University Press.
Stein, D.J. (2006). 'Evolutionary theory, psychiatry and psychopharmacology'. *Progress in Neuropsychopharmacology and Biological Psychiatry,* 30, 5, 766–73.
Stevens, A. (1995). *Private Myths: Dreams and Dreaming*. Cambridge, MA: Harvard University Press.
—— (2002). *Archetype Revisited: An Updated Natural History of the Self*. London: Routledge.
Stevens, A. & Price, J. (2000). *Evolutionary Psychiatry: A New Beginning*. London: Routledge.
Tooby, J. & Cosmides, L. (2005). 'Conceptual foundations of evolutionary psychology'. In *The Handbook of Evolutionary Psychology,* ed. D. Buss. New York/Edinburgh: John Wiley & Sons.
Tooby, J., Cosmides, L. & Barrett, H.C. (2005). 'Resolving the debate on innate ideas: learnability constraints and the evolved interpenetration of motivational and conceptual functions'. In *The Innate Mind,* eds. P. Carruthers, S. Laurence & S. Stich. Oxford/New York: Oxford University Press.
Walters, S. (1994). 'Algorithms and archetypes: evolutionary psychology and Carl Jung's theory of the collective unconscious'. *Journal of Social and Evolutionary Systems,* 17, 3, 287–306.

[Ms first received March 2009; final version December 2009]