

# KINDS OF BEHAVIOUR

Robert Aunger\* and Valerie Curtis\*

\* The Hygiene Centre  
London School of Hygiene and Tropical Medicine  
Keppel St.  
London WC1E 7HT

robert.aunger@lshtm.ac.uk (corresponding author), val.curtis@lshtm.ac.uk.

'The confusion and barrenness of psychology cannot be explained by calling it a 'young science'...in psychology there are experimental methods and conceptual confusion.' Wittgenstein, *Philosophical Investigations*.

IN PRESS, BIOLOGY AND PHILOSOPHY

## **Abstract**

Sciences able to identify appropriate analytical units for their domain, their natural kinds, have tended to be more progressive. In the biological sciences, evolutionary natural kinds are adaptations that can be identified by their common history of selection for some function. Human brains are the product of an evolutionary history of selection for component systems which produced behaviours that gave adaptive advantage to their hosts. These structures, behaviour production systems, are the natural kinds that psychology seeks. We argue these can be identified deductively by classing behaviour first according to its level of behavioural control. Early animals in our lineage used only *reactive* production, Vertebrates evolved *motivation*, and later Primates developed *executive* control. Behaviour can also be classified by the type of evolutionary benefit it bestows: it can deliver either immediate benefits (food, gametes), improvements in the individual's position with respect to the world (resource access, social status), or improvements in the ability to secure future benefits (knowledge, skill). Combining history and function implies the existence of seven types of behaviour production systems in human brains responsible for *reflexive*, *instinctual*, *exploratory*, *driven*, *emotional*, *playful* and *planned* behaviour. Discovering scientifically valid categories of behaviour can provide a fundamental taxonomy and common language for understanding, predicting and changing behaviour, and a way of discovering the organs in the brain – its natural kinds – that are responsible for behaviour.

*Key words: evolution; behaviour; emotion; brain; cognition; natural kinds*

## **Introduction**

Fruitful scientific investigation generally requires that entities, phenomena or processes be placed in meaningful classes called 'natural kinds'. Without such categorisation, empirical generalization, explanation and prediction can fail. (Allen & Bekoff, 1997; Dupré, 2000; Griffiths, 1997; LaPorte, 2004; Quine, 1969) Modern chemistry, for example, derives its explanatory power from the recognition that substances are composed of a small number of different, but unchangeable kinds of atoms. Its predecessor, alchemy, failed as a science because it assumed that a mysterious substance, the philosopher's stone, could transmute metals. As a result, alchemy could not correctly identify elements, or predict their immutability. Similarly, evolutionary biology could not explain the inheritance of information until genes, the correct units of inheritance, were described. (Wagner & Wagner, 2003) Atoms and genes are basic building blocks of their respective sciences, their natural kinds. (Wagner, 1996) Instances of a natural kind can be meaningfully grouped together because they share something fundamentally real, which can be discovered through objective investigation. (Boyd, 1991; Millikan, 1984) Science based on natural kinds thus allows us to map what matters in the world, giving us understanding and, hence, power over it. (Sterelny, 1990)

Psychology is the science of mental processes and behaviour. (Editors, 2006; Myers, 2006) However, psychology and the behavioural sciences have still not identified or agreed upon a set of component natural kinds. Scholars have argued that the brain is designed to perform functions such as: computation (Mountcastle, 1998), learning (Dayan & Abbott, 2001), the detection of discrepancies, (Barlow, 1994) reasoning (Johnson-Laird, 2006) or predicting the future (Hawkins & Blakeslee, 2004). However, from an evolutionary perspective, the primary function of the brain is to produce adaptive behaviour. (Churchland & Sejnowski, 1992; Freeman, 1999; Hebb, 1949; Newell, 1990; Skinner, 1938) Attention, learning, reasoning and prediction are of no use unless they facilitate the production of behaviours that help an animal stay alive a bit longer or reproduce itself. Animals are the only major kingdom of life to have adopted behaviour as their primary means of adaptation (Lorenz, 1965; Tinbergen, 1951) and they alone have brains. Further, those animals that live in more complex environments tend to have more complex brains to deal with them. (Godfrey-Smith, 1996) Since brains serve to produce behaviour, the natural kinds of psychology and the behavioural sciences must be the units in the brain that produce behaviour.

In biology, natural kinds such as cells or organs can be distinguished by their similar features. These features are shared because they result from a common history of natural selection for serving a particular purpose. Thus animal brains are a natural kind because they share a common phylogenetic history and a common function, that of producing behaviour. (Haslam, 2002) Similarly, neurons can be considered a natural kind because we know that they evolved to transfer information within brains. However, behaviour is controlled by structures which lie between the levels of individual neurons and of whole brains.

How, then, can we identify the component organs of the brain? To do so we have to be able to both identify its regular structures and describe their adaptive functions. Brains can only be meaningfully carved up by referring to the process that shaped behaviour, and hence brains, in the

first place -- that of evolution. Armed with such a classification, neuroscientists should be able to identify the structures responsible for producing these categories of behaviour within brains. These behaviour production units would then be the natural kinds that we seek.

We begin this paper by setting out what is meant by a natural kind in biology, and show how evolutionary kinds arise through a shared history of adaptation for function. We then set out a classification of animal behaviour by adaptive function. We propose that there are only three kinds of evolutionary function for behaviours: those that improve an organism's physiological state, those that improve its state with respect to the world, and those that allow it to improve its abilities to acquire evolutionary benefits. Evolution has invented three ways of carrying out those functions – in historical sequence, through reactive, motivated and executive behaviours. When combined, these conditions provide us with seven distinct classes of behaviour which we label *reflex*, *instinct*, *exploration*, *drive*, *emotion*, *play*, and *planning*. We have italicised our usage of these words to emphasise that we have assigned them principled new meanings, different from previous, often contested, histories of lay and scientific use. The definitions of these coinings, along with those of related terms, can be found in Table 1. We predict that these seven classes of behaviour will be produced by corresponding mechanisms, or behaviour production units, in the brain, which are its natural kinds for behaviour. In the discussion we reflect on the pitfalls of the approach we have taken and look at its practical implications both for neuroscience and for studying, predicting and changing behaviour.

Table 1 Glossary of terms

<b>CONCEPT</b>	<b>DEFINITION</b>	<b>EXAMPLES</b>
<b>Behaviour</b>	Self-propelled movement producing a functional interaction between an animal and its environment	Build shelter, Avoid predation, Hunt with others, Gossip
<b>Behaviour production unit (BPU)</b>	An evolved psychological mechanism for producing an optimized response to a cue, situation or goal	Eye-blink producer, Submission-to-social-dominant producer, Planner
<b>Evolutionary benefit</b>	A resource tightly correlated with increases in biological fitness	Food, Copulation
<b>Cue</b>	A signal of some evolutionarily salient state or variable (environmental or physiological)	Predator approaching, Low blood glucose
<b>Action</b>	Behavioural activity produced in response to a cue resulting from reactive-level control	Scratch, Suckle
<b>Reflex</b>	A class of BPUs and behaviours triggered by a cue and designed to attain a physiological end-state	Startle, Withdrawal
<b>Instinct</b>	A class of BPUs and behaviours triggered by a cue and designed to attain a situational end-state	Build nest

<b>Exploration</b>	Behaviour produced by a class of BPUs triggered by default and designed to attain an aptitudinal end-state through reactive control	Environmental reconnaissance
<b>Reward</b>	The value of feedback from behaviour (may be expressed as deviation from expectation)	Somewhat negative, Significantly positive
<b>Indicator</b>	A mental representation that indicates the degree of discrepancy between the current state of some variable (based on a cue) and its optimal state	Nutrient deficiency, Damaged social reputation
<b>Motivation</b>	A psychological state that arises when an indicator is greater than some threshold value	Low, High
<b>Need</b>	A task related to an evolutionarily significant aspect of an animal's ecological niche which requires goal-directed behaviour to solve	Invest in pair-bond, Maximize social status
<b>Episode</b>	A set of actions produced as a motivated response to a cue or situation	Display threat face → attack opponent → retreat
<b>Drive</b>	A class of BPUs and behaviours triggered by an indicator and designed to achieve a physiological end-state	Nutrient quest, Mate search
<b>Emotion</b>	A class of BPUs and behaviours triggered by an indicator and designed to achieve a situational end-state	Submission, Compete for status
<b>Play</b>	A class of BPUs and behaviours triggered by default and designed to achieve an aptitudinal end-state through motivated control	Object play, Pretend hunting
<b>Goal</b>	A mental representation of an end-state	Satiety, Copulate with mate
<b>Objective</b>	An arbitrarily distant or abstract (i.e., non-evolved) goal	Have a job
<b>Planning</b>	A BPU designed to achieve objectives through executive-level control	----
<b>Plan</b>	An action sequence produced by the planning BPU	Search want ads → fill out application → mail application, etc.

## ***Defining kinds of behaviour***

The original conception of natural kinds, due to Aristotle and pursued by the English philosophers Locke and Mill, viewed them as sets of things characterized by a necessary and sufficient suite of characteristics. These characteristics are shared by all members of the kind because they are subject to universally applicable laws. For example, water, gold and stars have the same composition, and always do the same kinds of things, wherever and whenever they are. This commonality is due to a shared essence which guarantees the identity of the natural kind and serves as its principal defining element, regardless of what those instances actually look like. (Kripke, 1972; Putnam, 1975)

This Aristotelian concept is now seen as being applicable to physical and chemical kinds, but not to biological or social kinds, which are more restricted in scope, both physically and temporally. (Boyd, 1991; Boyd, 1999b; Griffiths, 1999; Millikan, 1999) For example, any example of water will have exactly the same chemical constituents and structure (H<sub>2</sub>O), but all members of the human species do not have exactly the same composition and structure (e.g., some are male, others female). Conversely, there may be an organism on the planet Venus with exactly the same composition and structure of a human, which would not qualify as human because it does not have the same history (Ghiselin, 1974; Hull, 1978). The objects of theory in the biological and social sciences are thus historical in ways that the objects of physical and chemical sciences are not. As a result, biological and social kinds are 'fuzzier', and are subject to qualified laws. (Hacking, 2002; Millikan, 1999; Millikan, in press; Rieppel, 2005b)

The "homeostatic property cluster" (HPC) concept of natural kinds (Boyd, 1991; Boyd, 1999a) is now widely taken to be the best foundation for biological and social kinds. (Charland, 2002; Griffiths, 1999; Keller, Boyd & Wheeler, 2003; Kornblith, 2002; Millikan, 1999; Rieppel, 2005b) HPC kinds are 'homeostatic' because some force causes deviations in the qualities of members to return toward a central tendency. (Boyd, 1991; Boyd, 1999b) HPC kinds are the result of the same causal force acting on all of its members. While the Aristotelian, essentialist position assumes that the members of a kind are similar due to a shared essence or intrinsic property, HPC theory allows a broad range of properties and mechanisms jointly to constitute the kind. (Mallon, 2003) For example, species are subject to natural selection pressures from environmental factors which cause members of that species to have similar qualities (e.g., morphological regularities). A member of an HPC kind does not have to share every feature of its kind with all of its brethren; instead, individual members need only have enough of the important properties in the defining cluster to qualify as belonging. (Boyd, 1991; Boyd, 1999a) This weaker concept of membership seems appropriate for objects which do not share an essence, as in Aristotelian natural kinds, but rather similar causes.

HPC kinds for which the homeostatic mechanism is common descent constitute a class which we will call 'evolutionary kinds'. (Charland, 2002; Griffiths, 1997; Griffiths, 2004a; Rieppel, 2005a) In this case, the reason that members of the kind have a number of features in common is their shared history of subjection to natural selection pressures. Evolutionary kinds thus have law-like properties, but are limited in the range to which the kind extends, due to their historical specificity. (Boyd, 1991; Griffiths, 2004a; Rieppel, 2005b)

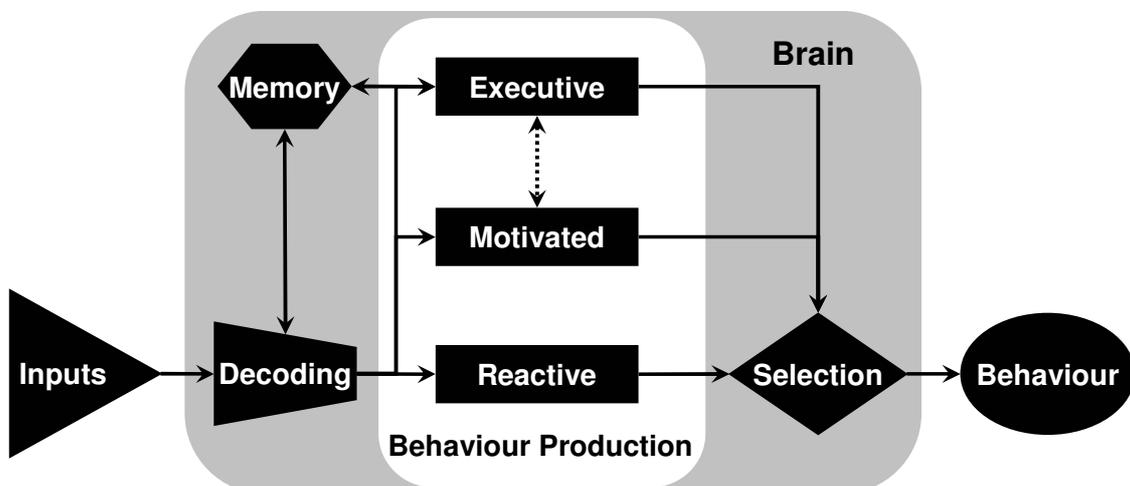
The members of evolutionary kinds produce effects which biologists recognize as phenotypes. These effects have been selected for the function that they performed in the past; those with higher fitness value tended to be retained, while those with lower value tended to disappear. Eyes are a quintessential example. (Fernald, 1997; Gehring & Ikeo, 1999) All eyes share a common function -- that of converting photons to signals that can be registered by nervous tissue. However, some structures perform this function better than others. The current types of eyes within a given lineage, such as Mammals, take the form they do as a consequence of a past history of selection for their function and are thus an evolutionary natural kind.

Natural kinds tend to appear in nested hierarchies: in chemistry, substances are made up of molecules, which are made up of atoms; similarly, in biology, organisms contain organs, which contain tissues which are made of cells. Thus functional units exist at multiple levels of organisation. (Bolker, 2000; Dawkins, 1976; Hartwell et al., 1999; Raff, 1996; Wagner, 1996) The following section proposes how a nested hierarchy of the brain-based units responsible for behaviour can be uncovered.

### ***Evolutionary kinds and behaviour***

Behaviour can be defined as self-propelled movement producing a functional interaction between an animal and its environment, such as finding a mate or fleeing from a predator. (Millikan, 2000) <sup>1</sup> Some organisms which lived in rapidly changing environments found it to their advantage to develop a mechanism which allowed their behavioural responses to be contingent on changing conditions. (Allman, 1999; Godfrey-Smith, 1996; Sterelny, 2003) Brains have been selected to produce behaviour adaptive to an animal's circumstances because mechanisms ensuring the production of such responses would have been favoured by selection. The result has been regularized structures within animals that consistently produce adaptive behaviour. We call these 'behaviour production units' (BPUs). We argue that these BPUs constitute the fundamental natural kind associated with behaviour.

**Figure 1: Behaviour Production Model**



Models of brain function in psychology and in artificial intelligence often divide the process of behaviour production into three parts: sensory systems at the 'front end' for recognizing stimuli, motor systems at the 'back end' for generating behaviour, and something in between, often called 'cognition', 'information processing', or 'central processing' (Harnish, 2001; Mesulam, 1998; Sloman, 2001). We argue that this middle box is at least partly composed of BPUs. In higher animals, BPUs are composed of sets of internal structures working in functional collaboration to produce behaviour. (Gershenson, 2001; Michaud, 2002) In higher organisms like humans, behaviour production encompasses the development of behavioural proposals by BPUs, the selection of one proposal by some mechanism, and its execution by the peripheral nervous system. We consider perceptions to be inputs to BPUs, while motor commands are their output.<sup>2</sup> Figure 1 presents a schematic picture of our behaviour production model. (We will later describe the three levels of control at which BPUs can function, represented as Reactive, Motivated and Executive.)

As with any evolutionary kind, these regular structures are produced, and can be distinguished, by their history and function. That is, members of a natural kind related to behaviour must be functionally specific, and second, they must be restricted to a class of organisms which have a shared evolutionary history (what biological systematists call a 'clade', due to being on the same branch of a phylogenetic tree (Hennig, 1966)). In what follows, we first outline the history of the development in the means through which behaviour has been produced, and then the classes of function which behaviour can accomplish, given these means. We shall see that this offers us a principled ways of carving behaviour at its joints, and hence of predicting the existence of corresponding BPUs in brains. Finally, we will observe that the types of behaviour so defined, perhaps unsurprisingly, approximately correspond to some established, if contested, psychological concepts.

### ***Types of behaviour production***

We first must restrict our examination of behavioural kinds to the history of evolutionary developments in a particular lineage. Psychologists are primarily concerned with explaining behaviour in our own species, which means we will restrict our attention to the animals of which we are the direct descendants.<sup>3</sup>

As with any historical phenomenon, the way in which things work can change over time. Just as cilia can become fins, which can evolve into legs, (Coates, Jeffrey & Ruta, 2002) so early mechanisms in brains for producing beneficial behaviour can be elaborated and improved upon in later descendants. Brains have shown remarkable increases in size and complexity in the lineage between early multi-cellular animals and our group of Primates. (Allman, 1999; Butler & Hodos, 2005; Northcutt, 2002; Streidter, 2005) What could have been the selective pressure responsible for this increase in investment in brains by later species in our lineage? With respect to behaviour, we believe that the advantage of devoting greater neural resources to the production of behaviour is that it increases the time horizon over which responses can be controlled. We therefore suggest

that there have been selective pressures on some lineages to increase the time span over which behavioural response is controlled, ranging from an instant reaction to complex and carefully planned sequences of behaviour. If an animal's time horizon is short (due to limitations of memory or information processing), then its behaviour at time  $t$  is purely responsive to stimuli at time  $t-1$ ; if its horizon is longer, then it can ignore a stimulus at time  $t-1$  to meet a need formed at time  $t-2$ , thus performing a different behaviour at time  $t$  than if it had been purely responsive. With long-term behavioural control, an organism can commit itself at an arbitrary time (i.e., time  $t-n$ ) to ignore stimuli arising during the intervening period in order to achieve a desirable end-state at time  $t$ . Animals which can forgo a present benefit can thus enjoy a higher return over the longer term by pursuing a different sequence of behaviours than would be dictated by responding to current conditions at each moment. The ability to calculate responses over longer time-frames can thus increase the adaptive value of behaviour.

It is also the case that the architecture of the human brain reflects its evolutionary history -- some control mechanisms are similar to those present in unicellular animals, some structures are similar to those of Invertebrates, and others are Mammalian in origin (Butler & Hodos, 2005; Rubenstein et al., 1999; Streidter, 1998). This gives us a reasonable expectation that there have been historical developments in the ways that animal brains work. The means for producing behaviour should therefore be cobbled together by evolution out of pre-existing parts, so that descendant species in a lineage tend to have the qualities of their ancestors, with a few changes. We thus argue that there have been three points at which the degree of control over behaviour production made significant advances in the lineage leading to humans (Table 2).

Table 2: Types of Behaviour Production Units

TYPE	LEVEL OF CONTROL OVER BEHAVIOUR	ORIGIN	TARGET OF ATTENTION	LEVEL OF MENTAL REPRESENTATION	TIME HORIZON*
<b>Simple</b>	Reactive	Metazoans (hydra)	cues	----	short
<b>Intermediate</b>	Motivated	Vertebrates (bony fish)	goals	representation	medium
		Mammals (rat)	strategic goals	meta-representation	
<b>Complex</b>	Executive	Primates (monkey)	objectives	meta-representation	long

\* Includes time before response (from time of stimulus), future time horizon over which response is calculated prior to enactment, and length of time over which behavioural response is controlled (see behavioural phenotype)

The simplest units of behaviour production are probably composed of one or more neural circuits. These units can be associated with parts of the simple nerve nets of the earliest multicellular animals, the Metazoans. The behaviour produced by these BPU's can be described as *reactive*,

because cues to evolutionarily significant situations are recognised and appropriate actions are engaged almost immediately (i.e., in response to conditions at time  $t-1$ ).

As brains become more complex, they show evidence of increasingly hierarchical structure. (Cziko, 1995; Geary & Huffman, 2002; Quartz, 2001; Swanson, 2003) The first animals to have a hierarchical organisation to their nervous systems were the Vertebrates, which have brains with multiple centralized ganglia as well as peripheral nerve circuits. (Streidter, 2005; Swanson, 2003) We postulate that units of control at this stage are composed of BPUs with intermediate degrees of complexity. These could divorce behaviour from immediate responses to cues, and hence get greater benefits (i.e., by considering conditions at  $t-2$ ). In effect, animals with this level of control can persist in a particular behaviour because they are pursuing a *goal*, or particular end-state. (Austin & Vancouver, 1996) This *motivation* provides a means of prioritisation, permitting an animal to keep to a course of action, in the face of cues to respond otherwise. (Berridge, 2004; Deci, 1975; Franken, 2001; Wong, 2000)

Motivated animals are responding to what we call an 'indicator'. *Indicators* are highly constrained, often subconscious mental representations of an abstract state which are based on inputs from the body and the environment. For example, *hunger* is a psychological state (which can be felt in humans) triggered by the level of ghrelin in the blood, (Druce, Small & Bloom, 2005) and by the presence of food and regulated by levels of gastric distension. (Gibbs, Maddison & Rolls, 1981) *Indicators* evolved to provide animals with information about the fact that their current situation deviates in some evolutionarily significant way from what is selectively advantageous. What indicators evolved to signal is a state of need – in particular, the need to solve an ecological problem. (Sterelny, 2003) *Needs* are related to some dimension of an animal's niche, such as its feeding or reproductive strategy, which presents an animal with a problem it must solve in order to survive or reproduce.

*Motivated* behaviour is typically manifested in *episodes* of goal-directed behaviour that lead to the extinction of the *indicator* through the meeting of a *need*. (Berridge, 2004; Franken, 2001; Murray, 1938) *Indicators* are constantly being calculated by the brain's attention or vigilance system, based on *cues* coming in from the environment or body. If an *indicator* exceeds its threshold value, *motivation* occurs, activating the relevant BPU to calculate behavioural options that will reduce the *indicator* below its threshold value at least cost and with highest reliability. An animal then engages in behaviour via a mechanism that selects the favoured behavioural option among the outputs of various BPUs. Stimuli representing changes in the environment and body due to the behaviour feed into the reward system, which calculates the value of that behaviour as a function of how much the *indicator* was reduced. Reduced *motivation* is positively rewarded, and motivated learning associates the behaviour with that reduction, increasing the likelihood of engaging in that behaviour the next time the *indicator* exceeds its threshold.

With Primates, we argue that a third stage of development in the complexity of BPUs was achieved. This 'executive' control is associated with a new kind of goal: 'objectives'. *Objectives* are mentally represented end-states which differ from *indicators* in being multi-dimensionally valued, highly contextualized, and long-term. They are robust mental representations, not the simple representation of internal (e.g., hormone) or environmental *cues*. *Objectives* probably arose when - other individuals had to be dealt with as agents, so that mental models of conspecifics within models of behavioural options were required. The adaptive advantage of this form of higher control

is that an *indicator* such as a hunger signal could temporarily be ignored -- for example, in the pursuit of trading for a tool that would lead to a better harvest later in the year. The type of behavioural control available to an animal with the ability to pursue objectives we call 'executive'. This allowed the *planning* of future action sequences to attain *objectives* through use of a new kind of cognitive control resident in the neocortex. (Koechlin, Ody & Kouneiher, 2003; Miller & Cohen, 2001; Wood & Grafman, 2003) Executive level control can also use motivation-level mechanisms for the hedonic valuation of imagined future states. (Gray, Braver & Raichle, 2002; Ochsner & Gross, 2005; Roseman, 1984)

The ability to disregard otherwise relevant concerns was made possible by increasingly sophisticated representations of the animal's situation in the world. The earliest animals interpreted in-coming stimuli in terms of what we narrowly define as *cues*. Thus *reactive* animals act directly on *cues* with no need for internal representations – that is, no ability to re-present sensations in an internal form for further processing. The ability to ignore nuisance stimuli in favour of particular *indicators* requires the ability to have mental representations of (some aspects of) the world. (Wood & Grafman, 2003) Thus, *motivated* animals were capable of first-order mental representations, the most important of which were *need*-states, body states, categories of objects in the world, and mental rewards.

To achieve strategic goals, Mammals evolved the ability to represent representations – that is, to 'meta-represent'. (Hughlings Jackson, 1958; Perner, 1991; Sterelny, 1998) This ability is associated with the neocortex, which first appeared in Mammals. (Allman, 1999; Northcutt & Kaas, 1995) The essence of meta-representation is the ability to hold a hierarchical structure of representations in memory and manipulate it while continuing to manage the relationships between representations, and hence maintain a consistent body of knowledge about the world. (Sperber, 2000) BPU's which involve the neocortex, and possibly other parts of the brain as well, are the most complex which have yet evolved. Animals with complex BPU's can represent their own representations to themselves. Some of them can even work towards *objectives*, rather than reacting to *cues* or *indicators*. (Rolls, 1999) Some animals (possibly only humans) can also objectivise their own mental states – that is, see themselves as having thoughts about their own thoughts. (Dennett, 1996; Proust, in press; Rosenthal, 2000) This allows animals with these special meta-representation abilities to engage in mental simulations; they have the ability to mark some mental representations as decoupled from states of the world, which allows temporary models of hypothetical situations to be created. (Cosmides & Tooby, 2000; Dienes & Perner, 1999; Sperber, 2000; Stanovich, 2004)

*Reactive* BPU's employ direct responses; unlike *motivated* behaviour, they do not refer to set-points (i.e., involve *indicators*) or require calibration against *indicators* via rewards. *Motivational* BPU's are iterative, constantly engaged in loops between the animal's current state and the previous state of the *indicator*. Motivated behaviour selection is based on how behaviours have been rewarded in similar previous situations. *Cognition*, on the other hand, is based on simulation of future outcomes. Unlike the iterative try-test-retry-exit system for goal achievement of motivational BPU's, planning can rely on a calculated sequence of options to achieve valuable *objectives*.

These different units of behaviour production have a number of features as a consequence of their different degrees of complexity. First, we suggest that this trend in increasing control has been accompanied by qualitative differences in the behavioural phenotypes produced by these types of

BPU. At first, *reactions* consisted of individual *actions* (which may be composed of multiple events), while *motivations* produce relatively short sequences of *actions* (which we will call *episodes*), composed of a number of habitual and automatic *actions*, often sequence-dependent, but controlled as a unit (such as getting dressed) (Barker & Wright, 1954; Schank & Abelson, 1977). *Executive* control produces potentially indefinite chains of *actions* that may last over a lifetime in the form of executed *plans*, which often accomplish multiple *goals* towards the achievement of an overall *objective*. (Koechlin et al., 2003; Zacks & Tversky, 2001)

It seems, therefore, that there has been a historical sequence in development of the mechanisms for the control of behaviour. Simple animals execute simple *reactive* behaviours in response to *cues*, Vertebrates evolved the ability to ignore immediate *cues* and pursue goals through *motivated* behaviour, and finally Primates added a new layer of *executive* control of behaviour which would allow them to achieve strategic end-states and *objectives* through planned sequences of behaviour.

These *reactive*, *motivated* and *planned* types of behavioural control coincide with brain structures which have long been identified by neuroscientists (as simple circuits or columns (Hebb, 1949; LeDoux, 2000; Mountcastle, 1957), modules (Damasio, 2003; Freeman, 2000; Panksepp, 1998; Tooby & Cosmides, 1992) and systems (Kelso, 1995; Thelen & Smith, 1994), respectively). All three of these structures are present in human brains. (Damasio, 2003; Rolls, 1999; Rolls, 2005) Similar levels of control have also been widely recognized in the literature on cognitive evolution. (Dennett, 1996; Ortony, Norman & Revelle, 2004; Rolls, 1999; Rolls, 2005; Sloman, 2001)

## ***Functions of behaviour***

Evolutionary kinds are also distinguished by their function. (Boyd, 1991; Millikan, 1989) Behaviour serves to put animals into states which have different types of benefits in evolutionary terms. Three categories of state can be distinguished:

- ‘physiological’ end-states, which provide changes to evolutionary benefits themselves
- ‘situational’ end-states, which produce a changed relationship with the world, such as access to territory or status, which tend to lead to future evolutionary benefits and
- ‘aptitudinal’ end-states, which produce a changed capacity to gain future situational or physiological benefits through the acquisition of knowledge or skill

The first category of ‘physiological’ end-states provides reproductive or survival benefits directly. These end-states tend to be focused on the condition of the body, because gaining immediate evolutionary benefits must involve the acquisition of some resource, or ameliorate the survival chances of the body. It is thus behaviour that is aimed at getting resources into the body, and wastes out (eating, drinking, excreting), keeping the body within a range of conditions for optimal functioning (move to suitable air, temperature, light, humidity conditions), at avoiding physiological damage (projectiles, cliff edges, parasites) or exchanging gametes (copulation).

The second way to use behaviour to get adaptive benefits is to manipulate one’s niche so as to put oneself into a *situation* where the acquisition of benefits becomes more likely in future. Effort can

be directed at improving the physical world (for example, by finding or building safe, productive habitats, or by hoarding resources), the biological world (for example by caching food or by avoiding pathogen habitats), or the social world (for example, by investing in offspring, by investing in a mate so they will help rear children, or by investing in improving social status as to get better access to resources). Biologists refer to this type of behaviour as 'niche construction'. (Odling-Smee, Laland & Feldman, 2003)

Our third category, 'aptitudinal' behaviours, is even more indirectly related to evolutionary benefits. In this case behaviour serves to improve the actor's own skills and abilities to carry out physiological or environmental tasks more effectively in future (for example, through practicing skills). The focus of aptitudinal behaviour is thus on changing the state of the brain, where memory and skill-based knowledge resides. (Deci & Ryan, 1985b; Maslow, 1943; White, 1959) <sup>4</sup>

This analysis suggests, then, that adaptive behaviour can be directed at one of three kinds of end-states; those that improve: the state of the body, the state of the world, or the state of the behavioural control system (i.e., brain) (Table 3). These three sorts of end-state define the ways in which behaviour can be functionally distinguished.

Our distinction between these three kinds of end-states is supported by evidence from psychology and neuroscience. The distinction between physiological and other end-states can be seen in brain imaging studies in humans that have shown that separate systems evaluate choices between immediate biological rewards and those delayed in time, (McClure et al., 2004) even when the delay is only a few minutes. (McClure et al., 2007) Different brain systems seem to be evolved in valuing primary versus secondary rewards; there is a clear demarcation in the brain's calculations between immediate benefits and future benefits at any remove from the present. The result of this bias, in behavioural terms, can be seen in paradoxes of inter-temporal choice: Faced with a choice between consuming \$10 today and \$11 tomorrow, some people will choose to consume the lesser quantity today. However, when these same individuals are faced with the choice between the same \$10 a year from now and \$11 a year and a day from now (the same time difference), they choose to wait and consume the greater quantity. (Frederich, Lowenstein & O'Donoghue, 2002) In effect, behaviours which immediately produce psychological rewards are hyper-stimulating.

Similarly, the division between aptitudinal and other kinds of end-states is supported by the literature on intrinsic motivation. (Deci, 1975; Ryan & Deci, 2000) Intrinsic motivation occurs when behaviour is performed for its own sake rather than to obtain material or social reinforcement (i.e., extrinsic motivation). Internally motivated behaviours tend to be felt as pleasurable, and are typically called playful, creative, or curious. (Lepper, Greene & Nisbett, 1973) Tangible rewards tend to have a substantial negative effect on intrinsic motivation – that is, providing people with money, food or other inducements to do a task which they find pleasure will tend to reduce their willingness to perform it, (Deci, 1971; Deci & Ryan, 1985a; Lepper & Henderlong, 2000) or to do so less creatively. (Amabile, 1996) This suggests that some behaviour is performed for the functional reward of skill-based learning for learning's sake – just the function we have postulated to characterize aptitudinal behaviour.

Table 3: Categories of behavioural end-state

END-STATE	DESCRIPTION	FOCUS
Physiological	Provides benefits directly (e.g., offspring, or resources for survival)	Body
Situational	Changes relationship of agent to the world such that ability to secure future evolutionary benefits is increased	World
Aptitudinal	Changes capability or capacity to carry out tasks more effectively in future	Brain

### ***Classes of behavioural kind***

We have now identified two ways of classifying behaviour – first, by the degree of sophistication in the way behaviour is controlled by brains (*reactive*, *motivated* and *executive*), and second, by the end-states which it is designed to produce (*physiological*, *situational* and *aptitudinal*). These criteria correspond to the two qualities which define an evolutionary kind: shared history and function. If we put the dimension of control together with the functional differences we have identified – that is, treat the levels of control as criteria orthogonal to the sorts of end-states achieved – the resulting two-dimensional matrix defines what we will call ‘classes’ of behaviour. Associated with these classes of behaviour, we argue, will be classes of BPU in the brain to produce them.

We present these classes as cells in a two-dimensional matrix in Table 4. Each class of behaviour can be defined succinctly by its function and level of control. We suggest a set of terms for these classes. Notice that the names we have chosen for these classes are familiar ones, such as *reflex* and *emotion*. We have opted for this strategy because behaviours that are commonly thought of as reflexive or emotional, for example, do mostly cluster within our new classification.

Because they have specific, succinct definitions, each class of behaviour can be easily described (see Tables 1 and 4). For example, *exploration* is a default behaviour (occurring when no significant threats or opportunities are detected) which can reduce inherent uncertainty through random movement that leads to the collection of information about the animal’s surroundings. A *reflex* is a reactive behaviour designed to protect an animal or its resources from threats (e.g., *startle*), or to take advantage of an opportunity which provides immediate benefits (e.g., *absorb nutrients*). *Reflexes* include the cervical contraction *reflex* in Mammals, and the suckle *reflex* in infants. Defensive *reflexes* include the annelid escape *reflex*, the emetic *reflex* in a crocodile, pupil dilation in response to light, and the human blink *reflex*.

Table 4: Classes of behaviour production units

END-STATE/ CONTROL LEVEL	PHYSIOLOGICAL	SITUATIONAL	APTITUDINAL
Reactive	reflex	instinct	exploration
Motivated	drive	emotion	play

Executive	planning
-----------	----------

*Instincts* are also reactive, but put an animal in a better position for survival or reproduction, without a directly linked physiological benefit. They can be built from combinations of *reflexes*. For example, fish can learn to associate cues of danger (e.g., conspecific injury signals) with predator cues. (Brosnan, Earley & Dugatkin, 2003) They can also learn to associate habitat dimension cues (e.g., water temperature) with presence of the predator (i.e., predator habitat recognition). Both of these mechanisms lead to flight strategies (a predator escape reflex). In combination, they produce an *instinct* to evade risky habitats, whether predators are present or not. (Kelley & Magurran, 2003)

Motivated behaviours include *drives*, which cause goal-directed sequences of behaviour which result in the direct acquisition of fitness-related benefits. <sup>5</sup> *Drives* are therefore concerned with achieving optimal levels of resources in an animal's body or within an animal's immediate control, because these actions provide immediate evolutionary benefits. The regulation of the internal milieu may involve consumption, or bringing external resources inside the body (*hunger, thirst, or acquiring male gametes*, if female); however, sometimes it involves expelling excess resources (e.g., waste products or, if male, gametes). What we call *play* is motivated behaviour with the primary function of acquiring new knowledge or skills. As with *exploration*, *play* is not a response to a particular *indicator*, but rather the lack of significant problems or opportunities; it is a default behaviour for those with an abundance of energy and no other pressing *needs* (Burghardt, 2005). *Play* not only results from the existence of surplus resources, it also creates new resources which can be used later. *Play* may be associated with a physiological or situational end-state, but its function is to enhance the organism's aptitudes, which can then be subsequently employed in the acquisition of evolutionary benefits. (Fagen, 1991)

Our definition of *emotions* as motivated behaviours which leave animals in situational end-states is different from those in the social psychological literature, where emotions have been conceived as:

- interrupting behaviour by focusing or redirecting attention (Sloman, 2001) or limiting the search for behavioural options (Evans, 2002)
- valuing events or possible behavioural options (emotion as any form of affect) (Dolan, 2002)
- finding the meaning of events (i.e., appraisal theory (Arnold, 1960; Lazarus, 1991; Scherer, 1984))
- helping people adjust to their local -- that is, cultural -- circumstances (i.e., social constructionism (Averill, 1980; Oatley, 1993))
- coordinating the physiological, expressive, subjective experiential and behavioural aspects of responses within a person (Levenson, 1999)
- interpreting feelings (Clore, 1994; Damasio, 1994; James, 1884; Prinz, 2004)
- solving recurrent problems important to biological fitness (Delancey, 2002; Ekman, 1999; Nesse, 1990; Tooby & Cosmides, 1990)

Our definition does bear a relationship to the last of these -- the evolutionary psychological definition, which is broader than ours. Nevertheless, many behaviours which would commonly be considered emotional fit within our categorization. For example, care for family members (*nurture*

and *pair-bond love*), the quest for social status (*ambition*), group membership (*belonging*), territory and money (*resource acquisition/greed*) all fit within our category of *emotions*.

*Planning*, however, is qualitatively different from these other classes of BPU. By *planning* we mean goal-directed behaviour designed to reach arbitrarily distant or abstract end-states, or 'objectives'. (Fincham et al., 2002; van den Heuvel et al., 2005) *Planning* requires the ability to generate representations of future action sequences and their likely outcomes. (Miller, Galanter & Pribram, 1960) By being able to represent goals explicitly as goals, end-states became a variable for mental manipulation, or part of the process of behaviour production itself, rather than simply the consequence of enacting a BPU. In effect, end-states were internalized, to be constructed, learned and freely chosen. (Cosmides & Tooby, 2000; Stanovich, 2004) For this reason, *planning* can be used to achieve end-states of any kind, including ones which cannot be specified before-hand. Thus *planning* can serve physiological or situational ends, but can also be used to enhance aptitudes. Indeed, *planning* can choose to pursue end-states which evolution would not select (i.e., maladaptive end-states). Thus, although the *planning* facility or BPU is an evolved structure, the outcomes it produces are not necessarily consistent with evolutionary logic.

Despite the sophisticated nature of this complex cognition in higher Primates, including humans, much of it goes on 'implicitly' in the sense of occurring below conscious awareness. (Bargh & Chartrand, 1999; Reber, 1996; Wilson, 2004) The ability to manipulate mental representations does not imply that one is consciously aware of one's own mental states. For example, Primates can engage in deception of conspecifics without necessarily congratulating themselves on their cleverness. (Whiten & Byrne, 1991). We therefore argue that there are two systems for *planning*: implicit and explicit. (Camerer, Loewenstein & Prelec, 2005; Rolls, 2005) Implicit cognition is inaccessible to consciousness due to its sub-symbolic, distributed representation in the brain, while explicit knowledge is captured (in computational models at least) by symbolic, localized representations, in which each processing unit is more easily interpretable and has a clearer conceptual meaning (Clark & Karmiloff-Smith, 1993; Sun, Slusarz & Terry, 2005) Different regions of the brain appear to be involved in the two different kinds of systems. (Keele et al., 2003; Posner, DiGirolamo & Fernandez-Duque, 1997) Brain imaging suggests the implicit evaluation system is primarily located in the orbitofrontal and cingulate cortex, while the explicit system is located in the lateral prefrontal and parietal cortex. (Rolls, 2005)

Formulating our classes of behaviour in this simple way allows us to identify relationships between them. For example, *instincts* and *emotions* share the characteristic of leaving animals in situational end-states, but differ from each other in their degree of control. We believe this is both a powerful and parsimonious way of classifying behaviour that should have considerable utility in reducing the confusion that surrounds claims about the psychological qualities of animal behaviour (e.g., whether seed-caching by birds is a form of planned behaviour (Raby et al., 2007; Suddendorf & Busby, 2003)).

### ***Individual kinds***

Thus far we have deduced how history and function, the two necessary characteristics of evolutionary kinds, should shape the structures that produce behaviour. However, the types and

classes of BPUs we have identified thus far are still unlikely to be the proximal cause of behaviour in a given situation. This is because animals recognize themselves to be in a particular context which requires a particular kind of response. The ability to recognize the nature of different situations and to select appropriate strategies must be the job of particular evolved mental algorithms. For example, within the class of situations that require a motivated response to change one's relationship to the world (i.e., an emotional situation), there is still a considerable difference in the kinds of computation, and hence behaviour, needed to repair one's social status and to defend one's territory. Hence, we expect that natural selection will have devised BPUs able to respond appropriately to situations of evolutionary significance which have been recurrent in an animal's phylogenetic history using algorithms specific to particular strategies. (Tooby & Cosmides, 1992) These mechanisms for generating behaviour must be present in all the members of the species which have faced similar ecological problems throughout their phylogenetic history. These mechanisms must be BPUs (i.e., natural kinds), because one can extrapolate from their presence in one member of a species to another. (Boyd, 1991; Millikan, in press)

Within the classes of BPUs we have discussed must therefore be 'individual' kinds designed to address particular evolutionary tasks set animals by their ecological niche. For example, *startle* is a *reflex* that helps people quickly redirect their attention in an appropriate direction. People also tend to recoil from faeces, which is an *instinct* to avoid the environmental sources of infectious disease. (Curtis, Aunger & Rabie, 2004) *Defending offspring* is an *emotion* because it conserves the situation in which kin can serve as instruments for further gene reproduction. *Pretend aggression* is a form of *play* because it is primarily concerned with acquiring skills at winning contested resources or competing for social status. (Panksepp, 1998)

The number of individual kinds which can be defined in this way will be large, but not very large, because the number of important evolutionary tasks is restricted to the dimensions of an animal's niche which distinguish it from that of other species in the same ecology. (Hutchinson, 1944) By the same token, individual kinds will vary from species to species, as the niches of different species cannot overlap completely. (Mayr, 1963)

A major exception to the argument that individual-level production mechanisms are BPUs is *executive* behaviours. Although the brain-based systems which implement such behaviours are themselves evolved, we do not believe that there are units within the neocortex dedicated to the achievement of particular *objectives*. The *plans* which executive kinds produce are therefore not evolved but constructed individually to solve perceived problems on the fly. Thus, *planning* is both a class and an individual BPU: there is only one BPU capable of meta-representation and the pursuit of *objectives*, most likely in the isomorphic prefrontal cortex.

## ***Discussion***

In this paper, we have used deduction from the necessary characteristics of evolutionary kinds to define natural kinds for producing behaviour at three levels. Our process of theory building by deduction (albeit inspired by current empirical findings) follows a common pattern in science. For example, the structure of atoms was predicted by Bohr before it was ever observed (Bohr, 1913); the structure of genes was deduced from patterns observed in X-ray crystallographic plates

(Watson & Crick, 1953); and the means by which genes express themselves was inferred from modelling prior to observation (Gamow, 1954). It is therefore reasonable to argue that deduction should also play a major role in the identification of natural psychological kinds.

What we call ‘types’ of BPUs are defined with respect to the complexity of their physical instantiation (i.e., implemented either as simple *reactive* BPUs, *motivational* BPUs of intermediate complexity, or truly complex *executive* BPUs). It is also possible to define ‘classes’ of BPUs within these types as a function of which kind of end-state such kinds produce (i.e., those providing *physiological*, *situational* and *aptitudinal* benefits). Finally, within classes are ‘individual’ BPUs which provide solutions to important evolutionary problems that living in particular kinds of ecological niches has required animals to solve (e.g., *hunger*, *thirst*, *get mate*). If statistical regularities in the world endure, they allow consistent selection pressures to operate over extended periods of time, and thus produce evolved structures within organisms – our natural kinds.

Our *types* of BPUs (e.g., *reactive*, *motivated*) are physically similar (e.g., of simple or intermediate complexity, respectively), although they may not overlap materially – they can be implemented by different parts of the brain. In this, they resemble *substances*, a chemical natural kind, which are composed of molecules, but not necessarily the same kinds of molecules (e.g., gold versus sodium chloride). Our *classes* of BPUs (e.g., *reflex*, *emotion*) are linked by function and history, but not necessarily structurally. This makes them similar to the biological kind of *organs* (e.g., heart, lung), which share functional characteristics, but are morphologically and physically distinct. *Individual* BPUs, our fundamental natural kind, are the equivalent of atoms or cells in these other sciences: they are the basic physical building blocks of psychology, with an isolatable physical presence in the brain; they are the way in which an individual brain physically produces an *action*, *episode* or *plan* that produces some kind of evolutionary benefit. These parallels with the kinds of other sciences are reassuring because they suggest we may have tapped into significant properties of the multiple levels of natural kinds more generally.

We are not the first to hypothesize the existence of psychological kinds. Debates concerning potential natural kinds in psychology have centred around cognition (Pylyshyn, 1984), consciousness (Hardcastle, 1995), knowledge (Kornblith, 2002), concept (Machery, 2005), psychiatric disorders (Zacher, 2000) particular emotions or the category of emotion (Barrett, 2006; Charland, 2002; Griffiths, 1997; Griffiths, 2004a; Griffiths, 2004b), and human kinds (Cooper, 2004; Hacking, 2002; Haslam, 1998). None of these proposals (except that for emotion) have met with much acceptance, probably because no clear way has been found to distinguish different members of the kind within these basic categories.

We are also not the first to postulate the existence of natural kinds related to behaviour. One influential predecessor in this endeavour has been ethology. In the 20<sup>th</sup> century, this major effort was devoted to the discovery of animal behavioural kinds with methods adopted from the natural sciences. (Lorenz, 1950) However, ethology produced no widely accepted taxonomy of behavioural kinds, even though numerous attempts have been made to construct schemes applicable to any species – that is, a ‘standardized’ description or ‘ontology’ of animal behaviour. (Catton et al., 2003; Golani, 1976; Schleidt et al., 1984); the Animal Behavior Ontology Project [<http://mlsource.ornith.cornell.edu/ethodata/>]. We hope that the scheme provided here can provide new direction to this project.

Our theory can be seen as a modification of expected utility theory in economics (Savage, 1954; von Neumann & Morgenstern, 1944) or expectancy value models in psychology (Bandura, 1986; Fishbein & Ajzen, 1975), where the value of a behaviour is determined by the odds the behaviour will put an animal in a particular state times the value it attaches to being in that state. In our case, value is measured not in terms of utility or perceived costs and benefits, but in terms of *evolutionary* benefits. We argue that behaviour is valued by the probability it will put the animal in a particular kind of end-state, influenced by the *evolutionary* value of that end-state (i.e., in terms of its likely contribution to biological fitness). Our primary hypothesis is that we can categorize end-states by the nature of their relationship to evolutionary benefits: direct provision, indirect provision or even more indirect provision.

Of course behaving animals don't explicitly calculate and compare expected evolutionary benefits before acting. (Utility is a similar convenient fiction, and psychologists often assume decision-making is only implicit as well.) The proximate measure of evolutionary benefits used by brains is psychological reward: natural selection has tuned brains to like feedback from adaptive behaviour and to dislike maladaptive behaviour. In this way, mechanisms that produce adaptive behaviour are reinforced, so that biological fitness tends to increase via behaviour.

### ***Practical Implications***

If we are right, and the BPUs we have deduced really serve as the fundamental building blocks of psychology and behavioural science, then identifying behavioural production mechanisms should give scientists a number of additional tools for empirical and theoretical work. We see the perspective we have developed as having practical implications in four primary spheres: for the neurosciences, behavioural sciences, and evolutionary biology, and with respect to interdisciplinary collaboration.

The first domain of implication concerns the brain sciences. The project to investigate the structure and function of the brain through neuroimaging badly needs a theoretical approach to behaviour. At present neuroscientists use common sense and intuition to determine the stimuli for their explorations of brain function using brain scanning technologies. These stimuli should rather be chosen from principle. For example, to determine how disgust works in the brain, scientists should first ascertain the purpose of the disgust BPU. Research suggests that disgust is a motivation to avoid the substances and situations that would have caused disease in our ancestors. (Curtis et al., 2004) Second, a set of stimuli should be found which, as far as is possible, mimic ancestral disease threats. Third, subjects should be exposed to these stimuli in fMRI tests, with the results being used to map the brain centres involved in disgust responses. The lack of principled sets of stimuli, due to a lack of a proper definition of the relevant domain of behaviour, is likely to be one of the reasons for continued controversy as to which parts of the brain are responsible for what behaviour.

Further, we can predict just how particular BPUs are implemented in the brain based on our argument about the order in which different classes of BPU originated. For example, being among the oldest BPUs, gene-based developmental programmes will have had time to evolve in association with particular *reflexes*, guaranteeing that they regularly appear as similar circuits in

different brains (to be found in older parts of the human brain). In particular, as short, dedicated 'arcs' of neurons, it is likely that there is little physical overlap in the circuitry used to implement different *reflexive* BPUs. Thus, the members of a particular kind of behaviour such as the *startle reflex* should reappear in similar fashion in different animals. Different *reflexive* BPUs (e.g., *startle* and *withdrawal*) will also share a similar, evolved structure – the reflex arc. So we can identify both a common history and form for the class of *reflexive* BPUs.

On the other hand, the *planning* BPU evolved relatively recently, and in humans is associated with the prefrontal cortex. (Adolphs, 2001; Christoff & Gabrieli, 2000; Miller, Freedman & Wallis, 2002) It is likely that the prefrontal cortex implements this *executive* BPU using information in distributed networks spread widely through the prefrontal cortex, while involving older parts of the brain as well. Therefore, the *executive* BPU may not be implemented in exactly the same way in the brains of different individuals. The BPU for *planning* probably has a low level of genetic specificity, and is implemented using functionally and morphologically similar tissue, the isocortex. (Fuster, 2003; Mountcastle, 1998) So there is a lower likelihood that *planning* will be associated with developmental programmes or even structurally distinct ones. (It is this generality of *executive* function, and the lack of evolved end-states toward which they are directed, which enables *planning* to address novel objectives without regard to specific goals.) We therefore expect the *executive* BPU to be less 'locatable' and display more connectivity to memory and motivational BPUs than *reflexes* (e.g., using brain imaging).

In between are the *motivated* BPUs. Although these are most likely modular, and so should have some aspects which are dedicated to specific purposes, they are also likely to overlap in their physical instantiation, with individual *drives* or *emotions* making use of components which serve as *reactive* circuits under other circumstances. Since evolution tends to tinker with pre-existing forms, they should invoke use of components that evolved earlier, and which can be shared between BPUs. Determining whether these deductions about the physical instantiation of different types of BPU are correct constitutes a rich and significant programme of research.

The second class of implications of our classification of behaviour concerns the study of behaviour itself. Behavioural scientists (e.g., ecologists or ethologists) want to know the adaptive function of particular behaviours; health psychologists and marketers want to better understand behaviour so as to change it. However, we suspect that it will never be possible to determine the level of control over individual behaviours. Observation of a single instance of any behaviour can be interpreted as the result of the lowest level of control, as a behaviour built into an animal's repertoire by genetic evolution. Even complex sequences can be found to be instinctive rituals if they are repeated on numerous similar occasions and are species-typical.

The behavioural scientist's job, then, is to look for statistical regularities in behaviour with respect to end-states. The stream of behaviour can be split into periods between physiological end-states which provide evolutionary benefits (e.g., from copulation to sleep to eating to predator defence to status improvement). Typically, the behavioural scientist will be looking for variation in the stream of behaviour from one kind of end-state to another. Only through comparison with the behaviour of the same animal or other animals in the species, resulting in the same category of beneficial end-state, can it be inferred that behaviour has likely been *reactive*, *motivated* or *planned*. A good indicator of the *motivated* nature of a stream of behaviour is that it exhibits a point at which some evolutionary benefit within immediate reach has been ignored; an indicator of planned behaviour,

that some end-state achievable through motivated action has been ignored. Multiple routes which reach the same category of end-state (e.g., copulation, fruit consumption, territory acquisition) will be exhibited if behaviour is *motivated*. *Plans* will exhibit patterns in the particular sequences in which beneficial end-states are reached. For example, *plans* that require high levels of skill to achieve will tend to require that multiple episodes of *playful* behaviour occur before the objective is achieved. *Plans* thus exhibit sequence dependence in motivated behaviours.

From the perspective of our own profession, health promotion, knowing which level of control is responsible for a behaviour is fundamental (and the reason that we undertook this project).. Recognizing that changing a behaviour will require overcoming motivation rather than reactive control, for example, increases the tool kit available for changing that behaviour. Motivated behaviour can be changed by providing incentives, knowledge, or physiological sensations, whereas changing reactive behaviour can only be achieved through the provision of evolved cues. Knowing which kind of end-state a behaviour results in can also have implications for the means available to change it. Behaviours providing physiological benefits, even when unhealthy (e.g., drug-taking, sugar consumption), are more difficult to overcome than situational behaviours such as not wearing seat belts, because immediate rewards have to be fought against.

Third, for evolutionary biologists, it should be possible to identify the behaviour kinds we have established in the brains of related species in the human lineage. For example, both rat and human brains should exhibit mental facilities to implement *drives* and *emotions* which are homologous structures. This project is already underway, with a number of such homologies having been discovered. (Panksepp, 1998; Streidter, 1998)

An empirical deduction from our approach is that animals in any given clade will only have certain types of production unit available for their behavioural repertoire. Thus, Chordates prior to Vertebrates will be 'instinctual creatures' (i.e., limited to *instincts*, *reflexes* and *exploration*), while Mammals will also display *drives*, *emotive* and *play* behaviour, and *planning* will be limited to Primates. These represent a few examples of the significant potential advances in the study of behaviour which we hope will materialize through future studies of the natural kinds we have introduced here.

We have also made specific claims about the order in which the different BPUs have evolved. Primitive functions tend to arise earlier in development and to be more similar among species in a clade (Kirschner & Gerhart, 2006; West-Eberhard, 2003). Hence among animals in the human lineage, later-evolved BPUs should develop later in brain ontogeny, and diverge more greatly between species in the human clade. We thus expect most animals will not exhibit higher cognitive kinds, causing them to diverge from Primates, and that these cognitive kinds will only develop relatively late in human ontogeny (i.e., with the frontal cortex, which develops last (Fuster, 2003)), while earlier-evolved BPUs, such as *instincts*, will be found in earlier-developing parts of the human brain.

Finally, we believe our approach has implications for interdisciplinary communication and collaboration. We have argued that psychology cannot advance as a science without terms that can attract agreement within and between the behavioural sciences. Our argument is that the best way to do this is to equate psychological constructs to biological natural kinds. Such kinds can form the foundation for interdisciplinary cooperation on the study of behaviour. It is unfortunate, for

example, that a term as basic to brain and behaviour as 'emotion' still has no agreed-upon definition. Our approach provides a short crisp definition, not from introspection or empirical observation, but from the perspective of function and evolutionary history. Emotion is a class of BPUs in the brain that motivate behaviour designed to improve an animal's situation with respect to its social world.

We can also compare predictions made from our approach about the nature of individual BPUs with those derived from a major alternative: folk psychology. Although we use non-standard means to classify behaviour into kinds, our typology does approximately map onto intuitive or 'folk' categorizations of behaviour in many cases. Hence, *hunger* is a *drive*, not because it causes a particular set of physiological states or feelings, but because it leads to evolutionarily significant returns through flexibly controlled, motivated consumption. Similarly, the *emotion* of *pair-bond love* leads an animal to work; it risks physical harm to defend rights of access to, or control of, a mate (a situational end-state). Loving behaviours can take many forms, and may endure until a particular strategic relationship is achieved. These two conditions define *pair-bond love* as an *emotion*.

However, a counter-intuitive deduction from our approach is that behaviours commonly called 'disgust' and 'fear' should often be considered *drives* rather than *emotions*. Fleeing from predation is an effort to avoid one's bodily resources becoming the predator's resources. Similarly, behaviours such as walking around excreta or rotten meat, or shunning someone who appears to be ill, involve a 'negative appetite' for avoiding resources crossing the body boundary (in this case wanting to avoid being eaten from inside by pathogens). Our presumption is that these behaviours don't involve the same complex meta-representation of prey or parasite as social interactions, and so are not *emotional* in nature. *Emotional fear* we reserve for behaviours which avoid conspecifics as threats to body or resources; *emotional disgust* is restricted to those motivated and contingent (i.e., strategic) behaviours which serve to ostracize, or shun, social 'parasites' (Curtis & Biran, 2001) Thus the overlap between our categories of behaviour and those given common-sensical names is not perfect. Such differences can be used to make unique predictions that can support or refute our approach.

## **Conclusion**

There is hardly a more powerful tool in science than what philosophers call 'natural kinds'. (Wagner & Wagner, 2003) Natural kinds are forms 'given by nature', not categories artificially constructed by the human mind (Boyd, 1991; Millikan, 1984). They are structures generated by processes that have distinctive intrinsic natures described by the causal factors at work in their production and maintenance. In this paper we have argued that natural kinds for psychology can be found by dividing behaviour into categories based on the kinds of behaviour production units which caused them. In particular, we have made use of a three-part division of the degree of psychological control over behaviour execution and an original distinction we have made between kinds of evolutionary end-states into which animals are put by this behaviour, to identify these kinds. We argue that the selective forces producing new levels of sophistication in the production of behaviour were the ability to control behaviour over longer periods, thus obtaining higher average benefits, and given a level of control, by the ability to devise more and more indirect means of acquiring evolutionary benefits by reaching different kinds of end-states. While our account necessarily

remains only plausible at present, we believe that it is consistent with a wide variety of evidence. We therefore believe this story lends support to our contention that natural kinds in animal behaviour exist, and have evolved in an understandable, orderly sequence.

We believe our approach has three major advantages over previous efforts to define classes of behaviour. First, it is uniquely constrained by both functional and historical considerations to tell a plausible story about how the various kinds of production first evolved, from one level to the next, with support from known developments in nervous systems, learning and animal task repertoires. By basing our approach in evolutionary theory, with particular reference to major transitions in complexity, we can make explicit claims about the order, timing, and means by which each kind of cause of behaviour arose. Second, our account combines production types and output types (end-states) into a comprehensive and interdependent model of the causes of behaviour; previous models have been limited to one or the other of these two dimensions. Third, we have provided new criteria for rigorously distinguishing among the various kinds of behaviour production systems. These criteria result in new claims about the functions of *instincts* and *emotions*, for example.

By redefining *reflexes*, *instincts*, *exploration*, *drives*, *emotions*, *play* and *plans* according to their evolved functions rather than by their mechanics or by the subjective states they create, we hope to provide a principled vocabulary that can be shared by behavioural scientists in any discipline, and applied to any species in our lineage. Most of these terms have long histories in psychology, of course, and have been abandoned by many precisely because of their imprecision. However, we believe that these concepts will continue to have value once couched in a natural kinds framework.

## ***Acknowledgements***

Thanks to Adam Biran, Kalina Christoff, Barbara Findlay, Carlos Gershenson, Ara Norenzayan, John Odling-Smee, Miguel Rubio-Godoy, Thomas Reydon, Beth Scott, Szymon Wichary and several anonymous reviewers for comments on earlier versions.

## References

- Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology* **11**, 231-39.
- Allen, C. & Bekoff, M. (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press, Cambridge, MA.
- Allman, J. M. (1999). *Evolving Brains*. Scientific American Library, New York.
- Amabile, T. M. (1996). *Creativity in Context*. Westview Press, Boulder, Co.
- Arnold, M. B. (1960). *Emotion and Personality*. Columbia University Press, New York.
- Austin, J. T. & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin* **120**, 338-375.
- Averill, J. R. (1980). A constructionist view of emotion. In *Emotion: Theory, Research, and Experience* (ed. R. Plutchik and H. Kellerman), pp. 305-339. Academic Press, New York.
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall, Englewood Cliffs, NJ.
- Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist* **54**, 462-479.
- Barker, R. G. & Wright, H. F. (1954). *Midwest and its Children: The Psychological Ecology of an American Town*. Row, Peterson and Company, Evanston, Ill.
- Barlow, H. B. (1994). What is the computational goal of the neocortex? In *Large-scale Neuronal Theories of the Brain* (ed. C. Koch and J. L. Davis), pp. 1-22. MIT Press, Boston, MA.
- Barrett, L. F. (2006). Emotions as natural kinds? *Perspectives on Psychological Science* **1**, 28-58.
- Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology and Behavior* **81**, 179-209.
- Bohr, N. (1913). On the constitution of atoms and molecules. *Philosophical Magazine* **26**, 1-25.
- Bolker, J. A. (2000). Modularity in development and why it matters to evo-devo. *American Zoologist* **40**, 770-776.
- Boyd, R. N. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* **61**, 127-148.
- Boyd, R. N. (1999a). Kinds, complexity and multiple realization. *Philosophical Studies* **95**, 67-98.
- Boyd, R. N. (1999b). Homeostasis, species, and higher taxa. In *Species: New Interdisciplinary Essays* (ed. R. A. Wilson), pp. 141-185. MIT Press, Cambridge, MA.
- Brosnan, S., Earley, R. & Dugatkin, L. A. (2003). Observational learning and predator inspection in guppies. *Ethology* **109**, 823-833.
- Burghardt, G. M. (2005). *The Genesis of Animal Play*. MIT Press, Cambridge, MA.
- Butler, A. B. & Hodos, W. (2005). *Comparative Vertebrate Neuroanatomy: Evolution and Adaptation*. 2nd ed. Wiley-Liss, New York.
- Camerer, C., Loewenstein, G. & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* **43**, 9-64.
- Catton, C., Wilson, C., Dalton, R. & Shotton, D. M. (2003). SABO: a Standard Animal Behaviour Ontology (alpha version 0.1). vol. 2006.
- Charland, L. C. (2002). The natural kind status of emotion. *British Journal for the Philosophy of Science* **53**, 511-537.

- Christoff, K. & Gabrieli, J. D. E. (2000). The Frontopolar Cortex and Human Cognition: Evidence for a Rostrocaudal Hierarchical Organization within the Human Prefrontal Cortex. *Psychobiology* **28**, 168-186.
- Churchland, P. & Sejnowski, T. J. (1992). *The Computational Brain*. MIT Press, Boston, MA.
- Clark, A. & Karmiloff-Smith, A. (1993). The Cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language* **8**, 487-519.
- Clore, G. (1994). Why emotions are felt. In *The Nature of Emotion: Fundamental Questions* (ed. P. Ekman and R. J. Davidson), pp. 103-111. Cambridge University Press, New York.
- Coates, M., Jeffrey, J. & Ruta, M. (2002). Fins to limbs: what the fossils say. *Evolution and Development* **4**, 390-401.
- Cooper, R. (2004). Why Hacking is wrong about human kinds. *British Journal for the Philosophy of Science* **55**, 73-85.
- Cosmides, L. & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In *Metarepresentation: A Multidisciplinary Perspective* (ed. D. Sperber). Oxford University Press, New York.
- Curtis, V., Anger, R. & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings: Biological Sciences (Biological Letters)* **277**, 131-133.
- Curtis, V. & Biran, A. (2001). Dirt, disgust and disease. *Perspectives in Biology and Medicine* **44**, 17-31.
- Cziko, G. (1995). *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*. MIT Press, Cambridge, MA.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. G. P. Putnam's Sons, New York.
- Damasio, A. R. (2003). *Looking for Spinoza*. Harcourt, New York.
- Dawkins, R. (1976). Hierarchical organisation: a candidate principle for ethology. In *Growing Points In Ethology* (ed. P. Bateson and R. Hinde), pp. 7-54. Cambridge University Press, Cambridge.
- Dayan, P. & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Boston, MA.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* **18**, 105-115.
- Deci, E. L. (1975). *Intrinsic Motivation*. Plenum Publishing Co., New York.
- Deci, E. L. & Ryan, R. M. (1985a). *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum, New York.
- Deci, E. L. & Ryan, R. M. (1985b). *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press, New York.
- Delancey, C. (2002). *Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence*. Oxford University Press, New York.
- Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. Basic Books, New York.
- Dienes, Z. & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences* **22**, 735-808.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* **298**, 1191-1194.
- Druce, M. R., Small, C. J. & Bloom, S. R. (2005). Minireview: Gut peptides regulating satiety. *Endocrinology* **145**, 2660-2665.
- Dupré, J. A. (2000). Natural kinds. In *A Companion to the Philosophy of Science* (ed. W. H. Newton-Smith), pp. 311-319. Blackwell, Oxford.

- Editors. (2006). *The American Heritage Dictionary of the English Language, Fourth Edition*. Houghton Mifflin, New York.
- Ekman, P. (1999). Basic emotion. In *Handbook of Cognition and Emotion* (ed. T. Dalgleish and M. Power), pp. 45-60. John Wiley & Sons, Ltd., Sussex, U.K.
- Evans, D. (2002). The search hypothesis of emotion. *British Journal for the Philosophy of Science* **53**, 497-509.
- Fagen, R. (1991). *Animal Play Behavior*. Oxford University Press, Oxford.
- Fernald, R. D. (1997). The evolution of eyes. *Brain Behavior and Evolution* **50**, 253-259.
- Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A. & Anderson, J. R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proc. National Academy of Sciences* **99**, 3346-3351.
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research*. Addison-Wesley, Reading, MA.
- Franken, R. (2001). *Human Motivation; 5th ed.* Brooks/Cole, Pacific Grove, CA.
- Frederich, S., Lowenstein, G. & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature* **40**, 351-401.
- Freeman, W. J. (1999). *How the Brain Makes up its Mind*. Weidenfeld and Nicholson, London.
- Freeman, W. J. (2000). Emotion is essential to all intentional behaviors. In *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development* (ed. M. D. Lewis and I. Granic), pp. 209-235. Cambridge University Press, Cambridge.
- Fuster, J. (2003). *Cortex and Mind: Unifying Cognition*. Oxford University Press, Oxford.
- Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. *Nature* **173**, 318.
- Geary, D. C. & Huffman, K. J. (2002). Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin* **128**, 667-698.
- Gehring, W. J. & Ikeo, K. (1999). Pax 6: Mastering eye morphogenesis and eye evolution. *Trends in Genetics* **15**, 371-377.
- Gershenson, C. (2001). Artificial Societies of Intelligent Agents. Ms Thesis thesis, Fundación Arturo Rosenblueth México.
- Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Zoology* **23**, 536-544.
- Gibbs, J., Maddison, S. P. & Rolls, E. T. (1981). The satiety role of the small intestine in sham feeding rhesus monkeys. *Journal of Comparative and Physiological Psychology* **95**, 1003-1015.
- Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press, Cambridge.
- Golani, I. (1976). Homeostatic motor processes in mammalian interactions: a choreography of display. In *Perspectives in Ethology*, vol. 2 (ed. P. P. G. Bateson and P. H. Klopfer), pp. 69-134. Plenum, New York.
- Gray, J. R., Braver, T. S. & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences, USA* **99**, 4115-4120.
- Griffiths, P. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. University of Chicago Press, Chicago.
- Griffiths, P. E. (1999). Squaring the circle: natural kinds with historical essences. In *Species: New Interdisciplinary Essays* (ed. R. A. Wilson), pp. 209-228. MIT Press, Cambridge, MA.
- Griffiths, P. E. (2004a). Is emotion a natural kind? In *Thinking About Feeling: Contemporary Philosophers on Emotions* (ed. R. C. Solomon), pp. 233-249. Oxford University Press, New York.

- Griffiths, P. E. (2004b). Emotions as natural and normative kinds. *Philosophy of Science* **71**, 901-911.
- Hacking, I. (2002). Making up people. In *Historical Ontology* (ed. I. Hacking), pp. 99-114. Harvard University Press, Cambridge, MA.
- Hardcastle, V. G. (1995). *Locating Consciousness*. John Benjamins, Amsterdam & Philadelphia.
- Harnish, R. (2001). *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science*. Blackwell, London.
- Hartwell, L., H. , Hopfield, J. J., Leibler, S. & Murray, A. W. (1999). From molecular to modular cell biology. *Nature* **402**, C47–C52.
- Haslam, N. (1998). Natural kinds, human kinds, and essentialism. *Social Research* **65**, 291-314.
- Haslam, N. (2002). Natural kinds, practical kinds and psychiatric categories: Review of Zachar on Biological-Psychiatry. *Psycoloquy* **13**, Biological Psychiatry (2).
- Hawkins, J. & Blakeslee, S. (2004). *On Intelligence*. Henry Holt, New York.
- Hebb, D. O. (1949). *The Organization of Behavior*. John Wiley and Sons, New York.
- Hennig, W. (1966). *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Hughlings Jackson, J. (1958). *Selected Writings of John Hughlings Jackson (Vols. 1 and 2)*. Staples Press, London.
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. Appleton-Century Co., New York.
- Hull, D. (1978). A matter of individuality. *Philosophy of Science* **45**, 335-360.
- Hutchinson, G. E. (1944). Limnological studies in Connecticut. VII. A critical examination of the supposed relationship between phytoplankton periodicity and chemical changes in lake waters. *Ecology* **26**, 3-26.
- James, W. (1884). What is an emotion? *Mind* **19**, 188-204.
- Johnson-Laird, P. N. (2006). *How We Reason*. Oxford University Press, Oxford.
- Keele, S., Ivry, R., Mayr, U., Hazeltine, E. & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review* **110**, 316-339.
- Keller, R. A., Boyd, R. N. & Wheeler, Q. D. (2003). The illogical basis of phylogenetic nomenclature. *Botanical Review* **69**, 93-110.
- Kelley, J. L. & Magurran, A. E. (2003). Learned predator recognition and antipredator responses in fish. *Fish and Fisheries* **4**, 216-226.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organisation of Brain and Behavior*. MIT Press, Cambridge, MA.
- Kirschner, M. W. & Gerhart, J. C. (2006). *The Plausibility of Life*. Yale University Press, New Haven.
- Koechlin, E., Ody, C. & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science* **302**, 1181-1185.
- Kornblith, H. (2002). *Knowledge and Its Place in Nature*. Clarendon Press, Oxford.
- Kripke, S. A. (1972). *Naming and Necessity*. Harvard University Press, Cambridge, MA.
- LaPorte, J. (2004). *Natural Kinds and Conceptual Change*. Cambridge University Press, Cambridge.
- Lazarus, R. (1991). *Emotion and Adaptation*. Oxford University Press, New York.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience* **23**, 155-184.
- Lepper, M. R., Greene, D. & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology* **28**, 129-137.

- Lepper, M. R. & Henderlong, J. (2000). Turning 'play' into 'work' and 'work' into 'play': 25 years of research on intrinsic versus extrinsic motivation. In *Intrinsic and Extrinsic Motivation: The Search for optimal motivation and performance* (ed. C. Sansone and J. M. Harackiewicz), pp. 257-307. Academic, San Diego, CA.
- Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition and Emotion* **13**, 481-504.
- Lorenz, K. (1950). The comparative method in studying innate behaviour patterns. In *Physiological Mechanisms in Animal Behaviour* (ed. J. F. Danielli and R. Brown). Cambridge University Press, Cambridge.
- Lorenz, K. (1965). *Evolution and the Modification of Behavior*. University of Chicago Press, Chicago.
- Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science* **72**, 444-467.
- Mallon, R. (2003). Social construction, social roles, and stability. In *Socializing Metaphysics* (ed. F. Schmitt), pp. 327-353. Rowman and Littlefield, New York.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review* **50**, 370-396.
- Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.
- McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience* **27**, 5796-5804.
- McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503-507.
- Mesulam, M.-M. (1998). From sensation to cognition. *Brain* **121**, 1013-1052.
- Michaud, F. (2002). EMIB -- Computational architecture based on emotion and motivation for intentional selection and configuration of behaviour-producing modules. *Cognitive Science Quarterly* **3-4**, 340-361.
- Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* **24**, 167-202.
- Miller, E. K., Freedman, D. J. & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts, and cognition. *Philosophical Transactions: Biological Sciences* **357**, 1123-1136.
- Miller, G. A., Galanter, E. H. & Pribram, K. H. (1960). *Plans and the Structure of Behaviour*. Holt Rinehart and Winston, New York.
- Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. MIT Press, Cambridge, MA.
- Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science* **56**, 288-302.
- Millikan, R. G. (1999). Historical kinds and the "special sciences". *Philosophical Studies* **95**, 45-65.
- Millikan, R. G. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge University Press, Cambridge.
- Millikan, R. G. (in press). Thoughts of real kinds. In *The Oxford Handbook of Philosophy of Psychology* (ed. J. Prinz). Oxford University Press, Oxford.
- Mountcastle, V. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology* **20**, 408-434.
- Mountcastle, V. B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Harvard University Press, Cambridge, MA.
- Murray, H. A. (1938). *Explorations in Personality*. Oxford University Press, New York.
- Myers, D. G. (2006). *Psychology (8th ed)*. Worth, New York.
- Nesse, R. M. (1990). Evolutionary explanations of emotion. *Human Nature* **1**, 261-289.
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.

- Northcutt, R. G. (2002). Understanding vertebrate brain evolution. *Integrative and Comparative Biology* **42**, 743-756.
- Northcutt, R. G. & Kaas, J. H. (1995). The emergence and evolution of mammalian neocortex. *Trends in the Neurosciences* **18**, 373-379.
- Oatley, K. J. (1993). Social construction in emotion. In *Handbook of Emotions* (ed. M. Lewis and J. Haviland), pp. 342-352. Guilford Press, New York.
- Ochsner, K. N. & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences* **9**.
- Odling-Smee, F. J., Laland, K. N. & Feldman, M. (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, Princeton.
- Olds, J. & Milner, P. M. (1954). Positive reinforcement produced by electrical stimulation of the septal area and other regions of the rat brain. *Journal of Comparative and Physiological Psychology* **47**, 419-427.
- Ortony, A., Norman, D. A. & Revelle, W. (2004). Affect and proto-affect in effective functioning. In *Who Needs Emotions? The Brain Meets the Robot* (ed. J. M. Fellous and M. A. Arbib), pp. 173-202. Oxford University Press, Oxford.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, New York.
- Perner, J. (1991). *Understanding the Representational Mind*. MIT Press, Cambridge, MA.
- Posner, M. I., DiGirolamo, G. J. & Fernandez-Duque, D. (1997). Brain mechanisms of cognitive skills. *Consciousness and Cognition* **6**, 267-290.
- Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press, Oxford.
- Proust, J. (in press). Metacognition and animal rationality. In *Rational Animals?* (ed. S. Hurley and M. Nudds). Oxford University Press, Oxford.
- Putnam, H. (1975). The Meaning of 'Meaning'. In *Mind, Language, and Reality* (ed. H. Putnam), pp. 215-271. Cambridge University Press, Cambridge.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press, Cambridge, MA.
- Quartz, S. (2001). Toward a developmental evolutionary psychology: Genes, development, and the evolution of human cognitive architecture. In *Evolutionary Psychology: Alternative approaches* (ed. S. Scher, M. Rauscher and (Eds).). Kluwer.
- Quine, W. V. (1969). Natural kinds. In *Essays in Honor of Carl G. Hempel* (ed. N. Rescher), pp. 5-23. Reidel, Dordrecht.
- Raby, C., Alexis, D., Dickinson, A. & Clayton, N. (2007). Planning for the future by western scrub-jays. *Nature* **445**, 919-21.
- Raff, R. A. (1996). *The Shape of Life*. Chicago University Press, Chicago.
- Reber, A. (1996). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford University Press, Oxford.
- Rieppel, O. (2005a). Modules, kinds, and homology. *Journal of Experimental Zoology B (Molecular and Developmental Evolution)* **304**, 18-27.
- Rieppel, O. (2005b). Monophyly, paraphyly, and natural kinds. *Biology and Philosophy* **20**, 465-487.
- Rolls, E. T. (1999). *The Brain and Emotion*. Oxford University Press, Oxford.
- Rolls, E. T. (2005). *Emotion Explained*. Oxford University Press, Oxford.
- Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In *Review of Personality and Social Psychology*, vol. 5 (ed. P. Shaver), pp. 11-36. Sage, Beverly Hills, CA.

- Rosenthal, D. M. (2000). Consciousness and metacognition. In *Metarepresentations: A Multidisciplinary Perspective* (ed. D. Sperber). Basil Blackwell, London.
- Routtenberg, A. & Lindy, J. (1965). Effects of the availability of rewarding septal and hypothalamic stimulation on bar pressing for food under conditions of deprivation. *Journal of Comparative and Physiological Psychology* **60**, 158-161.
- Rubenstein, J. L., Anderson, S., Shi, L., Miyashita, L. E., Bulfone, A. & Hevner, R. (1999). Genetic control of cortical regionalization and connectivity. *Cerebral Cortex* **9**, 524-532.
- Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* **25**, 54-67.
- Savage, L. J. (1954). *Foundations of Statistics*. Dover, New York.
- Schank, R. & Abelson, R. (1977). *Scripts, Plans Goals and Understanding: An Inquiry Into Human Knowledge Structures*. Erlbaum, Hillsdale, NJ.
- Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In *Approaches to Emotion* (ed. K. R. Scherer and P. Ekman), pp. 293-317. Erlbaum, Hillsdale, NJ.
- Schleidt, W. M., Yakalis, G., Donnelly, M. & McGarry, J. (1984). A proposal for a standard ethogram, exemplified by an ethogram of the bluebreasted quail (*Coturnix chinensis*). *Zeitschrift für Tierpsychologie* **64**, 193-220.
- Sheffield, F. D. & Roby, T. B. (1950). Reward value of a non-nutritive sweet taste. *Journal of Comparative and Physiological Psychology* **43**, 471-481.
- Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis*. Prentice Hall, Englewood Cliffs, NJ.
- Slooman, A. (2001). Beyond shallow models of emotion. *Cognitive Processing: International Quarterly of Cognitive Science* **2**, 177-198.
- Sperber, D. (2000). Metarepresentations in evolutionary perspective. In *Metarepresentations: A Multidisciplinary Perspective* (ed. D. Sperber), pp. 117-137. Oxford University Press, Oxford.
- Stanovich, K. E. (2004). *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. University of Chicago Press, Chicago.
- Sterelny, K. (1990). *The Representational Theory of Mind: An Introduction*. Basil Blackwell, Oxford.
- Sterelny, K. (1998). Intentional agency and the metarepresentation hypothesis. *Mind and Language* **13**, 11-28.
- Sterelny, K. (2003). *Mind in a Hostile World*. Oxford University Press, Oxford.
- Streidter, G. F. (1998). Progress in the study of brain evolution: From speculative theories to testable hypotheses. *The Anatomical Record* **253**, 105 - 112.
- Streidter, G. F. (2005). *The Principles of Brain Evolution*. Sinauer Associates, Sunderland, MA.
- Suddendorf, T. & Busby, J. (2003). Mental time travel in animals? *Trends in Cognitive Sciences* **7**, 391-396.
- Sun, R., Slusarz, P. & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review* **112**, 159-192.
- Swanson, L. W. (2003). *Brain Architecture: Understanding the Basic Plan*. Oxford University Press, Oxford.
- Thelen, E. & Smith, L. B. (1994). *A Dynamical Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, MA.
- Tinbergen, N. (1951). *The Study of Instinct*. Clarendon Press, Oxford.
- Tooby, J. & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* **11**, 375-424.

- Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (ed. J. H. Barkow, L. Cosmides and J. Tooby), pp. 19-136. Oxford University Press, Oxford.
- van den Heuvel, O., Veltman, D. J., Groenewegen, H. J., Cath, D. C., van Balkom, A. J. L. M., van Hartkamp, J., Barkhof, F. & van Dyck, R. (2005). Frontal-striatal dysfunction during planning in obsessive-compulsive disorder. *Archives of General Psychiatry* **62**, 301-309.
- von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.
- Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *American Zoologist* **36**, 36-43.
- Wagner, W. & Wagner, G. P. (2003). Examining the modularity concept in evolutionary psychology: The levels of genes, mind and culture. *Journal of Cultural and Evolutionary Psychology* **1**.
- Watson, J. D. & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press, Oxford.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review* **66**, 297-333.
- Whiten, A. & Byrne, R. W. (1991). The emergence of metarepresentation in human ontogeny and primate phylogeny. In *Natural Theories of Mind* (ed. W. A.). Basil Blackwell, London.
- Wilson, T. D. (2004). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Belknap Press, Cambridge, MA.
- Wong, R. (2000). *Motivation: A Biobehavioural Approach*. Cambridge University Press, Cambridge.
- Wood, J. N. & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience* **4**, 139-146.
- Zacher, P. (2000). *Psychological Concepts and Biological Psychiatry: A Philosophical Analysis*. John Benjamins, New York.
- Zacks, J. M. & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin* **127**, 3-21.

## Footnotes

---

<sup>1</sup> Movement can be of the whole body, of body parts (e.g., swimming, tail-flicking, flint-knapping), or ejections of matter or energy into the environment from inside the body (e.g., defecation, speech). Behaviour must be self-powered (i.e., locomotory) rather than dependant on external forces (e.g., drifting in a current of water or air). We do not count physiological movements such as heart-beats or peristalsis as behaviour because they do not interact with the environment. Such internal movements are also controlled by the autonomous nervous system rather than the central nervous system, and so are independently regulated.

<sup>2</sup> The behaviour production unit lets animals turn salient external and internal stimuli into adaptive outputs or behaviour. We have thus set aside input side topics such as sensation, perception, categorization and concept discrimination. We similarly also avoid issues such as selection among competing behavioural options and motor control on the output side. Finally, we have set aside, at least for the moment, some of the special complexities of human behaviour -- difficult issues such as temperament, mood, expression, and the role of culture. For the most part, these either modulate processes we do discuss or are late additions to behaviour production units.

<sup>3</sup> Since behaviour does not fossilise, we cannot reconstruct the behavioural abilities of our ancestors directly; we can only make inferences based on extant animals that may, or may not, be representative of the behavioural capabilities at a certain point in our evolutionary past, given that contemporary species have had many millions of years (in some cases) to evolve new abilities. Our argument regarding the timing of new types of production systems is based not on the most advanced contemporary exemplar in some clade, but rather on the inferred qualities of a *prototypical* ancestral species in that group.

<sup>4</sup> Just about any behaviour is likely to be accompanied by learning. The knowledge acquired in this way can modify the way in which a particular BPU works in future. However, this does not make all behaviour aptitudinal in our sense because the primary function of those behaviours is to achieve other kinds of end-states, rather than being devoted to learning *per se*.

<sup>5</sup> Experimental psychologists nowadays avoid use of the word 'drive', largely because a number of results have been interpreted as refuting Hull's classic drive reduction theory (Hull, 1943). However, in each case, these experimental results can be interpreted as 'tricking' the reward system in a way which natural selection cannot be expected to have foreseen. First, rats will work to obtain saccharine rewards even when not calorie deficient (i.e., not in a state of 'hunger') (Sheffield & Roby, 1950). However, saccharine constitutes an artificial, non-nutritive source of sweet taste, which is a proximal cue of ripe fruit in an animal's evolved niche. Second, hungry rats will work to obtain direct electrical stimulation of their mesotelencephalic dopamine system, even over the option to receive food (Olds & Milner, 1954; Routtenberg & Lindy, 1965). This surgical procedure short-circuits the evolved reward system; selection is unlikely to predict a technology which provides rewards in the absence of the requisite biological resources following behaviour. Showing that maladaptive outcomes can be obtained through such artificial procedures does not refute the validity of the general claim that *drives* have evolved to produce need-directed behaviour

---

which improves an animal's fitness, and that such behaviour tends to appropriately manage internal resource levels under 'natural' ecological conditions.