David Rose

Three Dimensions of Psychosemantics

What is meaning? The question has a long history in both 'western' analytic philosophy (e.g. Ogden & Richards 1923; Putnam 1975; Garfield 2000; Millikan 2004) and in the 'continental' phenomenological tradition. Perhaps the most commonly used definition involves the aboutness or intentionality of mental states (Brentano 1874). How such states acquire their meaning is another matter, however. Here, I will argue there is no single mechanism, and hence no simple answer or definition. I will follow first the approach taken by analytic philosophy and cognitive science, then show how this has been extended recently by adding extra factors to form a more complex theory structure. Specifically, I will describe three conceptual dimensions to psychosemantics that need to be recognised.

One of the specific issues driving this research is the question of whether meaning can exist sub-personally, or only at or above the level of the individual person or mind. From the positivists to the phenomenologists, the primacy of human experience as the basic source of meaning has been a fundamental tenet. Yet this does nothing to solve the explanatory gap between mind and body, hence modern research has explored ways in which the meaning of 'meaning' can be revised in ways that avoid such species chauvinism, relativism, subjectivism, dualism and the man in the head fallacy (e.g. Dennett 1969; Millikan 2004; Rose 2006; Breidbach 2007; Mandik 2008). The overall aim is to naturalise the mind by integrating it with biology in a series of smaller, more comprehensible steps (e.g. Revonsuo 2006; Rose 2006) rather than leaving an unbridgeable gulf between the two, as in the traditional approaches. Here, I debate how meaning fits within such a complex multi-level theory.

The paradigm case-study in the field stems from Lettvin et al.'s (1959) paper titled "What the frog's eye tells the frog's brain". While the notion of one part of the nervous system 'telling' something meaningful to another part is a useful metaphor, and is commonly used productively throughout cognitive neuroscience, the question is how literally neural interactions can be described as 'telling' or 'communicating'. [Note that Lakoff & Johnson (1980) have argued that metaphors are commonly used throughout human discourse; hence when we are performing scientific research ('exploring unknown territory'), and postulating new mechanisms or hypothetical entities such as those that

GESTALT THEORY

(ISSN 0170-057 X)
Vol. 34, No.3/4, 259-286

might bridge the explanatory 'gap' between mind and brain, we cannot avoid using metaphorical language, even if only for a while. As evidence accrues over time, such metaphorical terms might come to be accepted as referring literally to reality (e.g. Dennett 2011).] These words imply that meaningful content is being transmitted internally within the nervous system. The same question arises within mental systems as within neural, since both are divisible into component parts or modules that interact with one another; hence the principles developed to answer the frog's brain question are assumed or proposed to apply also to the human mind.

Cognitive theories of stimulus identification posit the existence of mental representations as the internal bearers of meaning (Fodor 1975; Sterelny 1990; Clapin et al. 2004; Bechtel 2008; Nielsen 2010). These representations are normally understood as playing a symbolic role, in that they stand for something else, ranging from a concrete object to an abstract idea. [Representations may or may not be in some way isomorphic with what they represent; see further in the Discussion. Also, to avoid prejudging the issue of the origins of meaning, I will take it that any mental state may represent both immediate sensory awareness and interpreted or conceptual knowledge; hence I do not use the terminology of 'presentation' versus 'representation' for these respectively as in the phenomenological tradition (e.g. Albertazzi 2006, xi; 2007; Poli 2006).] Representationalism is central not only to perceptual psychology but also to fields as diverse as cognitive science, linguistics, art, philosophy of mind, semantics and semiotics in general. It is also implicit in the neurophysiology of perception, where feature detectors, grandmother cells and labelled lines (Rose 1996; 1999a) are presumed to represent a stimulus or aspects of it (for example: "The essence of all neural keys or decoding systems is the notion that the activity of a particular cell always *means* the same thing and that the decoder *knows* what that thing is: the cell is labelled ...": Harris 1997, 165; my italics). But how does — how can — neural activity have meaning? What connects it to its referent and determines its exact meaning? How can the frog retina 'tell' the brain "there is a bug" and the brain 'understand' what the retina is 'saying' (Lettvin et al. 1959)?

Dimension 1: Atomism versus Holism

In traditional analyses, two broad theories have been prominent. Table 1 illustrates the diversity of ways in which this fundamental dichotomy has been expressed, which depend on emphasis and field of origin. For present purposes I will firstly outline the overall gist of each approach, which I will for simplicity call 'atomism' and 'holism' (without intending to imply that these terms describe their key features).

atomism	—	holism
reference	—	sense
natural signs	—	interpretations
indicators	—	inferences
causal role	—	conceptual role
extension	—	intension
broad/wide content	—	narrow content
correspondence	—	coherence

Table 1 Characterisations of the traditional dichotomy in psychosemantics.

Philosophers familiar with psychosemantics may be able to skip through this section, but for others it should provide useful background to the rest of the paper. These two theories are nowadays well-known to be flawed, yet it is valuable to present each in its full light, firstly for fairness to each theory, secondly to introduce the criteria and test cases by which any subsequent theory has come to be judged, and most importantly for the present paper to lay them out as the extreme opposite poles of a dimension which defines one axis of the conceptual space of possible theories of psychosemantics. Once so drawn the line between them can first be searched for intermediate positions with various weightings or combinations of elements from both extremes, and, later, combined with other dimensions similarly analysed to see if these together enable us to construct more coherent and comprehensive theories. (See Rose & Dobson 1985; 1989; for examples from vision and learning of how such multi-dimensional concept spaces are useful in mapping and understanding fields in which there is a multitude of conflicting theories, and how these spaces then can facilitate the generation of appropriate tests and new theoretical syntheses.)

Atomism

Atomism says a representation means, for example, 'bug' because it is caused to exist by the existence of a bug in the outside world. When a bug is in front of the eye, neural processes are put in chains that in some way evoke, generate or create the representation. This causal, physical and deterministic link from the referent ensures the representation means 'bug' because bugs activate it. Each such link is independent of all others, meanings are isolated nuggets of truth, and representations are univocal — each having a single clear meaning (Figure 1).

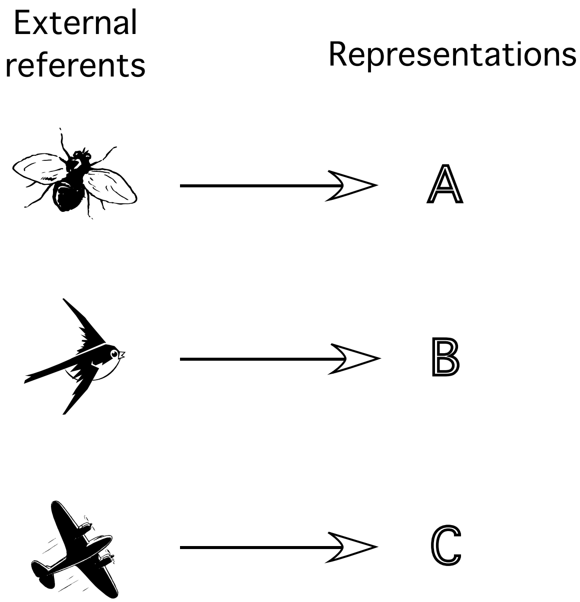

Fig. 1 Atomism posits that particular objects evoke particular representations (A-C) on a one-to-one basis.

In psychology and AI it is common to adopt a two-stage approach, in which finding the meaning of a stimulus is the same as identifying it, which is accomplished by fitting the immediate sensory representation to a stored template or a category held in long-term memory. The match may be exact or to the nearest approximation, and may be described variously at the neural level as prolonged resonance or as the settling of the sensory system into a minimal entropy attractor state (depending on exact theory: reviewed by Rose 2006, ch. 9). These approaches nevertheless still assume there is ultimately a single correct meaning/identity for every stimulus, hence they can be classified as atomist in spirit.

There is intuitive appeal and simplicity to such an approach, but several problems face atomist theories.

Problems with Atomism

First is their inability to explain mis-representation, a condition wherein a representation does not indicate the true state of the world. Atomism thus does not account for hallucinations, illusions, delusions or just plain mis-identifications. These seem to be 'meaningful' to the mind in question, even though they are not connected by causal links to the true state of the world. For example a small moving spot of black metal or cardboard can trigger activity in a frog's 'bug detector', as can a blackbird flying past in the distance. Thus it is not only bugs that activate bug representations, and so these cells do not simply indicate 'bug' (e.g. Fodor 1987).

In language, synonyms by definition have the same referent. But mental representations may not be known by their possessor to be synonymous. Thus they can refer to the same object yet play different roles in the cognitive system. For example, like Oedipus you may know that Jocasta is attractive, and that you cannot marry your mother, without necessarily knowing that Jocasta is your mother, i.e. that your thoughts about these two people are co-referential. Hence you may pursue different actions consequential to each of those two thoughts (in this case, proposing marriage to Jocasta but not your mother). Yet atomism would ascribe the same meaning to each thought that in fact has the same external referent (Fodor 1987, 74). The duality of meanings for Oedipus is not captured by atomism.

Both these mis-representational problems arise because correct meaning is equated with absolute truth, and so the meaningfulness of our thoughts can only be defined objectively from an inaccessible, transcendent point of view (from 'God's eye'; this assumption is also called 'objectivism' by Lakoff & Johnson 1980). This is the 'meaning for whom' issue — who or what understands the meaning: who reads the message, interprets the signal or decodes the information?

Even more obvious however are the problems of accounting for abstract concepts (truth, morality, numbers and so on) and for counterfactuals in general. The latter include thoughts about non-existent entities (Superman, unicorns) and states of affairs (the future, what the world would be like if X had occurred or would occur, what is not true) about which we can think and reason meaningfully despite their 'unreal' or non-physical status.

Similarly, many of our commonest concepts concern categories, which may not have objective and precisely definable referents (e.g. how do you define a 'game'? Wittgenstein 1953) and which can vary between individuals in different cultural or linguistic groups (Lakoff & Johnson 1980; but cf. Millikan 2000, 42-48; Shapiro 2011, ch. 4; for recent reviews see Cohen & Lefebvre 2005). Returning to our toy example: as usually understood, a true 'bug' detector responds not just to one particular bug but to any, so ideally its selectivity should be sufficiently general for it to be evoked by any member of the entire category of bugs *but* by no other stimulus (iff bugs; but cf. Millikan 2000, 24-32; 2005). At the implementational or neural level, it is however problematic whether there is, or could be, a mechanism to give such perfect selectivity. From the early ideas of Gestalt isomorphism to modern attempts to discover the neural wiring underlying feature selectivity (e.g. even the simple case of orientation selectivity in V1, let alone something as sophisticated as a truly bug-selective cell) there have been no convincing answers, only those compatible with coarse tuning and noisy neural encodings (e.g. Quian Quiroga et al. 2008; cf. Bowers 2010). However, if representations are only roughly selective (fuzzy categories or 'bug-like' detectors) then meanings are approximate too; yet surely some meanings are exact (propositional rather than analogue, binary rather than metric), so a complete answer still evades us.

Further, our knowledge is not a random collection of isolated atoms but shows structure (systematicity). It is highly organised into a hierarchy of categories, and some pairs of ideas are clearly more similar than others (e.g. compare red in turn with pink, green, sweet, elephant and justice).

Finally there is the indeterminacy of causation. Mental activity is always consequential to a chain of causes, and we have no reason to pick any single one as *the* single true source of the representation and thus its meaning. For example in the case of a bug detector is it the distal stimulus (the bug), the proximal stimulus (the retinal image), transduction in the receptor cells, the chain of afferent action potentials and transmitter releasings, transmitter contacting the dendrites of the bug detector itself, binding and changes in conformation of its glutamate receptors, ion flows, current passing along the dendrites, or what? All of these can truthfully be said to cause the cell to fire (or to be part of a converging set of causes: Reichenbach 1928; Schaffer 2007); so, to which cause does the firing refer?

Holism

The alternative approach, holism, says that representations cannot have meaning in isolation, but must instead be part of a system of inter-related concepts. Any given semantic unit is always situated in a context that defines its significance. Contra atomism, representations of abstract concepts, fictions, generalisations, counterfactuals and so on can have meaning since they are no different in kind from representations of concrete objects and events.

As an analogy, consider the word "burro". This means 'butter' — if you happen to be speaking Italian. But it means 'donkey' if you are speaking Spanish; and it has no meaning within most other languages. That is, meaning depends, at least in the case of words, upon which language you are speaking. Similarly within a language: synonyms are differentiated by their context (cf. does 'bark' refer to a dog or a tree? In 'they are eating apples', are 'they' people or apples?). By analogy, mental representations only have meaning in the context of a system that exists in someone's mind.

Historically, both associationist and cognitive psychologists have used holistic conceptions. First, workers as diverse as James (1890), Hebb (1949) and Crick & Koch (2003) have suggested a role for associations in generating meaning. Phenomenally, stimuli evoke not just one percept but also linked or correlated ideas and a sense of anticipation or expectancy as to what will happen next, and it is these that together tell us what a stimulus 'means'. They place it in a scenario or context: a spatial scene and a temporal narrative. This process is underpinned by sensory input arousing a chain of activations that spread from one representation to another throughout the perceptual and mnemonic systems, which together constitute a 'semantic network' of linked ideas. Subjectively, this leads not only to closely related concepts but also to a range of subtle connotations, overtones and implications. For example, 'red' makes us think not just of specific red objects we have seen (e.g. blood, a fire engine, a ripe tomato), but also of danger, heat, communism, sex, guilt, etc., depending on context. For an English speaker, 'burro' may evoke a word which is similar in some way, such as 'burrow' or 'borrow', giving 'burro' one or more implied or subliminal meanings even in the absence of a specific definition. Meaning arises automatically and spontaneously via these links.

Under associationism, meaning thus derives from the spread of activity to other representations. But how do these representations in turn get their own meanings? If the processing were purely serial, they would have to depend upon yet further associations, which would lead to an infinite regress. Instead, there could be a recursive system of connections, including feedback to earlier stages of processing, in which meanings would be mutually defining (like looking up a word in a dictionary: what do you find? — other words) and the entire set of representations would form a coherent integrated system. Under this holistic model the meaning of any individual unit depends both on its place within the context of the complete mental system and on its role within that system: the effects it has on other representations and vice versa.

Thus in modern analytic philosophy and cognitive science, holism sees a representation as having meaning insofar as it plays a role in the chains of inference and reasoning consequent upon its existence. For example a sensory detector's firing means 'bug' (Lettvin et al. 1959) if it is appropriate for dealing with bugs. But this response is not automatic, like a reflex. Instead, there is a choice of possible behaviours (a human might engage in catching, swatting, ignoring, observing, indicating, naming, etc.; and even frogs and toads are flexible in their responses to bugs: Neander 2006). The flexibility arises firstly because a variety of other representations may interact with any given representation before motor output arises, and secondly because there are several types of interaction (such as antithesis, category membership, causation, ownership and so on; Figure 2).

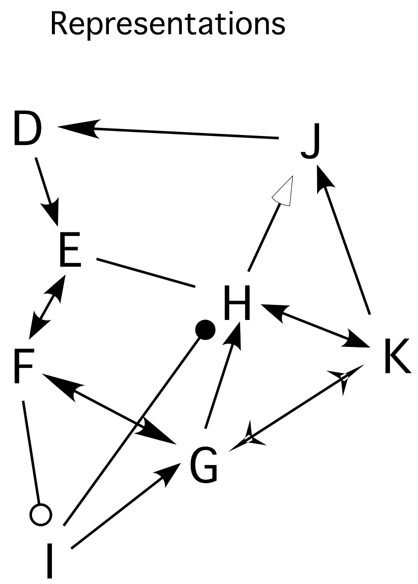

Fig. 2 Holism: all the representations (D-K) are connected in a semantic network. (Here, the connectives are of various types, as posited by cognitive rather than associative or semiotic theories.)

This last point crucially refers to the idea that any intelligent system must necessarily process 'symbols' according to 'rules', and the interactions within the system cannot be merely of one type, as postulated by associationism (e.g. Fodor 1975; 1987; 1994a; 1998). This 'computational' metaphor proposes that many different rules or operators exist that can be applied selectively to transform one or more representation to generate another. This process is akin to making an inference or drawing a logical conclusion from a set of premises or axioms. The mere passive spread of activity via associative links that form a network is far too simplistic a model to explain the creativity, generativity, compositionality and imaginativeness of our thought processes.

However, despite these explanatory virtues there are also problems with holistic theories of semantics (see also Carnap 1928; Putnam 1975; Millikan 1984; Fodor & Lepore 1992; Fodor 1998).

Problems with holism

Under holism, no sign or stimulus has univocal meaning. It always initiates a chain of associations or inferences which may go off in a variety of directions (the endless play of meaning: Derrida 1967). Hence, holism is akin to atomism in that, for both, meaning depends on an indeterminately long chain of causation (though in holism's case it runs downstream from the representation). Worse, it is more complicated in that the chain is not linear/serial but a convoluted knot of mutually-modulating alternatives.

Since a mental representation only has meaning within its context, i.e. the entire mind, it follows that its particular meaning depends on which mind it is. But because everybody has a different semantic network, two people cannot hold the same belief (Fodor 1987, 55-57). This makes precise communication and interpersonal understanding impossible (Quine 1960).

Further, since a representation's meaning depends upon, and is in turn affected by, links with other representations — indeed, indirectly with the entire network — its meaning is not fixed across time, because any single change anywhere in the system (e.g. due to learning) alters, if only subtly, the meaning of every representation within that system (Quine 1951). In practice, one can argue, such perfect interconnectedness never happens. New information does not lead to a comprehensive check-through for its consistency with the existing knowledge base and appropriate re-organisation if inconsistency occurs, so within any given mind there are always many mutually contradictory beliefs (Millikan 1993). But perhaps we have instead a number of smaller holistic nets, each forming a 'module' with its own proprietary inbuilt knowledge and skills/mechanisms, e.g. one for vision, another for hearing, language and so on (Fodor 1983; 2000; Coltheart 1999; Neander 1999; Callebaut & Rasskin-Gutman 2005). Each acts as a 'frame' circumscribing the knowledge that is relevant for creating the representation's meaning. However, this leaves the question of how these modules communicate with each other, as well as all the above problems (Rose 2006, 143-147).

The question of how arbitrary symbols (i.e. those non-isomorphic to their referent) can have their meanings fixed is known as the grounding problem (Harnad 1990; Marcus 2001; de Vega et al. 2008). An internal network of representations, even a perfectly coherent one with no inconsistencies or contradictions, has no intrinsic orientation relative to the world. Such a system could be rotated, as for example in the classic inverted spectrum conundrum (what you see as 'red' might be what I experience as 'blue') or in multi-dimensional scaling (Clark 1993). There needs to be some way of anchoring the network to prevent this, so that particular representations are tied to particular elements of meaning or referents. Several such validating anchors are required to orient and scale the whole network.

A Hybrid Theory

An obvious reply to the grounding problem is to combine aspects of atomism and holism, for example by postulating two types of representation. Perceptual representations acquire their meaning as specified by atomism, i.e. their existence is directly caused by their referents, so they have the content they do in virtue of the particular stimulus and deterministic physical mechanisms that give rise to each of them. Conceptual representations, on the other hand, have their meanings determined by the functional relationships within the holistic system of representations. This system includes perceptual representations, so concepts can be indirectly evoked by stimuli, and have their ultimate meanings derived from those, hence grounding the higher-level cognitive symbols (Figure 3). The meanings of 'horse' and 'horn' may derive from the external environment, but putting them together to form 'unicorn' is purely an internal mental process. Also, 'horse', 'cow' and so on can generalise/abstract to the category 'animal' (and firefly, mosquito, etc. to 'bug'). It solves the inverted spectrum puzzle and the communication problem (since two people can have representations with the same external referent). The division into two types of representation accords with our intuitive distinction between perceptual experience and conceptual thought, with the division between perceptual and cognitive illusions (e.g. Gregory 2009), and is in the traditional empiricist spirit of Locke and Hume which has been revived in various forms in answer to the above single-factor theories (e.g. Dennett 1969; Dretske 1988; Sterelny 1990; Neander 1995; 1999; Garfield 2000).

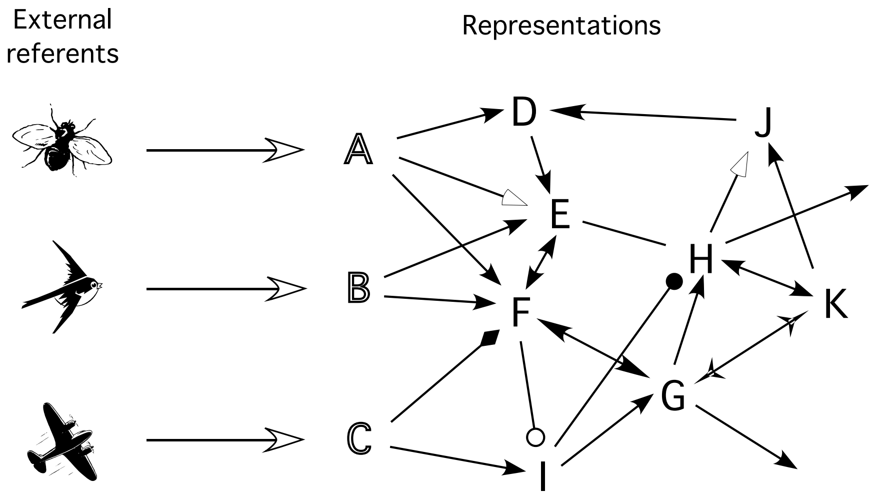

Fig. 3 Hybrid model with two types of representation.

Problems with the Hybrid Theory

First, there are questions about where the boundary between the two types of representation lies, and whether the division is sharp or gradual (see Rose 2006, 143-147). Indeed, some have argued against the existence of any such dichotomy, though mainly because they think it could lead to Cartesianism (e.g. Merleau-Ponty 1945; Dennett 1991). However, such moves simply take us back to the problems of the single-factor theories.

Most crucial is that, even within analytic philosophy, the hybrid theory itself still doesn't avoid all of the problems described above for the traditional two theories, especially that of mis-representation. Although it reduces the endless play of meaning among concepts (including abstractions and counterfactuals) that follows from pure holism, it does not prevent it altogether, nor the sequence of changes caused by forming (by inference or learning) a new concept, nor the communicability of concepts from person to person. Worse, it adds together the two indeterminacy problems: from atomism, the causal chain leading to the representation; and from holism, the endless play of the representation's effects throughout the system.

So, how do we get out of this? Atomism, holism and the hybrid theory all appear as unsatisfactory positions along the dimension. As in many impasses, resolution comes by recognising that the situation is in fact more complicated and that there is an orthogonal dimension which needs to be taken into account (a process akin to the delineation of search spaces in problem solving tasks: Rose & Dobson 1985; 1989; Shrager & Langley 1990). Advocates of this second dimension have included atomists, holists and hybrid theorists.

Dimension 2: Synchronic versus Diachronic Functions

The traditional theories assume that a representation's function reflects or describes what it does, i.e. what it is doing now (synchronically). New theories however suggest that something's function can be defined as what it is there to do in the future. We can assign it such a teleological purpose based upon what it has done in the past. This diachronic view is known as teleological functionalism and is the cornerstone of recent 'teleosemantic' theorising in psychosemantics and neurosemantics, particularly aimed at solving the problem of mis-representation (Wright 1973; for recent reviews see Buller 2005; Macdonald & Papineau 2006; Rose 2006; Lycan & Neander 2008; Neander 2008; and for a specific application to vision see Jacob & Jeannerod 2003, ch. 1).

The general thesis here is that representations (and/or the mechanisms that create them) exist because they subserved a useful role in the past. Natural selection permitted the survival only of creatures that possessed them. That past role determines the current representation's 'normal', 'proper' or ideal function, i.e. what it is *supposed* to do. Just as a screwdriver can be used to open a tin can yet it is still a screwdriver, and a failed heart that does not pump blood is still a heart, so a bug-detector activated by a distant blackbird is still a bug-detector — because it evolved to detect bugs not blackbirds.

Under this remit, the meaning of a representation is grounded by (or reduced to) its function, which in turn is its purpose, which has been determined by the way it has been shaped through its history. That history has ensured that the representations present in the brain/mind fit their environment, so stimuli activate them like a key fitting into a lock (to put it metaphorically — e.g. the causal interaction may be informational rather than mechanical, depending on level; see below). Evolution has shaped each lock so it receives and responds appropriately to a particular stimulus that is likely to exist in the organism's environment — where 'appropriately' means in such a way as to increase survival and reproduction.

This grounding naturalises meaning and thus circumvents the need for a transcendent, God's eye view, which is inherent in the traditional truth-based definitions (although without rendering truth obsolete: see Discussion). Mis-representation is explained by recognising the distinction between what the representation is supposed to do and what it actually does. A bug detector may actually respond to any moving dark spot of light in its receptive field, whatever that spot's cause, but if it evolved to detect bugs, in that activation of ancestral homologues has historically aided survival along the germ line of that frog by detecting bugs, then it can legitimately be called a bug detector. Thus mis-representation can occur without negating what the mechanism's evolved proper function is.

Unlike traditional synchronic holism, teleofunctional meaning is not instantly mutible by changes elsewhere in the semantic network. Holism assumes that functional meaning is defined by the immediately acting web of causes and relationships between representations, hence making it alter as soon as any change occurs within the network, which makes the system unstable and unreliable. Teleofunctionalism explains how meaning can be stabilised — yet without creating an unworkable inflexibility. Meaning can be changed, indeed created ab initio, but over slower, evolutionary (and ontogenetic — see below) time-scales. This also makes meaning comparable across individuals who have comparable evolutionary histories.

Although based on the axiomatic presuppositions of Darwinian natural selection, it has also been argued that empirical evidence and comparison across species (e.g. finding homologous structures) can provide a solid grounding for particular ascriptions of teleofunctionality (Amundson&Lauder 1994). Moreover, as a research tactic it works to reason top-down: begin by assuming the current environment is much like the past and adaptation is perfect; in other words, act as if the system has been 'designed' (Dennett 1987; 2011). This then gives us testable hypotheses about how the given system might be working as well as why. Such top-down constraints should be used when generating hypotheses: we need to reason a priori from the cybernetic necessities that dictate what a system must be like in order for it to survive and reproduce (Dobson & Rose 1985; Schroeder 2004).

Problems with Diachronic Teleofunctionalism (and some Replies)

How can evolutionary history explain our ability to recognise doorknobs, film stars, flying saucers and national flags? Surely such ideas cannot be innate? The conventional answer among teleofunctionalists is that similar principles of selection apply during individual learning history as over evolutionary time-scales. Trial-and-error learning shapes memory by processes analogous to the way Darwinian natural selection shapes the organism (Dennett 1969; Papineau 1987, 65-67; Dennett 1995, 373-381; Garson 2011). Another answer is that it is the mechanisms of representation construction that evolved with the teleological purpose of generating representations appropriate for the occurrent environment (Millikan 1984, 46-47; Fodor 1994b, 76-77; 1998, 142; Kingsbury 2008). Millikan has emphasised that both the producing mechanisms and those that process the representations further (the 'consumers') must co-evolve, since there is no advantage in having knowledge that is not usable (e.g. Millikan 1984; 2006). Further, Millikan (2004) sees evolution as first developing representations akin to Gibsonian affordances, in which the emphasis is on what they do rather than what activates them, but which nevertheless represent both the stimulus and the response. However, she also argues that later evolution necessarily gave rise to pure perceptual and pure motor representations in addition to the affordances. These can be flexibly combined and configured so we can learn to deal with novel and complex situations.

Now, no biological system is perfect — for a number of reasons. These include not merely that it might be broken, unavoidably inefficient or noisy, but also because it takes time to adapt to a new environment — and if the environment keeps changing the biology may never catch up. This idea does however explain how we often possess multiple mechanisms with similar functions (e.g. for depth perception we can use disparity, perspective, haze, occlusion and so on: Ramachandran 1985). At different phases in evolutionary history, various traits needed to be selected, so over time we have acquired a tool-kit of assorted devices, some of which have overlapping functions. Each can do only an imperfect job on its own in the current environment, but together they can achieve a much higher degree of performance than any individual mechanism. This in turn reduces the selection pressure for perfecting any of the individual mechanisms. Teleofunctionalism can thus accommodate the atomist problem of fuzzy implementation and categorisation, and the holistic problem that our thoughts are sometimes mutually contradictory. Note that while an imperfect mechanism individually conveys imprecise meaning, multiple overlapping representations can give not just increased precision (Ramachandran 1985) but also flexibility of choice, depending on which representation (or which combination) is used by the other representations active in the occurrent context.

Nevertheless, recognising that our evolutionary history has been long, chaotic and complex raises the question of how we can possibly specify what the historical origin of any particular representation has been. Despite its virtue in explaining imperfection, taken to the extreme teleosemantics is too flexible. If the whole environment changes, or even if only one small aspect changes, this can lead to completely new uses for a representation, taking it away from the purpose for which it previously evolved (Gould & Lewontin 1979) and leaving an inescapable indeterminacy (Dennett 1995, 401-412). If we try to ascribe univocal meaning via history (e.g. that a representation evolved to detect bugs) we are merely speculating. Therefore, some have rejected teleosemantics in favour of the idea that meaning can only be defined relative to the current environment, i.e. synchronically (proposed by Cummins 1975 and supported by Davies 2001 and Craver 2007). However a reasonable compromise is conceivable in which both factors should be considered (Buller 1998; Garson 2011). In this hybrid view, which definition is relevant depends on which question you ask: evolution and ontogeny determine whether a mechanism is supposed to be a bug detector, but whether it is actually detecting a bug now depends on whether there is a bug in the current environment.

Dimension 3: Level of Analysis

A remaining major question is that of meaning indeterminism versus the univocality of representation. This is best understood by bringing in a third orthogonal dimension, that of level of organisation. [Note that here I am using the word 'level' to refer to a layer or stratum of organisation, description or function, and that I distinguish it from 'stages', which are the successive processes that occur along a sensory pathway within bottom-up, serial processing models of perception. (For a review of levels see Craver 2007, ch. 5; for a phenomenological approach see Poli 2006, who asserts that the psychological level is fundamentally different from higher and lower ones.) Unlike Poli (2006) but like Revonsuo (2006) I posit a commonality of organisational principle across all levels, such that phenomenality can be understood as a biological emergent phenomenon — indeed that the biological, psychological and social levels are all thus united.] The poles of this dimension assume that consciousness originates from either of two sources: our physical and biological natures (e.g. Crick 1994), or our social, cultural and linguistic context (e.g. Cooley 1902; Mead 1934; Lakoff & Johnson 1980; Burkitt 1991; Fireman et al. 2003). [Associated with these theories are bottom-up and top-down determinism respectively. However, attempts to explain how causation can operate between levels have proved unsatisfactory, from Descartes to Kim (1998). Instead, causation, to the extent it exists at all (Dobson & Rose 1985; Schaffer 2007), runs horizontally between entities within a level, while the relationship between phenomena at different levels is one of identity (Mackay

1980; Rose 2006; Fazekas & Kertész 2011).] The patent obviousness of both sources however demands hybrid theories which accept that all levels are relevant to the understanding of each other (e.g. 'social cognitive neuroscience': Ochsner & Lieberman 2001). Moreover, some theories posit more than just three levels (physical-mental-social), since within each of these there are obviously several, forming complex structures (e.g. Lycan 1987; Scott 1995; Revonsuo 2006; Rose 2006, ch.4; Craver 2007; Bechtel 2008; Cacioppo & Decety 2011; Feinberg 2011). The exact number of levels in total is an open question, hence this approach is simply known as multi-levellism. In modern versions of this theory, Nature is seen as spontaneously forming into a nested hierarchy of systems (where a system is defined as a collection of interacting parts organised so as to generate new emergent properties not possessed by any of the parts in isolation). Each part is itself a system that can be analysed similarly into its component sub-systems plus their organisational structure (Figure 4). For each system, its immediate environment is the surrounding units at the same level, with which it interacts to form a higher level system. Within a system or module the connections are in relevant ways closer and stronger than those between systems at the same level — a configuration known as 'nearly decomposable' (Simon 1962) and which can be shown to be optimal on a number of quantitative criteria (e.g. 'small world network' analysis: Sporns 2011; and evolutionary adaptability: Lorenz et al. 2011).

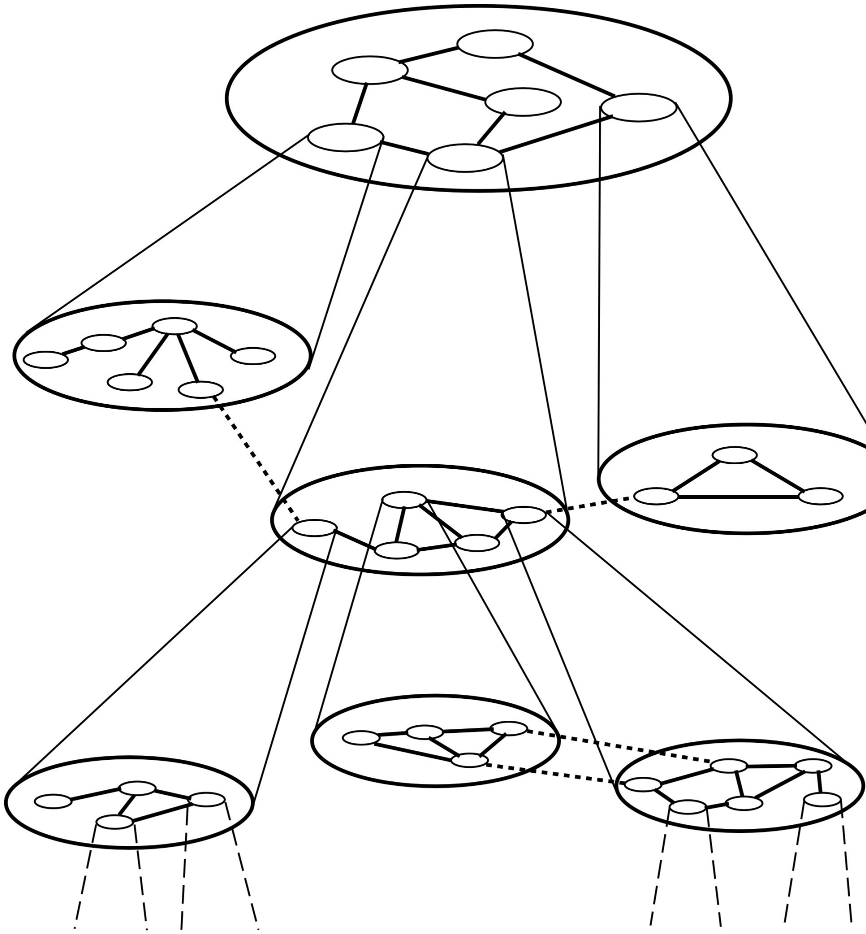

Fig. 4 A fragment of a nested hierarchy of systems, illustrating the increasing number of systems at increasingly lower levels. Note the dotted connections indicating inter-module interactions within a level. Most intralevel connections are two-way.

Now, where and how does 'meaning' fit within this worldview? Within a multi-levelled understanding of 'meaning' there is still not necessarily a single meaning to any given activity, but the indeterminacy is vertical as well as horizontal. This sounds worse if you think meanings should be univocal, but the additional complexity is well motivated and all the more realistic for it, and makes sense of some of the above problems. Thus previously we saw that a problem within the atomistic explanation of bug detection is the indeterminately-grained chains of serial and converging causes and events that can be traced back to the real bug (and in principle, determinism says we could also track back along the chains that lead to the bug's existence there and then, which stretch all the way back to the big bang!). Holism includes internal loops which indefinitely expand, elaborate or confuse the representation's meaning. Teleological functionalism adds even more distal functions, i.e. it asserts that evolved representations aid survival and reproduction, so therefore they all have the same meaning. Putting all these together would apparently lead to the conclusion that a bug detector firing really means not just 'bug' but also 'survival', 'sex', 'food', 'moving dark spot', 'glutamate release onto my dendrites', etc. (Lycan 2001; Rose 2006). However, this is not quite right.

Thus, processes do not just run at a single level. Every event exists at multiple levels (Figure 5). For example a given incident of 'release of neurotransmitter' may be just one small part of a series of events that together constitute the higher level event 'transmitting a message from neuron A to neuron B'. That in turn is one small event in a series that mediates 'activating neuron B' .... and so on to 'detecting a spot in the retinal image', 'detecting a bug', 'finding food' and to 'surviving'. At higher levels each event has longer duration and is comprised of a larger number of lower-level events integrated to form a system (Simon & Ando 1961; Salthe 1985). Thus under multi-levellism, each element interacts with other elements at the same level to construct a system which has new, emergent properties — so that, to coin a phrase, the whole is greater than the sum of its parts, which is to say qualitatively different. With regard to semantics, these properties include new levels of meaning. Putting these ideas together: as stimulus-evoked activity spreads through successive stages of neural processing, recurrent activity integrates those stages to form a larger and larger system (e.g. Pollen 1999; Lamme 2004; Roelfsema 2005; Clarke et al. 2011a; 2011b). Hence new levels of meaning may emerge progressively over time as the system grows and self-organises or settles into one new configuration after another. The more stages that become involved, so the more levels are created, and the deeper and more qualitatively novel the understanding of what the occurrent situation 'means' becomes (Rose 1999b; 2006, 107-110, 345-348 and 392-395).

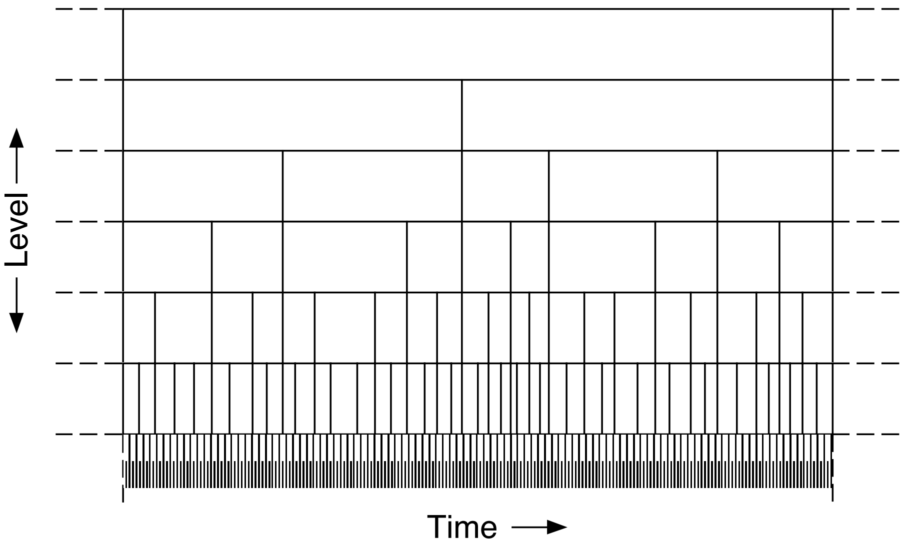

Fig. 5 A nested hierarchy of events across time. An event is a relatively stable state separated from the preceding and following events at its level by relatively rapid transitions (Kauffman 1993; Sporns 2011).

In general, a higher level emergent system suppresses the variability and noise of its lower level components, which it constrains in top-down fashion (e.g. Dobson & Rose 1985; Salthe 1985; Kaufmann 1993; Feinberg 2011). In semantic terms, the emergence of new understanding causes disambiguation of the lower levels. Each lower-level component is forced to settle onto one interpretation of each of its direct inputs from within the same system (i.e. within its immediate context of components at the same level interacting with it to form the higher-level system). Thus univocal meanings (e.g. propositional or binary) can arise, but only within a single level: the relationship between two directly connected elements must admit of only one interpretation of the producer/transmitter by the consumer/receiver, which can happen if these together form (part of) a higher level system that has adopted a stable conformation or Gestalt. That higher level state still admits however of different interpretations by the modules connected to it at its own level. The endless play of meaning only occurs if no higher level state arises which has only a single stable equilibrium state, or if stimulus analysis by further integration of representations into an ever-enlarging system never halts.

Discussion

Theory Spaces

The three Dimensions I have described may be seen as delineating a 3-dimensional theory space (Figure 6). This way of organising the numerous extant ideas is intended as a visual heuristic for understanding the complexities and variations on ‘psychosemantics’. I certainly do not wish to claim three is the definitive number, or to exclude the addition of further dimensions in the future to extend the space of possibilities sketched here.

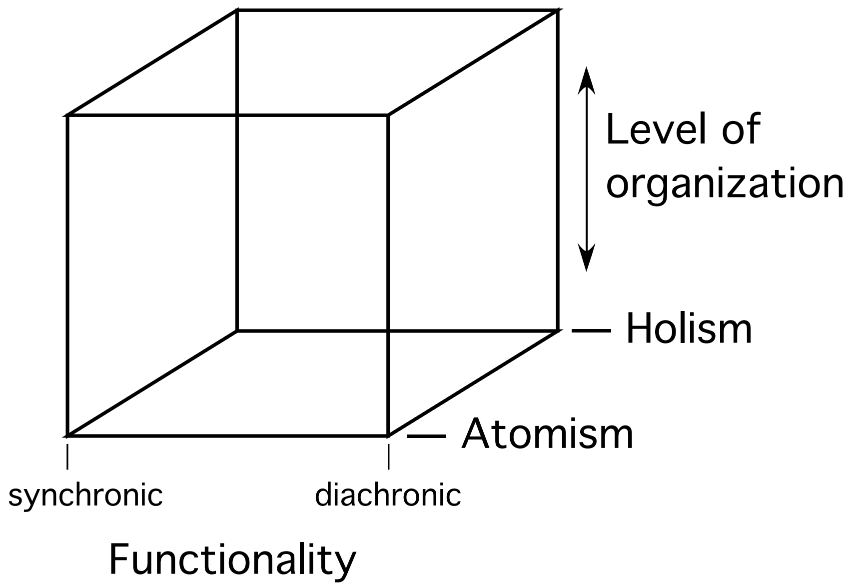

Fig. 6 Three-dimensional state-space for psychosemantics.

The Dimension 2 and 3 theories are mutually compatible (e.g. Lycan & Neander 2008), but given that these were developed in reaction to the shortcomings of the Dimension 1 approach (e.g. Millikan 1984), do they not replace any such understanding? Should Dimension 1 be scrapped? I have here retained the ‘traditional’ truth-based approach to meaning as one axis of the theory space for two reasons. Firstly, its hybrid view that there are two types of representation with different grounding in the environment (direct or basic versus indirect, derived, or inferential) is still valid within the Dimension 2 approach (e.g. Millikan 1984, ch. 19; Dretske 1988; Sterelny 1990; Neander 1995; 1999; Garfield 2000). This is even more so within Dimension 3, which posits a whole series of levels of meaningful representation with varying degrees of extension into the environment (where ‘the environment’ is relative to the size and level of the representing system under consideration).

Secondly, although alternatives to traditional representationalism have been proposed, recent thinking within analytic philosophy and cognitive science has supported the compatibility of all approaches rather than replacement of the old by the new. The truth value of inner representations (where truth is defined as exact correspondence, however arrived at, between a representation and some aspect of an objective external reality) is not so easily eliminated. For instance, even abstract concepts such as numbers must have real referents, else the demonstrable usefulness of those concepts would be a miracle (Jenkins 2008). In particular, theories of embedded or embodied perception-action are often presented as antithetical to the (Dimension 1) representational view postulating arbitrary symbols intervening between the sensory input and motor output (for a useful review see Shapiro 2011). Note however that representationalism does not depend upon the arbitrariness of the symbols (although the linguistic analogy within holism may imply it). Isomorphic analogue representations may ground the symbol level (Barsalou 2008; de Vega et al. 2008), as may spatially generated 'image schemata' (Lakoff 1987; Hampe 2005). I agree with Clark (2008) and Shapiro (2011) that such embodied grounding is fully compatible with symbolic representations-and-rules models of cognitive architecture and psychosemantics. Thus, it is possible to see embedded theories as merely an obvious variant of the holistic view which has been extended to include both mind and environment, which together form a unified system (e.g. Clark 2008; Gangopadhyay & Kiverstein 2009; Menary 2010; but cf. Fisher 2007; Rupert 2010). Recurrent feedback exists between both: stimuli affect our actions and our actions alter both the immediate sensory input (reafference) and the state of the world.

Internal Meaning

This view is consistent with the traditional idea that meaning comes ultimately from personal interaction with the world; but how can it be argued that meaning can be internal too (Lettvin et al. 1959; Dennett 1969; Harris 1997; Millikan 2004; Rose 2006; Breidbach 2007; Mandik 2008)? The answer is best approached via multi-levellism.

First, remember that the various parts of the nervous system self-organise during development and learning to form functional units — systems, assemblies, ensembles, groups, modules, etc. (e.g. Simon 1962; Edelman 1987; Sporns 2011). This occurs at multiple scales to generate a nested hierarchy (Figure 4). The almost ubiquitous phenomenon of neural plasticity and learning means that connected units not only affect each other instantaneously but also over longer time-scales, mutually altering one another's very structures, and thus their inbuilt 'knowledge bases', with potentially life-long effects. So if one asks "what does a module know" (or any other internal part of a person, even a single neurone) then the intentionality of its representational contents can be seen to be *directly* relative to its immediate environment at the same level, and *indirectly* related to further connected systems (which at some point connect onto the extra-personal world).

The developmental process ensures that the meaning of a system's activation is to subserve the teleological purpose of that system's existence as a system within a higher level system. Because these systems have thus co-evolved, the meaning of the generated representation is taken by the consuming mechanisms (the linked systems) to be its proper function, i.e. what it is supposed to be 'saying'. This is what it means for the consumers — at any level. This avoids any privileging of human-world interaction as the sole basis of meaning; any system that interacts with other systems, with which it has evolved in interaction, has meaning that is derived from (or implicit within) those current interactions. The meaning of 'meaning' is relative to level.

How does this answer the question of how the firing of a bug detector apparently has many meanings (Dimension 3)? A particular mechanism or system is only a bug detector within a certain level of organisation, where it connects (let us suppose, directly) on to a 'bug catcher' module, for which the bug detector has developed, and thus to which the detector's firing means 'bug present now' (see Dimension 2). Within the bug detector is a (lower-level) mechanism for detecting small black spots of light on the retina (perhaps, an 'off' centre-surround bipolar cell), and that mechanism's activation means 'dark spot on retina' to its receiving mechanisms (e.g. ganglion cells, tectal cells and so on). Within each of these mechanisms are those generating action potentials, and so on down the levels. But a single retinal action potential does not mean 'bug', nor does activity in the spot detector. These activities may over time become progressively integrated into higher level systems, including a bug detector, food detector, survival promoter, and so on as the levels are ascended. Above the level where the bug detector exists, 'the stimulus' is no longer the bug alone, but includes the bug's surrounding scene or context, which has widened over spatial and temporal (memory, prediction) horizons to form a scenario and narrative. Each of the additional meanings however only arises if and when each relevant new level emerges, and only applies within its own level.

Put another way: what uses, consumes, receives, decodes or 'reads' a mental representation? Not the whole person, not the Cartesian mind, not the ego and not the 'self'. A complete person may consume an external stimulus or sign, but at sub-personal scales one has to assume that functional modules, networks, neurones, receptor molecules and so on do the 'reading' (depending on level). Each of these elements has only partial knowledge and capabilities, compared to a whole person, so the depth of semantic analysis each can undertake on its own is limited (Attneave 1961; Dennett 1978; Rose 2006, ch. 4). If two elements or modules receive in parallel from the same representation, they interpret it differently, since they have different internal knowledge and powers of 'inference'. So, it 'means' different things to each of them — just as two different people can interpret the same external object differently. The more elements that can work together, the more profound their collective 'understanding' can become. When the entire set of relevant mental/neural elements is engaged, the system is at its maximal competence and performs as well as a complete mind. There is no difference in kind, no dualism between intra- and extra-personal meaning. Any distinction between types of representation is just a matter of degree rather than any duality in intrinsic quality. Or rather, there is a multiplicity of qualitative, Gestalt-like jumps; yet the principles by which these arise do not differ between sub-personal, personal and supra-personal scales. A society has its emerged Zeitgeist, consensus, mores, morals, values and linguistic (nested hierarchies of) systems which constrain the thoughts and percepts as well as the behaviour of the lower level units (i.e. us) that form the society, just as our minds constrain the activity of our neurons and our neurons constrain the molecules that form them. Similarly, at different levels the nature of the interaction changes. At low levels it may be described as physical (e.g. electrostatic forces between transmitter and receptor molecules), then as we ascend: informational, then symbolic, then cognitive, then cultural — from mechanical causation to mapping, decoding, associating, inferring, informing, conversing, persuading, inspiring, etc. There is no simple dichotomy between hard-wired, mechanical cause-and-effect, and cognitive, mental reasoning and experiencing; nor is there one between such intrapersonal thinking and interpersonal social interacting.

Perception and Meaning

I could have started this paper with the question "What is perception?" — but there are limits to what can be covered in a single article, so I have so far approached the question of how meaning and perception are related from one direction only. However, both meaning and experience (qualia, phenomenality) are supposed to relate to reality, at least sufficiently accurately on sufficiently frequent occasions. Hence there may be commonalities of principle — and of neural mechanism — underlying both; and indeed many have proposed their intimate relationship or co-occurrence (e.g. Lycan 2001; Pinna & Reeves 2009; Rowlands 2010); but see also Rose (2006, 361-375). If this overall thesis is valid, our understanding of perception should thus come to converge with our understanding of meaning, at least in part.

For instance, can the multi-dimensional analysis I have sketched here apply to perception as well as to meaning? That experience is multi-levelled (Dimension 3) is eminently reasonable (e.g. Pinna & Reeves 2009); and a nested hierarchy of perception-like internal and external scanning or monitoring mechanisms, teleologically developed (Dimension 2), has also been proposed (Lycan 1987; 1996); indeed several neural models of consciousness can be interpreted in this way (Rose 2006). Finally, the atomism-holism Dimension 1 is present in the distinction between the empiricist notion of perception as construction from simple elements or qualia versus the progressive differentiation of a fundamentally holistic phenomenology. This distinction has been made for perceptual development and learning (e.g. Gibson & Gibson 1955; Dennett 1988), but also for immediate ongoing processing, where stimuli may be conceived either as creating novel experiences from nothing, or as disturbing or focussing an ongoing phenomenal 'field' (e.g. Searle 2000). A modern hybrid theory however combines both ideas, in that afferent sweeps of neural activity are postulated to underlie different aspects of perception from those of recurrent, cyclic or ongoing processing in an internal network of representations (gist versus detail; bottom-up versus top-down attention; mismatches versus matches to expectations, stored knowledge, aims or needs), depending on contextual task demands. The initial feedforward sweep is thought to underlie only preconscious processes, whereas subsequent recurrence and phenomenal states may evolve progressively over time (e.g. Lamme & Roelfsema 2000; Hochstein & Ahissar 2002; Rose 2006, ch. 9; Epshtein et al. 2008; Gregory 2009; de Graaf et al. 2010; Koivisto et al. 2011; McManus et al. 2011; Schmidt et al. 2011).

Conclusions

Meaning can be ascribed to sub-personal activity, which includes neural firing, percepts and thoughts. This opens the possibility of diffusing the homunculus fallacy and building stepping-stones across the explanatory gap. Perceptual meaning is not simple or single, and any external stimulus or internal representation of a stimulus can have many meanings. New meanings or understandings emerge progressively as new systems are created at higher levels by continual integration of stored knowledge into the analysis of the stimulus.

Summary

To understand the meaning of 'meaning' and whether the concept can be applied at sub-personal levels, its nature must first be elucidated. Full understanding of psychosemantics requires a three-dimensional state-space of possible explanations, with axes (1) atomism-holism, (2) synchronic-diachronic functionality, and (3) level of analysis. Atomism proposes that meaning depends on causal links with the world, whereas under holism, meaning is relative to the current context or system of representations. These can be combined in a hybrid atomist-holist theory which contains elements of both. To solve the problems of mis-representation however the second dimension is needed. Diachronic teleofunctionalism links meaning to what has worked in the past, and synchronic functionalism to what works in the current system or context. Which is more relevant depends on which question you ask: evolution and ontogeny can tell you a mechanism's activity is supposed to mean for example 'bug', while whether it actually means bug now depends on whether there is a bug in the current environment. So again, both are relevant. However to understand fully the problem of indeterminacy the third dimension must be introduced. Multi-levellism allows multiple layers of meaning, both within and between persons. The function of any representation operates over larger spatio-temporal scales at higher levels of system organisation and integration. Thus 'meaning'

can be ascribed equally validly to internal acts of communication within the nervous system as to interactions with external objects or people, and between groups of people.
Keywords: Psychosemantics, meaning, levels of organisation, emergence.

Zusammenfassung

Um die Bedeutung von „Bedeutung" zu verstehen und zu entscheiden ob das Konzept auf die intrapersonelle Ebene angewendet werden kann, müssen wir zuerst seine Eigenschaften erläutern. Umfassendes Verständnis der Psychosemantik benötigt einen dreidimensionalen Zustandsraum für mögliche Erklärungen mit den folgenden Achsen (1) atomar – ganzheitlich, (2) synchrone – diachronische Funktionalität, und (3) Ebene der Analyse. Das Atomare nimmt an, dass „Bedeutung" von der kausalen Verbindung mit der Welt abhängt, wohingegen Ganzheitlichkeit „Bedeutung" in Bezug zum gegenwärtigen Kontext oder System der Repräsentation erklärt. Diese Konzepte können in einer hybrid atomar-ganzheitlichen Theorie kombiniert werden. Eine zweite Dimension ist jedoch nötig, um das Problem der Fehlrepräsentation zu lösen. Diachronischer Teleofunktionalismus verbindet „Bedeutung" mit dem, was in der Vergangenheit funktioniert hat, wohingegen synchronischer Funktionalismus die Bedeutung mit dem jetzigen System und Kontext in Verbindung bringt. Welche dieser Erklärungen relevanter ist hängt von der Fragestellung ab: Evolution und Ontogenie können die beabsichtigte Bedeutung der Aktivität des Mechanismusses erklären, zum Beispiel „Käfer", wobei die Frage, ob die Aktivität jetzt wirklich Käfer bedeutet davon abhängt, ob ein Käfer in der momentanen Umgebung existiert. Wieder sind beide Pole wichtig. Um jedoch das Problem der Undeterminiertheit vollständig zu verstehen muss eine dritte Dimension eingeführt werden. Vielschichtigkeiten erlauben viele Ebenen von Bedeutungen, sowohl intra- als auch interpersonell. Die Funktion jeder Repräsentation operiert über größere räumlich-zeitliche Distanzen auf höheren Ebenen der Systemorganisation und -integration. Demzufolge kann „Bedeutung" genauso interner Kommunikation innerhalb des Nervensystems, der Interaktion mit Gegenständen und Personen, und zwischen Gruppen von Personen zugeschrieben werden.
Schlüsselwörter: Psychosemantik, Bedeutung, Ebenen der Organisation, Auftauchen.

References

Albertazzi, L. (2006): *Immanent Realism: An introduction to Brentano.* Dordrecht: Springer.

Albertazzi, L. (2007): At the roots of consciousness. Intentional presentations. *Journal of Consciousness Studies 14* (1-2), 94-114.

Amundson, R. & Lauder, G.V. (1994): Function without purpose: the uses of causal role function in evolutionary biology. *Biology and Philosophy 9*, 443-469.

Attneave, F. (1961): In defense of homunculi, in Rosenblith, W.A. (ed.) (1961): *Sensory Communication*, 777-782. Cambridge, MA: MIT Press.

Barsalou, L.W. (2008): Grounded cognition. *Annual Review of Psychology 59*, 617-645.

Bechtel, W. (2008): *Mental Mechanisms. Philosophical perspectives on cognitive neuroscience.* New York: Routledge.

Bowers, J.S. (2010): More on grandmother cells and the biological plausibility of PDP models of cognition: a reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychological Review 117*, 300-308.

Breidbach, O. (2007): Neurosemantics, neurons and system theory. *Theory in Biosciences 126*, 23-33.

Brentano, F.C. (1874): *Psychologie vom empirischen Standpunkt, Band 1.* Leipzig: Duncker und Helmblot.

Buller, D.J. (1998): Etiological theories of function: a geographical survey. *Biology and Philosophy 13*, 505-527.

Buller, D.J. (2005): *Adapting Minds. Evolutionary psychology and the persistent quest for human nature*. Cambridge, MA: MIT Press.
Burkitt, I. (1991): *Social Selves: theories of the social formation of personality.* London: Sage.
Cacioppo, J.T. & Decety, J. (2011): Social neuroscience: challenges and opportunities in the study of complex behavior. *Annals of the New York Academy of Sciences 1224*, 162-173.
Callebaut, W. & Rasskin-Gutman, D. (eds.) (2005): *Modularity. Understanding the development and evolution of natural complex systems*. Cambridge, MA: MIT Press.
Carnap, R. (1928): *Scheinprobleme in der Philosophie. Das Fremdpsychische und der Realismusstreit.* Berlin-Schlachtensee: Weltkreis Verlag.
Clapin, H., Staines, P. & Slezak, P. (eds.) (2004): *Representation in Mind. New approaches to mental representation*. Amsterdam: Elsevier.
Clark, A. (1993): *Sensory Qualities*. Oxford: Oxford University Press.
Clark, A. (2008): *Supersizing the Mind. Embodiment, action and cognitive extension*. Oxford: Oxford University Press.
Clarke, A., Taylor, K.I. & Tyler, L.K. (2011a): The evolution of meaning: spatio-temporal dynamics of visual object recognition. *Journal of Cognitive Neuroscience 23,* 1887-1899.
Clarke, A., Taylor, K.I., Deveurex, B., Randall, B. & Tyler, L.K. (2011b): From perception to conception: how the meaning of visual objects emerges over time. *Perception 40 (supplement),* 32.
Cohen, H. & Lefebvre, C. (eds.) (2005): *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.
Coltheart, V. (1999): Modularity and cognition. *Trends in Cognitive Science 3*, 115-120.
Cooley, C.H. (1902): *Human Nature and the Social Order.* New York: Scribner.
Craver, C.F. (2007): *Explaining the Brain. Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
Crick, F. (1994): *The Astonishing Hypothesis: the scientific search for the soul*. London: Simon and Schuster.
Crick, F. & Koch, C. (2003): A framework for consciousness. *Nature Neuroscience 6*, 119-126.
Cummins, R. (1975): Functional analysis. *Journal of Philosophy 72*, 741-765.
Davies, P.S. (2001): *Norms of Nature. Naturalism and the nature of functions.* Cambridge, MA: MIT Press.
de Graaf, T.A., Roebroeck, A., Goebel, R. & Sack, A.T. (2010): Brain network dynamics underlying visuospatial judgement: a fMRI connectivity study. *Journal of Cognitive Neuroscience 22*, 2012-2026.
Dennett, D.C. (1969): *Content and Consciousness*. London: Routlege & Kegan Paul.
Dennett, D.C. (1978): *Brainstorms. Philosophical essays on mind and psychology*. Montgomery, VT: Bradford Press.
Dennett, D.C. (1987): *The Intentional Stance*. Cambridge, MA: MIT Press.
Dennett, D.C. (1988): Quining qualia, in Marcel, A.J. & Bisiach, E. (eds.) (1988) *Consciousness in Contemporary Science*, 42–77. Oxford: Oxford University Press.
Dennett, D.C. (1991): *Consciousness Explained*. Boston, MA: Little, Brown & Co.
Dennett, D.C. (1995): *Darwin's Dangerous Idea. Evolution and the meanings of life*. London: Allen Lane.
Dennett, D.C. (2011): Homunculi rule: reflections on Darwinian populations and natural selection by Peter Godfrey Smith. *Biology and Philosophy 26*, 475-488.
Derrida, J. (1967): *Voix et le Phénomène. Introduction au problème du signe dans la phénoménologie de Husserl*. Paris: Presses Universitaires de France.
de Vega, M., Glenberg, A. & Graesser, A. (eds.) (2008): *Symbols and Embodiment. Debates on meaning and cognition.* Oxford: Oxford University Press.
Dobson, V.G. & Rose, D. (1985): Models and metaphysics: the nature of explanation revisited, in Rose, D. & Dobson, V.G. (eds.) (1985): *Models of the Visual Cortex,* 22-36. Chichester: J. Wiley & Sons.
Dretske, F.I. (1988): *Explaining Behavior. Reasons in a world of causes*. Cambridge, MA: MIT Press.
Edelman, G.M. (1987): *Neural Darwinism. The theory of neuronal group selection.* New York: Basic Books.
Epshtein, B., Lifshitz, I. & Ullman, S. (2008): Image interpretation by a single bottom-up top-down cycle. *Proceedings of the National Academy of Sciences of the USA 105*, 14298-14303.
Fazekas, P. & Kertész, G. (2011): Causation at different levels: tracking the commitments of mechanistic explanations. *Biology and Philosophy 26,* 365-383.
Feinberg, T.E. (2011): The nested neural hierarchy and the self. *Consciousness and Cognition 20,* 4-15.
Fireman, G.D., McVay, T.E. and Flanagan, O.J. (eds.) (2003): *Narrative and Consciousness: literature, psychology, and the brain.* New York: Oxford University Press.
Fisher, J.C. (2007): Why nothing mental is just in the head. *Noûs 41*, 318-334.
Fodor, J.A. (1975): *The Language of Thought*. Hassocks: Harvester Press.

Fodor, J.A. (1983): *Modularity of Mind. An essay on faculty psychology*. Cambridge, MA: MIT Press.
Fodor, J.A. (1987): *Psychosemantics. The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.
Fodor, J.A. (1994a): Fodor, Jerry A., in Guttenplan, S. (ed.) (1994): *A Companion to the Philosophy of Mind,* 292-300. Oxford: Blackwell.
Fodor, J.A. (1994b): *The Elm and the Expert*. Cambridge, MA: MIT Press.
Fodor, J.A. (1998): *Concepts. Where cognitive science went wrong*. Oxford: Oxford University Press.
Fodor, J.A. (2000): *The Mind Doesn't Work That Way. The scope and limits of computational psychology*. Cambridge, MA: MIT Pres.
Fodor, J.A. & LePore, E. (1992): *Holism. A shopper's guide*. Oxford: Blackwell.
Gangopadhyay, N. & Kiverstein, J. (2009): Enactivism and the unity of perception and action. *Topoi 28*, 63-73.
Garfield, J.L. (2000): The meanings of "meaning" and "meaning": dimensions of the sciences of mind. *Philosophical Psychology 13*, 421-440.
Garson, J. (2011): Selected effects and causal role functions in the brain: the case for an etiological approach to neuroscience. *Biology and Philosophy 26,* 547-565.
Gibson, J.J. & Gibson, E.J. (1955): Perceptual learning: differentiation or enrichment? *Psychological Review 62*, 32–41.
Gould, S.J. & Lewontin, R.C. (1979): The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B 205*, 581-598.
Gregory, R.L. (2009): *Seeing Through Illusions.* Oxford: Oxford University Press.
Hampe, B. (ed.) (2005): *From Perception to Meaning. Image schemas in cognitive linguistics.* Berlin: Mouton de Gruyter.
Harnad, S. (1990): The symbol grounding problem. *Physica D 42*, 335-346.
Harris, L. (1997): The coding of self motion, in Jenkin, M. & Harris, L. (eds.) (1997): *Computational and Psychophysical Mechanisms of Visual Coding,* 157-183. Cambridge: Cambridge University Press.
Hebb, D.O. (1949): *The Organization of Behavior*. New York: Wiley.
Hochstein, S. & Ahissar, M. (2002): View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron 36*, 791–804.
Jacob, P. & Jeannerod, M. (2003): *Ways of Seeing. The scope and limits of visual cognition*. Oxford: Oxford University Press.
James, W. (1890): *The Principles of Psychology*. New York: Holt.
Jenkins, C.S. (2008): *Grounding Concepts. An empirical basis for arithmetical knowledge*. Oxford: Oxford University Press.
Kauffman, S.A. (1993): *The Origins of Order. Self organization and selection in evolution*. New York: Oxford University Press.
Kim, J. (1998): *Mind in a Physical World.* Cambridge, MA: MIT Press.
Kingsbury, J. (2008): Learning and selection. *Biology and Philosophy 23*, 493-507.
Koivisto, M., Railo, H., Revonsuo, A., Vanni, S. & Salminen-Vaparanta, N. (2011): Recurrent processing in V1/V2 contributes to categorization of natural scenes. *Journal of Neuroscience 31*, 2488-2492.
Lakoff, G. (1987): *Women, Fire, and Dangerous Things. What categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, G. & Johnson, M. (1980): *Metaphors We Live By.* Chicago: University of Chicago Press.
Lamme, V.A.F. (2004): Separate neural definitions of visual consciousness and visual attention: a case for phenomenal awareness. *Neural Networks 17*, 861-872.
Lamme, V.A.F. & Roelfsema, P.R. (2000): The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience 23*, 571-579.
Lettvin, J.Y., Maturana, H.R., McCulloch, W.S. & Pitts, W.H. (1959): What the frog's eye tells the frog's brain. *Proceedings of the IRE 47*, 1940-1951.
Lorenz, D.M., Jeng, A. & Deem, M.W. (2011): The emergence of modularity in biological systems. *Physics of Life Reviews 8*, 129-160.
Lycan, W.G. (1987): *Consciousness*. Cambridge, MA: MIT Press,.
Lycan, W.G. (1996): *Consciousness and Experience*. Cambridge, MA: MIT Press,.
Lycan, W.G. (2001): The case for phenomenal externalism, in Tomberlin, J.E. (ed.) (2001): *Philosophical Perspectives, 15, Metaphysics,* 17-35. Malden: Blackwell.
Lycan, W.G. & Neander, K. (2008): Teleofunctionalism. *Scholarpedia 3 (7)*, 5358. *http://www.scholarpedia.org/article/Teleofunctionalism* (30.01.2011).
Macdonald, G. & Papineau, D. (eds.) (2006): *Teleosemantics*. Oxford: Oxford University Press.

Mackay, D. (1980): The interdependence of mind and brain. *Neuroscience 5*, 1389–1391.

Mandik, P. (2008): Neurosemantics bibliography. *http://www.petemandik.com/blog/neurosemantics-bibliography* (31.3.2011).

Marcus, G.F. (2001): *The Algebraic Mind. Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

McManus, J.N.J., Li, W. & Gilbert, C.D. (2011): Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences of the USA 108*, 9739-9746.

Mead, G.H. (1934): *Mind, Self and Society from the Standpoint of a Social Behaviorist.* Chicago: University of Chicago Press.

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

Merleau-Ponty, M. (1945): *Phénoménologie de la Perception*. Paris: Gallimard.

Millikan, R.G. (1984): *Language, Thought, and Other Biological Categories. New foundations for realism*. Cambridge, MA: MIT Press.

Millikan, R.G. (1993): *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Millikan, R.G. (2000): *On Clear and Confused Ideas. An essay on substance concepts*. Cambridge: Cambridge University Press.

Millikan, R.G. (2004): *Varieties of Meaning*. Cambridge, MA: MIT Press.

Millikan, R.G. (2005): Why (most) concepts aren't categories, in Cohen, H. & Lefebvre, C. (eds.) (2005): *Handbook of Categorization in Cognitive Science*, 305-315. Amsterdam: Elsevier.

Millikan, R.G. (2006): Useless content, in Macdonald, G. & Papineau, D. (eds.) (2006): *Teleosemantics*, 100-114. Oxford: Oxford University Press.

Neander, K. (1995): Misrepresenting and malfunctioning. *Philosophical Studies 79*, 109-141.

Neander, K. (1999): Fitness and the fate of unicorns, in Hardcastle, V.G. (ed.) (1999): *Where Biology Meets Psychology, 3-26*. Cambridge, MA: MIT Press.

Neander, K. (2006): Content for cognitive science, in Macdonald, G. & Papineau, D. (eds.) (2006): *Teleosemantics*, 167-194. Oxford: Oxford University Press.

Neander, K. (2008): Teleological theories of mental content: can Darwin solve the problem of intentionality? in Ruse, M. (ed.) (2008): *The Oxford Handbook of Philosophy of Biology*, 381-409. Oxford: Oxford University Press.

Nielsen, K.S. (2010): Representation and dynamics. *Philosophical Psychology 23*, 759-773.

Ochsner, K.N. & Lieberman, M.D. (2001): The emergence of social cognitive neuroscience. *American Psychologist 56*, 717–734.

Ogden, C.K. & Richards, I.A. (1923): *The Meaning of Meaning. A study of the influence of language upon thought and of the science of symbolism.* London: Routledge & Kegan Paul.

Papineau, D. (1987): *Reality and Representation*. Oxford: Blackwell.

Pinna, B. & Reeves, A. (2009): From perception to art: how vision creates meanings. *Spatial Vision 22*, 225-272.

Poli, R. (2006): Levels of reality and the psychological stratum. *Revue Internationale de Philosophie 61*, 163-180.

Pollen, D.A. (1999): On the neural correlates of visual perception. *Cerebral Cortex 9*, 4-19.

Putnam, H. (1975): The meaning of 'meaning', in Gunderson, K. (ed.) (1975): *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science, vol. VII,* 131-193. Minneapolis: University of Minnesota Press.

Quian Quiroga, R., Kreiman, G., Koch, C. & Fried, I. (2008): Sparse but not 'Grandmother cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences 12*, 87-91.

Quine, W.V.O. (1951): Two dogmas of empiricism. *Philosophical Review 60*, 20-43.

Quine, W.V.O. (1960): *Word and Object*. Cambridge, MA: MIT Press.

Ramachandran, V.S. (1985): The neurobiology of perception. *Perception 14*, 97-103.

Reichenbach, H. (1928): *Philosophie der Raum-Zeit-Lehre.* Berlin: de Gruyter.

Revonsuo, A. (2006): *Inner Presence. Consciousness as a biological phenomenon.* Cambridge, MA: MIT Press.

Roelfsema, P.R. (2005): Elemental operations in vision. *Trends in Cognitive Sciences 9*, 226-233.

Rose, D. (1996): Some reflections on (or by?) grandmother cells. *Perception 25*, 881-886.

Rose, D. (1999a): The historical roots of the theories of local signs and labelled lines. *Perception 28*, 675-685.

Rose, D. (1999b): Creativity, intentionality and the conscious/unconscious distinction: a neural theory. *Journal of Intelligent Systems 9*, 407-443.

Rose, D. (2006): *Consciousness. Philosophical, Psychological and Neural Theories.* Oxford: Oxford University Press.

Rose, D. & Dobson, V.G. (1985): Methodological solutions for neuroscience, in Rose, D. & Dobson, V.G. (eds.) (1985): *Models of the Visual Cortex,* 533-545. Chichester: Wiley.

Rose, D. & Dobson, V.G. (1989): On the nature of theories and the generation of models of the circuitry of the primary visual cortex, in Kulikowski, J.J., Dickinson, C. & Murray, I.J. (eds.) (1989) *Seeing Contour and Colour*, 651-658. Oxford: Pergamon.

Rowlands, M. (2010): *New Science of the Mind. From extended mind to embodied phenomenology.* Cambridge, MA: MIT Press.

Rupert, R.D. (2010): *Cognitive Systems and the Extended Mind*. New York: Oxford University Press.

Salthe, S.N. (1985) *Evolving Hierarchical Systems. Their structure and representation.* New York: Columbia University Press.

Schaffer, J. (2007): The metaphysics of causation. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/causation-metaphysics (25.10.2011).

Schmidt, T., Haberkamp, A., Veltkamp, G.M., Weber, A., Seydell-Greenwald, A. & Schmidt, F. (2011): Visual processing in rapid-chase systems: image processing, attention, and awareness. *Frontiers in Psychology 2*: 169. doi: 10.3389/fpsyg.2011.00169.

Schroeder, T. (2004): New norms for teleosemantics, in Clapin, H., Staines, P. & Slezak, P. (eds.) (2004): *Representation in Mind: new approaches to mental representation*, 91-106. Elsevier, Amsterdam.

Scott, A. (1995) *Stairway to the Mind: The controversial new science of consciousness.* New York: Springer.

Searle, J.R. (2000): Consciousness. *Annual Review of Neuroscience 23*, 557–578.

Shapiro, L. (2011): *Embodied Cognition.* Abingdon: Routledge.

Shrager, J. & Langley, P. (eds.) (1990): *Computational Models of Scientific Discovery and Theory Formation.* San Mateo, CA: Kaufmann.

Simon, H.A. (1962): The architecture of complexity: hierarchic systems. *Proceedings of the American Philosophical Society 106,* 467–482.

Simon, H.A & Ando, A. (1961): Aggregation of variables in dynamic systems. *Econometrica 29*, 111-138.

Sporns, O. (2011): *Networks of the Brain.* Cambridge, MA: MIT Press.

Sterelny, K. (1990): *The Representational Theory of Mind. An introduction*. Oxford: Blackwell.

Wittgenstein, L.J.J. (1953): *Philosophical Investigations.* (Transl. Anscombe, G.E.M.) Oxford: Blackwell.

Wright, L. (1973): Functions. *Philosophical Review 82*, 139-168.

David Rose is a Visiting Reader in Psychology at the University of Surrey. He holds degrees in Psychology (BSc), Neurophysiology (PhD) and the History and Philosophy of Science (MPhil) from the Universities of Bristol and Cambridge. His empirical research has been mainly in visual psychophysics, particularly binocular vision, but also in several areas of neuroscience. His philosophical investigations centre on theories of consciousness.

Address: Department of Psychology, University of Surrey, Guildford, Surrey GU2 7XH, U.K.
E-Mail: d.rose@surrey.ac.uk