

Copyright © Imprint Academic 2002
Reproduction permitted for single personal use only
Unauthorised transmission, publication and web uploads prohibited

Liane Gabora

Amplifying Phenomenal Information

Toward a Fundamental Theory of Consciousness

Abstract: *Fundamental approaches bypass the problem of getting consciousness from non-conscious components by positing that consciousness is a universal primitive. For example, the double aspect theory of information holds that information has a phenomenal aspect. How then do you get from phenomenal information to human consciousness? This paper proposes that an entity is conscious to the extent it amplifies information, first by trapping and integrating it through closure, and second by maintaining dynamics at the edge of chaos through simultaneous processes of divergence and convergence. The origin of life through autocatalytic closure, and the origin of an interconnected worldview through conceptual closure, induced phase transitions in the degree to which information, and thus consciousness, is locally amplified. Divergence and convergence of cognitive information may involve phenomena observed in light e.g. focusing, interference, and resonance. By making information flow inward- biased, closure shields us from external consciousness; thus the paucity of consciousness may be an illusion.*

Keywords: abstract thought, amplification, closure, combination problem, concept, consciousness, context, double aspect theory, edge of chaos, episodic mind, focus, holographic memory, information, interference, light, modern mind, origin of life, panpsychism, phase transition, quantum mechanics, resonance, superposition, symbolic threshold.

Approaches to consciousness can be divided into two camps: (1) *reductionist approaches* that attempt to explain how consciousness could arise out of non-conscious components, and (2) *fundamentalist approaches* (for example, *panpsychism*)¹ that take consciousness to be a universal primitive, and attempt to explain why human consciousness has the distinctive qualities it has. The notion

Correspondence: Liane Gabora, Center Leo Apostel for Interdisciplinary Studies (CLEA), Free University of Brussels (VUB), Krijgskundestraat 33, Brussels, B1160, Belgium.
Email: lgabora@vub.ac.be

Journal of Consciousness Studies, 9, No. 8, 2002, pp. 3–29

that consciousness is fundamental has been with us for centuries, and it has been drawn on the basis of philosophical, religious, and scientific argument alike (e.g. Chalmers, 1996; Dyson, 1979; Feigl, 1958/1979; Foster, 1989; Ghose & Aurobindo, 1998; Griffin, 1998; Hartshorne, 1968; Lockwood, 1989; Montero, 2001; Nagel, 1979; Russell, 1926; Stoljar, 2001; Strawson, 2000; Whitehead, 1929). The fact that the same conclusion has been reached by influential thinkers from very different angles does not mean it is *right*, but does suggest that it should not be dismissed lightly. Rather than repeat arguments that have been presented elsewhere, this paper tackles some issues that must be faced once we accept the premise that the position is plausible.

Fundamental approaches bypass the problem of how complex an entity must be to be conscious, and what sort of complication *grants* it consciousness, by positing that consciousness — or at least a primitive form of it, perhaps quite unlike that with which we are familiar — is in the building blocks from the lowest level. But this opens up other problems. One is the *combination problem* (Seager, 1995): how do you get from the sort of consciousness that could be present in rocks and trees and everything else to the real McCoy, human consciousness? Another is that they strike most people as counterintuitive. Rocks and trees and so forth don't *seem* conscious; what reason is there to believe they *are*? These are the challenges responded to in this paper.

I: Is the Combination Problem Solvable in Principle?

Let us begin with the first challenge. If consciousness were ubiquitous, one might expect it to be uniformly distributed. Why then do there appear to be gradations of it, such that some entities are more conscious than others (or the same entity more conscious in some states than others)? A related question is: how does primitive consciousness get combined into more interesting forms of consciousness?

1. *The double aspect theory of information*

A relatively straightforward way of dealing with this falls out of Chalmers' (1996) concept of *naturalistic dualism*, which posits that phenomenal experience or consciousness is a fundamental feature of the universe like mass or charge, not logically supervenient on the physical facts, but entirely compatible with them. One way this could work is whenever an information state is realized physically it is also realized phenomenally. This idea that information has a conscious aspect he refers to as the *double aspect theory of information*. Chalmers uses the

[1] Chalmers suggests the term *panprotopsychism* to refer to the idea is that everything possesses, not consciousness necessarily, but proto-phenomenal properties which collectively generate consciousness when organized into a certain kind of structure such as a brain. (I have not adopted this term here because the sort of consciousness that, say, an elementary particle or a thermometer *could* have is so different from our own that at this stage of our understanding it probably doesn't matter too much whether you call it conscious or in possession of proto-phenomenal properties. But in the future this distinction may be useful.)

term information in the Shannon (1948) sense as a difference that makes a difference to some interpretive system. Seager (1995) stresses that information must have semantic significance, and that it may be noncausal, i.e. signal a relation based in correlation rather than cause, and these points are incorporated into the present use of the term.

If we accept the double aspect theory of information as plausible, we need to determine what sort of architecture could coerce bits of information to combine their individual subjectivities into a single, concentrated subjectivity that maintains its integrity, and blocks out the subjectivity of everything else, such that the consciousness of non-self parts of the world appear relatively negligible by comparison. Chalmers suggests that this might be accomplished through the *amplification* of phenomenally-endowed information. Following up on this suggestion, it seems reasonable to propose that the degree to which an entity is conscious is a function of the degree to which it locally amplifies information. Although this is not an explicitly threshold-type proposal, but rather a gradualist one, some physical processes are posited to induce phase transitions in degree of amplification. The basic idea is that biological and cognitive systems accomplish this by trapping information through autocatalytic closure, and maintaining the dynamics at the edge of chaos through simultaneous processes of divergence and convergence. Others have suggested that consciousness is indeed fundamental, but that its non-uniform distribution is best explained in terms of some variable *other* than information such as randomness (see Goertzel, 1995) or energy (see Taylor, 2001). The basic scheme laid out here still largely follows through.

2. *Localized amplification of information*

If we are to posit that certain entities such as human brains are conscious in virtue of their capacity to amplify information, it is useful to begin by looking at how information gets amplified in simpler systems. The best-known amplification techniques are probably those that work with light, e.g. lasers (light amplification by stimulated emission of radiation). Information can be conveyed through the frequency and amplitude of a light wave, and through its timing or phase relations relative to other waves. However, light tends to travel forever onward in a straight line, whereas we are interested in how information could be amplified *locally* in entities that tend to stay more or less in the same place.

(1) *Amplifying Information by Trapping It* The light-amplifying architecture we consider is a sphere composed of a light-penetrating substance with a higher refractive index than air.² Light more readily enters the sphere (Figure 1a) than leaves it (Figure 1b). Since the surface is concave, light rays refract as they enter the sphere (Figure 1a), and reflect on the interior surface (Figures 1b), thereby becoming more focused. Only light that passes through the centre can avoid reflecting back into the sphere (Figures 1c and 1d). Therefore, light is trapped and thus locally amplified (Figure 1e) such that the direction of informing is

[2] This includes pretty much anything.

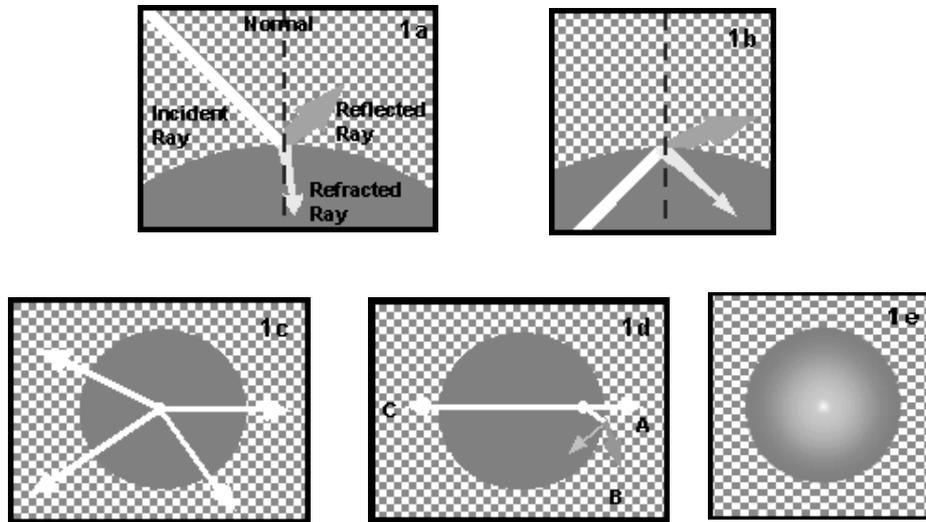


Figure 1. (a) The incident ray separates into a refracted ray and a reflected ray when it meets the surface. Because the refractive index of the sphere is higher than that of air, light approaching the sphere is more likely to refract and pass through it, than to reflect off. The refracted ray bends toward the normal, and since the surface is concave, it becomes more focused. (b) Since the refractive index of the sphere is higher than that of air, light approaching this interface from within the sphere is more likely to reflect than refract. Here the refracted ray bends away from the normal. Since the surface is concave, this reflecting light also becomes more *focused*. (c) Light in the centre of the sphere can radiate outward in any direction with no refraction and minimal reflection because wherever it contacts the sphere it is perpendicular to it. (d) Light near the surface can radiate through only a narrow region without reflection and refraction, when it hits the sphere at a perpendicular angle, as in rays A and C. If the angle is just a bit off, as in ray B, it reflects and refracts. (e) Because the effects in a-d are taking place simultaneously throughout the sphere, light that enters the sphere is, to a great extent, trapped.

nonsymmetrically inward-biased. This does not mean that information never flows out of the system, just that more flows in. The effect is even greater if the refractive index of the material the sphere is made of gradually increases from the periphery to the center of the sphere. In this case, light no longer travels in straight lines but in arcs, so it is even less likely to escape.

(2) *Edge of Chaos through Divergence and Convergence* Now we have a structure that locally traps light, but as it stands it is not very informative because the signal does not change. This brings us to the issue of *informability*. Some entities have greater potential to inform one another than others; for example, one may have a lot to say to one person and nothing to another, a key can open some doors but not others, an enzyme can catalyse some reactions but not others, and so forth. There are entities whose dynamics are *relevant* to one another, and others whose are not. To some extent, informability is enhanced through *similarity*. For example, people who speak the same language can convey more to one another than people who cannot. Yet to some extent, informability requires *difference*. If I knew everything you know, then even if we speak the same language, when you spoke to me you would not inform me.

It turns out that the information-carrying capacity of a system is highest when the degree to which its parts are correlated or causally connected falls within a narrow regime between order and chaos (Kauffman, 1993; Langton, 1992). To use an analogy that conveys intuitively why this is, think of a sequence of drum beats. Each beat provides information, but because a sequence of beats can be recoded as the single instruction ‘loop drum’, it isn’t very informative. On the scale from complete order to complete chaos, it lies at the extreme order end. On the other hand, if each instrument in a band plays notes randomly without regard to what preceded or what will follow, or to what the other instruments are doing, this can be recoded as ‘play anything’. Thus, at the extreme chaos end, things are no more informative. An intermediate degree of connectedness can arise and be maintained when each musician is partly focused on doing their own thing, and partly focused on being in harmony with the others. In general, for the dynamics of a system to be poised at the edge of chaos there must exist simultaneously processes of *divergence* (promoting differentiation) and processes of *convergence* (promoting sameness or similarity). The information carrying capacity is enhanced when differentiation and convergence are occurring at multiple hierarchical levels of resolution or time scales.

(3) *Resonance and Integration* Let us see how the simple light amplifying structure described above could maintain its dynamics at the edge of chaos through the divergence and convergence of information. First we ask: what causes a wave (such as a ray of light) to inform a physical object? It should first be noted that all objects are potential oscillators. The more closely the distribution of frequencies present in the wave match the natural frequency distribution of a potential oscillator (such as an electron bound in an atom), the more likely it will oscillate. If their frequency distributions match, they are said to *resonate*. Thus a signal can inform a potential oscillator by matching its frequency distribution with the resonant frequencies of the oscillator.

Now consider the situation where the resonant frequencies vary across the surface of the sphere, and the surroundings vary dynamically with respect to the intensity of these different frequencies (as in the natural world). A structure composed of differentiated parts can be *integrated* to function as a whole by ensuring that, for any given part, there are other parts that *inform* it, and other parts that are informed *by* it. Therefore in order for each region of the sphere to have the potential to inform and be informed, the variation in resonant frequencies diminishes as one penetrates deeper into the sphere. The focal point — the point where light converges — now shifts from one instant to the next, depending on the distribution of incoming light of different frequencies at a given instant. If a bird flies by, its shape and position change as it moves — thus there is change — yet even from another angle it is still the bird — thus there is sameness. The divergence and convergence of information in the exterior world is reflected in the interior of the sphere.

(4) *Superposition and Interference of Information Pathways* It is of course more interesting to consider entities wherein the forces promoting divergence

and convergence do not merely reflect the exterior world but arise in part through the internal dynamics of the entity itself. Actually, this is present in a very limited sense in the sphere, because the location and strength of the point of convergence at one instant are affected by where it converged at the previous instant, so it has a degree of built-in temporal self-similarity. But it would be better able to stay poised at the edge of chaos if it could tune its properties dynamically in response to its internal state and external situation, maximizing its potential to inform and be informed. This could be accomplished, for example, by slightly altering its shape such that light passed through at a different angle, thereby changing the location and manner in which incoming rays converge.

To see what effect this could have, it is necessary to look briefly at a phenomenon that occurs when light waves combine or are *superposed*. The new wave that results is generally more informative than its constituents (Figure 2).

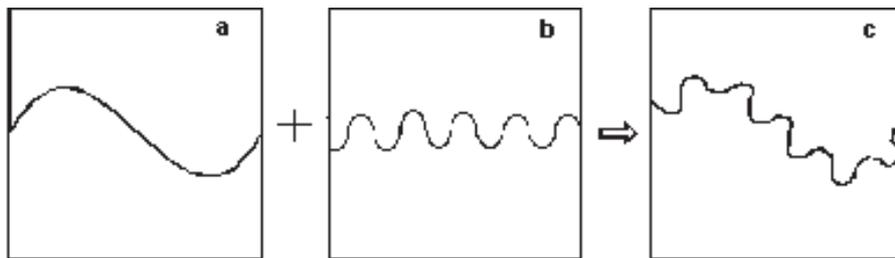


Figure 2. Wave **c** is the sum of the superposed waves **a** and **b**.

As a consequence of how waves superpose, when two waves are present in phase in the same medium at the same time, they undergo *interference*. *Constructive interference* occurs when they are aligned crest to crest and trough to trough, such that the resultant wave has the same frequency but is greater in amplitude (Figure 3).

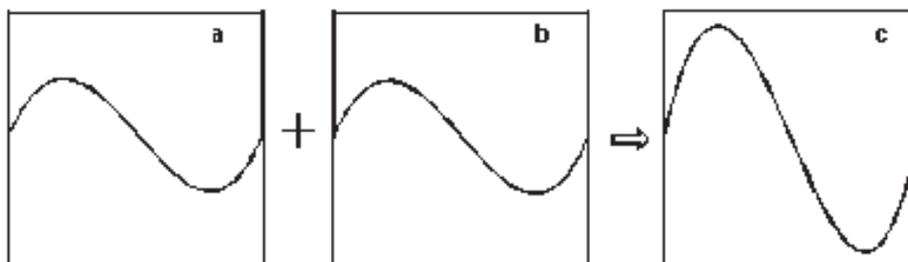


Figure 3. The amplitude of wave **c** is equal to the sum of its component waves **a** and **b** due to constructive interference.

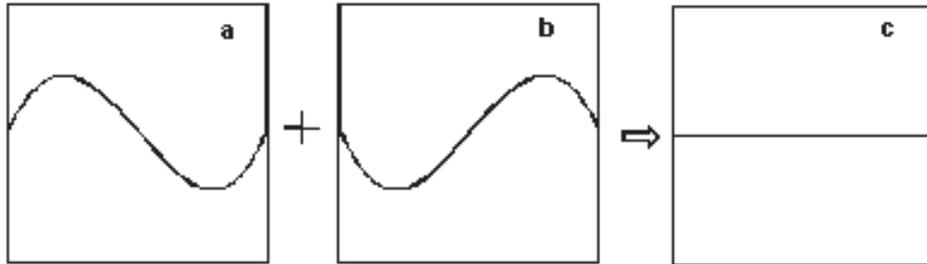


Figure 4. Waves **a** and **b** exhibit destructive interference such that the result is no wave at **c**.

Destructive interference occurs when the upward displacement of one wave cancels the downward displacement due to another (Figure 4)

Thus once light is trapped (such as through the spherical structure we have been looking at) the informing process can be maintained through superposition (including constructive or destructive interference) with its reflected self, or with other rays. Information is carried back and forth throughout the sphere through the dynamical pattern of phase relationships.

The structure we have been looking at is simple enough to illustrate clearly how information could be locally amplified, and how this amplified state could be maintained. If the double aspect theory of information holds true, it follows that phenomenality is heightened as well. Thus the combination problem is solvable in principle, and we have a means of addressing the question of how consciousness could be primitive and universal yet nevertheless graded.

II: Amplification of Organic Information

Now that we have seen how it is possible for information to be locally amplified in relatively simple systems, let us explore how this might occur in more complex systems such as living organisms, and finally in the structure we are most certain is conscious, the human mind. These systems are more complicated than the simple physical systems we have been looking at, but they use some of the same basic mechanisms to amplify information.

1. *Trapping and integrating information through autocatalytic closure*

A crucial step from simple physical structures to human minds was the origin of life. The origin of life presents a chicken-and-egg problem: if living things come into existence when other living things give birth to them, how did the first living thing arise? How could something able to reproduce itself come to be? Over the last century we have learned that self-replication is orchestrated by an intricate network of interactions between proteins and DNA; proteins are made by decoding DNA, and DNA requires the catalytic action of proteins to get decoded. So the question becomes: how could a system composed of complex, mutually-dependent parts come into existence?

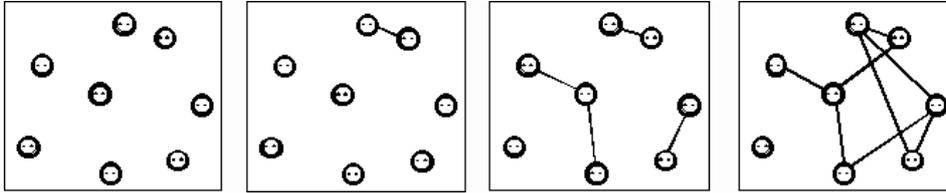


Figure 5. From left to right: (a) A set of loose buttons. (b) Tie two randomly chosen buttons together with a thread. (c) Repeat over and over. Occasionally lift a button and see how many connected buttons get lifted. (d) Increasingly large clusters emerge, and eventually reach a point where they form one giant cluster containing most of the buttons.

The solution proposed by Kauffman (1993), inspired by graph theorists Erdos and Renyi (1959; 1960), is that life began with the emergence of a set of autocatalytic polymers. Kauffman (1999) conveys this basic idea intuitively as follows (Figure 5). Spill some buttons on the floor. Tie two randomly chosen buttons together with a thread. Repeat this again and again. Every once in a while, pick up a button and see how many connected buttons get lifted. After a while, clusters emerge. The clusters get larger. Eventually they join together forming one giant cluster that contains most of the buttons. In fact, when the ratio of nodes (buttons) to edges (strings) reaches approximately 0.5, the probability that they form an interconnected cluster goes from extremely unlikely to almost inevitable. Thus whichever button you choose to pick up, you end up lifting the entire web of connected buttons.

Kauffman adapts this as follows. Say the nodes (buttons) are various different catalytic polymers such as those that would have been around at the time of the origin of life. Say the edges (strings) are catalysed reactions. When polymers interact, the number of different polymers increases exponentially. However, the number of reactions by which they can interconvert increases faster than their total number. Therefore, as their diversity increases, so does the probability that some subset of the total reaches a critical point where there is a catalytic pathway to every member. Just as with the buttons and strings, when the ratio of polymers to reactants reaches approximately 0.5, the percolation threshold, the probability that one giant cluster emerges goes from extremely unlikely to almost inevitable. Some subset of the total reaches a critical point where there is a catalytic pathway to every member. The set is *autocatalytically closed* because although none of the polymers can catalyse its own replication, each can catalyse the replication of some other polymer in the set, and likewise, its own replication is catalysed by some other member of the set.

Thus, as separate elements transform into a closure structure, they generate not just new information-processing components, but *exactly those* that will be exploited by what is already in place such that they collectively function as a whole. This is consistent with the notion that living systems are self-maintaining (Maturana & Varela, 1987) and entropy-defying (Prigogine & Stengers, 1984). Since the parts are specifically suited to inform *each other*, the direction of

information processing is nonsymmetrically inward-biased. Information is more likely to stay within the confines of the system rather than leak out; it is trapped, as it was in the sphere example. Thus, if the double aspect theory holds true, the emergence of autocatalytic structure at the organic level brings about a phase transition in degree of consciousness.

2. *The closure structure is poised at the edge of chaos*

Clearly, in order for autocatalytic closure to be achieved, each polymer should catalyse at least one reaction. However, if each polymer catalysed *every* reaction, they would quickly form an interconnected whole, but the system would be stagnant. We saw that the information-carrying capacity of a system is highest when the degree to which its parts are correlated or causally connected falls within a narrow regime between order and chaos. It turns out that many aspects of living systems fall squarely in this regime (Kauffman, 1993; Langton, 1992). In the origin of life scenario, the requisite intermediate degree of connectedness arises naturally as a consequence of the fact that the shapes and charges of polymers endows them with the ability to catalyse some reactions and not others; thus the distribution of reactions any given polymer can catalyse is approximately Gaussian. In a computer simulation, this was approximated by assigning each polymer a low *a priori* random probability P of catalysing each reaction (Farmer *et al.*, 1987).

3. *Maintaining the information-amplifying structure through replication*

How did the first autocatalytically closed structure self-replicate? It is commonly believed that the primitive self-replicating set was enclosed in a small volume (such as a coascervate or liposome) to permit the necessary concentration of reactions (Cemin & Smolin, 1997; Morowitz, 1992; Oparin, 1971). This of course furthers the effect of trapping and localizing information. Since each polymer is getting duplicated somewhere in the set, eventually multiple copies of all polymers exist. The abundance of new polymers exerts pressure on the vesicle walls. This often causes such vesicles to engage in *budding*, where part pinches off and the vesicle divides in two. So long as each contains at least one copy of each kind of polymer, the set can continue to self-replicate indefinitely. Replication is far from perfect, so an ‘offspring’ is unlikely to be identical to its ‘parent’. Different chance encounters of polymers, or differences in their relative concentrations, or the appearance of new ‘food’ polymers, could all result in different catalysts catalysing a given reaction, which in turn alters the set of reactions to be catalysed. So there is plenty of room for heritable variation. Because this structure is able to replicate itself with endless variety and increasing complexity, this process of locally amplifying information is self-perpetuating. By continuously generating variety — novel information — and selectively replicating variants that manage to survive as a unified whole for some time, evolving systems explore the space of possible information-amplifying architectures.

4. Convergence (increased fidelity of replication) via the genetic code

Let us look briefly at how other transitions in the structure of living organisms (see de Duve, 1995; Maynard-Smith & Szathmary, 1994, for reviews) would have enhanced the maintenance and proliferation of locally amplified information. We have seen that early life lacked explicit instructions for how to self-replicate, yet according to Kauffman's widely-accepted scenario, collectively its parts had all the instructions necessary. Kauffman goes on to describe how, given an autocatalytic set of RNA-like polymers, a replicator–interactor distinction could come about. Some 'offspring' might have a tendency to attach small molecules such as amino acids (the building blocks from which proteins are made) to their surfaces. Some of these attachments inhibit replication and are selected against, while others favour it and are selected for. That is, we have our first indication of a division of labour between the part that interacts with the environment (the proteins), and the part of the organism concerned with replication (in this case, an RNA-based code).

The advent of this code is significant not only because it enabled the processes involved in replication of the information amplifying structure to be carried out recursively and with greater precision, but also because acquired characteristics could no longer be passed on to the next generation. Prior to coded replication, there was nothing to prohibit the inheritance of acquired characteristics. A change to one polymer would still be present in one of the offspring after budding occurs, and this could cause other changes that have a significant effect on the lineage further downstream. Thus the genetic code promoted convergence of information by increasing the fidelity of replication. The effect was magnified with the evolution of DNA proofreading enzymes.

5. Divergence through sexual reproduction

Following the appearance of DNA, but prior to sexual reproduction, each replicant was so similar to the others that the degree to which they manifest *new* information was limited. This is also the case, of course, for present-day asexual organisms. Each living entity is a bit like one of the drum beats that could be described with the simple instruction 'loop drum'. The onset of sexual reproduction did for organic information what the skilled drummer does for a rhythmic beat: maintain the overall pattern while varying the specifics. Sexual reproduction thus enabled a much more thorough exploration of the space of possible information-amplifying entities.

6. Multicellularity, organs, nervous systems

The division of labour at the level of cells, organelles, and organs provided the means for coadaptation between informers and informees, and put them into closer contact, thus facilitating spatiotemporal consolidation of informative events. In the nervous system, differentiation is accompanied by convergence to spectacular effect. To return to Kauffman's analogy, one of the buttons becomes much more widely cross-connected than any of the others (Figure 6).

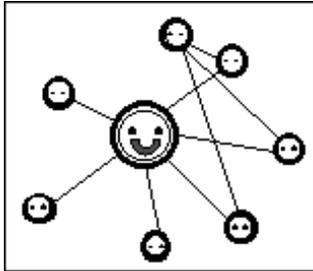


Figure 6.

Once autocatalytic closure has been established, further concentration of phenomenal information is achieved by funneling the majority of information-providing pathways into a region that constitutes a small fraction of the total system.

Sensory apparatuses increase the capacity to interiorize information from the outside world, while brains funnel these diverse inputs into a region much smaller than the system at large, thereby facilitating their integration.

III: Amplification of Cognitive Information

Nervous systems could not only coordinate experienced stimuli with appropriate responses but also store them as memories. Consider what would happen if these memories were to self-organize into an interconnected internal model of the world, or worldview. The ‘stepping stones’ between the memories would be concepts of varying levels of abstraction (i.e. cup, container, thing, etc.). Memories could then be ‘bounced back and forth’ throughout this interwoven web, re-defined and re-described (Karmiloff-Smith, 1992) according to how they relate to (or how they appear in the context of) one another, each evoking the next recursively in streams of abstract thought. Such an architecture yields endless possibilities for refining elements of the worldview in terms of their actual and possible inter-relations, and in the process, maintaining a stream of concentrated phenomenal information. It thereby also preserves dynamics at the edge of chaos through a combination of convergence (such as the realization that different experiences of cats are instantiations of the abstract concept ‘cat’) and divergence (such as the creative application of information from one domain to another). But this brings us to a paradox not unlike that presented by the origin of life: to generate concepts you have to be able to engage in a stream of associative thought, but a stream of associative thought will not get very far without concepts. How then do memories weave themselves into a relationally-structured model of the world?

The problem of how concepts — which are defined in terms of *other* concepts — came into existence, is often said (and particularly in the context of its implications for consciousness) to lead to a problem of infinite regress (e.g. Kurak, 2001). The problem is reminiscent of the problem of how the very first self-reproducing living organism came to exist. Again self-organized autocatalytic closure provides a tentative solution. That is, once brains came along, the stage was set for further trapping of phenomenal information through the emergence of a second autocatalytic system embedded in the first (Figure 7). In this section we look at this proposal in more detail.

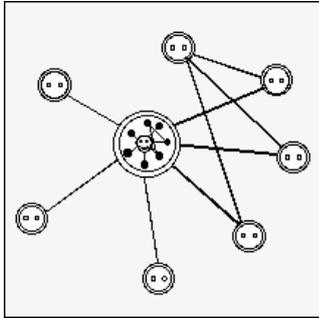


Figure 7.

The mind as a second autocatalytic system embedded in the first, further amplifying phenomenal information.

1. *Trapping and integrating information through conceptual closure*

Donald (1991) argues convincingly that before the arrival of *Homo erectus*, the human memory system was like that of a primate, largely limited to the storage and cued retrieval of specific episodes. Accordingly, he uses the term *episodic* to designate such a mind. With the first signs of culture approximately two million years ago, we see evidence of another kind of cognitive architecture. Because it can form abstract concepts, and thus engage (to some extent) in abstract, symbolic thought, it is referred to as the *abstract mind*. The transition to this kind of cognition has been referred to as the *symbolic threshold* (Deacon, 1997).

The application of graph theory to the origin of life problem has been adapted to yield a tentative solution to the origin of the cognitive architecture that gave birth to the evolution of culture. The basic idea is that the kind of cognitive architecture capable of sustaining cultural evolution may, like biological evolution, have originated in a phase transition to a self-organized web of catalytic relations between patterns, through a process of closure referred to as *conceptual closure*. The theory is not merely analogous to Kauffman's theory, but rather it posits a genuine new manifestation of a closure structure. The proposal has similarities to Goertzel's (1993a,b) application of graph theory to cognition. Again, since this is discussed at length elsewhere (Gabora, 1998; 2000; submitted), here I just outline the core idea and how it pertains to consciousness. Although this account focuses on integration of the worldview through the abstraction of deeper, more general concepts, the principles apply also to the integration of the psyche through the purification of intentions and emotions.

It must first be said that memories of episodes are *distributed* across multiple locations in assemblies of neurons, and can thereafter be evoked by stimuli that are similar (Hebb, 1949; Marr, 1969). According to the doctrine of *neural re-entrance*, the same memory locations get used and reused again and again, and this happens through a process of signaling back and forth along reciprocal connections (Edelman, 1987; 1992; Edelman & Tononi, 2000; Sporns *et al.*, 1989; 1991). Also, memory is *content addressable*; there is a systematic relationship between the state of an input and the place it gets stored. Thus memories and concepts stored in overlapping regions of conceptual space are correlated, or share features.

Here, memories of past experiences play the role of nodes (buttons), and the role of edges (strings) is played by associative pathways enabling one memory to evoke a reminding of another. The first step was the appearance of one or more minds with a tendency toward more widely distributed storage and retrieval of memories (which would likely have had a genetic basis). Given that the region activated and searched from at any instant is wider, and because the memory is content addressable, similar memories are stored in overlapping regions of conceptual space, and sometimes get retrieved simultaneously. Thus more memory locations both (1) participate in the etching of an experience into memory, and (2) provide ingredients for the *next* instant of experience. Much as catalysis increases the number of different polymers, which in turn increases the frequency of catalysis, reminding events increase the number of items stored in the mind by triggering the emergence of abstractions such as concepts (as well as attitudes, stories, theories, and so forth). The number of memories (buttons) increases as new experiences are had and new concepts formed. However, the number of associative pathways (threads) increases faster because concepts by their very nature have the potential to be associatively linked to many instantiations of them, as well as to other concepts. (For example, ‘container’ can be associated with different sorts of containers, as well as fluids and so forth that are kept in containers, as well as the materials they are made of, etc.) One can

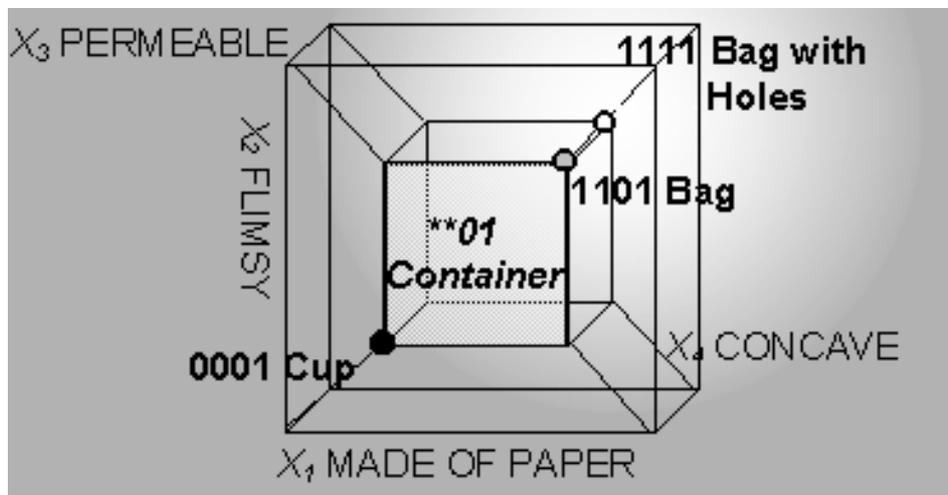


Figure 8. Four-dimensional hypercube that schematically represents a portion of conceptual space. The figure ignores some important details of concept representation (such that some features of a concept are more strongly activated than others) in order to illustrate the relationship between an abstraction and its instances. The features ‘MADE OF PAPER’, ‘FLIMSY’, ‘PERMEABLE’ and ‘CONCAVE’ lie on x_1 , x_2 , x_3 , and x_4 axes respectively. Dots represent centers of distributed hyperspheres where concepts are stored. Three concepts are stored here: ‘Cup’, ‘Bag’ and ‘Bag with Holes’. Black-ringed dots represent centres of distributed regions where they are stored. Fuzzy white region indicates portion of space activated by ‘Bag with Holes’. Note that activation falls off with distance from ‘Bag with Holes’, as described in section IV.2. Emergence of abstract concept ‘container’ implicit in the conceptual space (central white square) is possible if activation is distributed yet the degree of distribution is constrained.

express this by saying they cover more of conceptual space. This is illustrated very simplistically (in order just to get this basic point across) in Figure 8.

Reminding events *themselves* begin to evoke reminding events recursively, thus generating streams of associative thought, which increase in both duration and frequency. In the course of these streams of thought, more abstractions emerge, which themselves become connected in conceptual space through higher-level abstractions. Just as catalytic polymers undergo a phase transition to a state where there is a catalytic pathway to each polymer present and together they constitute a self-replicating set, memories and concepts undergo a phase transition to a state where each is retrievable through a path of associations. Together they now constitute an autocatalytically closed relationally structured conceptual network or worldview, a transmittable internalized tapestry of reality that both weaves, and is woven by, threads of experience.

Thus in the abstract mind, the information amplifying effect of autocatalytic closure happens all over again at another level when memories and abstractions are woven into an autocatalytically closed, interrelated web. Once again, the direction of information flow becomes nonsymmetrically inward-biased because new elements are generated (in this case, concepts, categories, stories, analogies, and other sorts of abstractions) that inform and get informed by (in this case through associative pathways) those that existed already (stored memories and previously-formed abstractions). Once again the newly generated elements are exactly those needed to *integrate* existing elements into a united whole.

2. *The conceptual closure structure is poised at the edge of chaos*

There is increasing evidence that the dynamics of cognition and consciousness are, like the dynamics of (other) organic systems, poised at the edge of chaos (e.g. Orsucci, 1998). Combs (1996) suggests ‘consciousness [is] a system near the edge of chaos, and states of consciousness [are] chaotic attractors by which the various aspects of the mind-body system, and especially the brain, are drawn into patterns of activity, or basins’ (p. 168). The implications of the edge of chaos principle for cognition is clear. If memories were distributed so narrowly (that is, if the cell assemblies that stored particular experiences were so tiny) that they never overlapped, the current experience would have to be identical to a previously-stored one to evoke it. On the other hand, if they were *very* widely distributed, successive thoughts would not necessarily be meaningfully related to one another.³

In neural networks, the need for constraining distributions is well-known (Hinton, 1990). If *every* input activates *every* memory location, as in Figure 6b, the system is subject to crosstalk; nonorthogonal patterns interfere.⁴ Techniques for constraining the degree to which memories are distributed have proven

[3] The free-association of a schizophrenic seems to correspond to what one might expect of a system like this (Weisberg, 1986).

[4] This phenomenon is also referred to as ‘superposition catastrophe’, ‘false memories’, ‘spurious memories’ or ‘ghosts’ (Feldman & Ballard, 1982; Hopfield, 1982; Hopfield *et al.*, 1983; von der Malsburg, 1986).

highly effective. One way of constraining distributions is to use a radial basis function (RBF) (Hancock *et al.*, 1991; Holden & Niranjana, 1997; Lu *et al.*, 1997; Willshaw & Dayan, 1990). Each input activates a hypersphere⁵ of memory locations, such that activation is maximal at the center k of the RBF and tapers off in all directions according to a (usually) Gaussian distribution of width s . Thus one part of the network can be modified without affecting the capacity of other parts to store other patterns. The further a stored concept is from k , the less activation it not only *receives* from the input but in turn *contributes* to the output, and the more likely its contribution is cancelled out by that of other simultaneously evoked locations. A small value for s corresponds to the situation where the activation function is tall and narrow in shape; few memory locations get activated, but these few are hit hard. A large s corresponds to the situation where the activation function is wide and flat; locations in the vicinity of k are activated almost as much as k itself. The wider the activation function, the more memory locations activated, and the greater the likelihood that something is retrieved from the memory and a stream of associative thought ensues. Thus the probability that one item evokes activation of another is determined by s rather than a random probability P as in Kauffman's OOL model, but the idea is the same.

In brains, much as in these neural networks, an intermediate degree of connectedness arises because the arrangement of neuron interconnectivity ensures that a particular constellation of stimulus dimensions perceived at any instant causes some memory locations to be activated, to varying degrees, and not others. Thus a given instant of experience activates (and retrieves 'ingredients' for the next instant of experience from) not just *one* location in memory, nor *every* location to an equal degree, but an assembly of neurons containing many memory locations. Thus although the storage and retrieval of memories is distributed the degree of distribution is *constrained*. As a result, the degree of association amongst items stored in memory is intermediate. At a cognitive level, this is consistent with Rosch's (1978) work on basic level categories. She holds that the way we organize information is not arbitrary but emerges in a way that maximizes explanatory power, and that this is the case when for each category or concept there exists an intermediate number of instances. So cognitive space is recursively clustered with hierarchical levels of entities that are similar enough to be categorized as members of the same concept, yet different enough to be distinguished as different instances (or as Bohm, 1980, put it, 'similar differences and different similarities').

Dynamically, the consequence of constrained, distributed storage and retrieval from memory is that there tends to be an intermediate degree of similarity between one instant of experience and the next; it underlies our capacity for a stream of coherent yet potentially creative thought. It is worth stressing that during a stream of thought there is no explicit *search* taking place, just information flowing through a system displaced from equilibrium. The current instant of experience activates certain neurons, which in turn activate certain other

[5] A hypersphere is a sphere with more than three dimensions.

neurons, which leads to the distributed storage of that experience in the memory locations in these neurons, which activates whatever *else* is stored in those locations, which then merges with salient sensory and motivational or goal-oriented information to form the *next* instant of experience, etc., recursively. What emerges is that the system appears to retrieve experiences that are similar, or concepts that are relevant, to the current experience. But that is simply a side effect of the fact that correlated qualia patterns get stored in overlapping locations. The semantic continuity of a stream of thought has a unifying effect on the memory at large. When experience *A* evokes a memory of experience *B*, *B* gets tinged by *A* and vice versa. The subjectivities of these events become intermixed.

3. *Maintaining the amplification structure through social exchange*

A conceptually-closed worldview replicates, not according to a code like biological organisms do now, but in the same piecemeal manner as did pre-DNA life, when we express and assimilate ideas (Gabora, in press). An adult shares concepts, ideas, stories, and experiences with children (and other adults), spreading his or her worldview little by little. The children expose their ‘copy’ of these fragments of what was originally the adult’s worldview to different experiences, different bodily constraints, and thus sculpt different worldviews, unique internal models of the relation of self to world. Over the course of childhood, situations arise that sooner or later provide exposure to the most formative and useful elements of a culture. These concepts, stories, procedures, and so forth get fitted together in the child’s mind in a somewhat (but not altogether) new way. Thus the information amplifying conceptual closure structure is not only maintained, but each ‘copy’ is a little different. Just as we saw earlier with sexually reproducing organisms, the overall informativeness is higher than it would be if replication were accomplished through the precise duplication of a single ‘parent’ entity.

4. *Convergence and divergence through variable focus*

It has been suggested that the transition from episodic mind to modern mind occurred in two stages. The origin of culture approximately two million years ago is overshadowed in the archeological record by another cultural transition during the Middle/Upper Paleolithic, following the appearance of *Homo sapiens sapiens* 100,000 years ago. Various possible explanations for the explosion of creativity we see at this time have been put forth, such as the advent or more generalized use of complex language (Aiello & Dunbar, 1993; Bickerton, 1990; 1996; Carstairs-McCarthy, 1999; Davidson, 1991; Dunbar, 1993; 1996) that material culture came to function as an externalized memory (Donald, 1991; 2001), or that domain-specific modules got connected (e.g. Mithen, 1996). Though each of these has potential merit, none provides a satisfactory account of why this transition occurred (for full explanation — as well as full presentation of the alternative proposal below — see: Gabora, submitted).

The abstract mind, as we left it at the end of section III.3, could redescribe an experience in terms of previous experiences. But since it activated regions of

conceptual space of fixed size, it had limited ability to focus. Even if it had some form of language, it could not use it in a sophisticated manner to unearth relationships that previously went unnoticed, and progressively zero in on the aspects that are most relevant. Thus, although the abstract mind was capable of creativity, that capacity was limited. What kind of functional change would make the creativity of the modern human mind possible?

To understand what caused this unprecedented explosion of creativity it is useful to examine briefly the psychology of creativity. Creativity is thought to require the capacity for both a fluid, intuitive, associative, or correlation-based form of thinking, and a more controlled, logical, causation-based form of thinking (e.g. Boden, 1990; Dartnell, 1993; Dennett, 1978; Gabora, 2002; James, 1890; Johnson-Laird, 1983; Neisser, 1963; Piaget, 1926; Rips, 2001; Sloman, 1996). There is in fact a considerable body of experimental evidence that creativity is associated with, not just cognitive fluidity, nor control, but both (Eysenck, 1995; Feist, 1999; Fodor, 1995; Richards *et al.*, 1988; Russ, 1993). Feist, for example, writes: 'It is not unbridled psychoticism that is most strongly associated with creativity, but psychoticism tempered by high ego strength or ego control. Paradoxically, creative people appear to be simultaneously very labile and mutable and yet can be rather controlled and stable' (Feist, 1999, p. 288). He notes that, as Barron (1963) put it: 'The creative genius may be at once naïve and knowledgeable, being at home equally to primitive symbolism and rigorous logic. He is both more primitive and more cultured, more destructive and more constructive, occasionally crazier yet adamantly saner than the average person' (Barron, 1963, p. 224).

This suggests that, once memories were distributed widely enough to permit abstraction and worldview closure, the next step would be to make them neither more nor less widely distributed, but more variable in the degree to which they are distributed. It is proposed that the Paleolithic revolution took place because of onset of the capacity for *variable focus*, where focus refers specifically to the shape of the distribution of memory locations activated and retrieved from memory at any instant. Recall from section III.2 that a flat activation function causes activation of a large cell assembly and thus many locations in memory, whereas a narrow one activates a small cell assembly with few locations. The idea here is that new experiences touch more or fewer memory locations, and to a greater or lesser degree, according to the situation and how far along in the creative process one is. When new information is inconsistent with the worldview (as it stands), a rough idea is hewn by becoming receptive to remote or subtle associations through a widening of the activation function, and the idea is thereafter refined through a narrowing of the activation function. Thus we acquired the capacity to shift back and forth at will from what Kauffman refers to as a *supracritical state* — conducive to disrupting inadequate structure — to a *subcritical state* — conducive to patching structure back together.

So after the mind became conceptually closed, it still had to acquire the capacity to self-organize its dynamics to best accommodate new information, such that its modules could work together in a coordinated manner to recursively

redescribe experiences. Variable focus meant that the fruits of divergent, creative thinking could be systematically refined through focused, rational thought, and the worldview could be conceptually closed at multiple, hierarchical, interpenetrated levels of abstraction. Thus the information amplifying effect of conceptual closure became even more pronounced.

5. *Holographic memories and contextualized concepts*

Whereas the bulk of neuroscientific research has focused on spatial coding, there is considerable evidence from many sensory modalities that much information processing in the brain is temporally coded (e.g. Abeles *et al.*, 1993; Campbell & Robson, 1968; Cariani, 1995; 1997; Emmers, 1981; Lestienne, 1996; Lestienne & Strehler, 1987; Metzinger, 1995; Mountcastle, 1993; Perrell & Bullock, 1968; Riecke & Warland, 1997; Stumpf, 1965; for reviews see De Valois & De Valois, 1988; Pribram, 1991). Different kinds of information, and different dimensions, are carried by different frequencies, much like a radio broadcast system. Static stimuli elicit no response; unless there is a *change*, nothing registers (Mackay, 1986). As Cariani points out, temporal coding drastically simplifies the problem of how the brain coordinates, binds, and integrates information. Wave amplitudes determine the size of the cell assembly activated in response to a signal. Perceptual and cognitive information combines through *multiplexing*, which works according to the principles of superposition and interference discussed in section two. So in effect, a stream of experience can be viewed as a series of changes in the phase relations of a bundle of waves of various frequencies.⁶

Many who take a dynamical approach to cognition have commented on the similarity between distributed memory storage in the brain, and how information is spread across a holographic plate (Alkon, 1989; Landfield, 1976; Mishkin & Appenzeller, 1987; O'Keefe, 1986; Pribram, 1991, 1997; Pribram & Bradley, 1998). Let us look briefly at how spatiotemporal information is integrated in a holography-inspired model of memory such as Pribram's. Recall that a hologram is created when a *signal wave S* carrying the image interacts with a *reference wave R* of coherent light in an optically sensitive material such as film, crystal (Psaltis & Mok, 1995), or organic polymers (Sincerbox, 1999)⁷ to generate an interference pattern, or grating. Either of the two original waves can then be used to reconstruct the other. Re-orienting the reference wave into the holographic material at the same angle as before regenerates the input image. Likewise, illuminating the holographic material with an image similar to the original regenerates the reference wave. Thus, every time a ray of light is shined from a different angle, a different stored hologram is elicited. (Specifically, the output is a

[6] These accelerations/decelerations are in the same direction as the wave itself for transverse waves such as sound, and perpendicular to it in the case of longitudinal waves such as light registered by the eyes.

[7] Sincerbox (1999) shows that organic materials can act as holograms when photosensitive monomers undergo polymerization in regions of constructive interference, thus setting up a concentration gradient, and recruiting more monomers to these regions, which too polymerize. This sets off a pattern of modulated refractive index.

function of the correlation between the input and each of the stored holograms, and depends also on how closely their directions match that of the input. Just as in neural networks, the phenomenon is referred to as associative retrieval.) In Pribram's physiologically explicit model, attractors activate one another much as do the interference rays that create a hologram, flowing in a smooth progression from one to the next. Pribram suggests that interference patterns in the brain are recorded through changes in the conformations of biomolecules at membrane surfaces. Since the extent to which representations are distributed is constrained, he refers to his model as not holographic but *holonomic*. Different inputs resonate with, and therefore activate and evoke retrieval from different distributions of memory locations.⁸ If the nervous system indeed works through the matching of resonance frequencies, wave superposition, and interference, as Cariani, Pribram (and others) suggest, it would make sense that neurons in the association cortex have a wider range of resonant frequencies than neurons in other regions. Then, just as in the sphere example, the simultaneous existence of both regions specialized for different kinds of information (sensory cortex) and regions appropriate to their convergence (association cortex) make it easier to stay poised at the edge of chaos.

The extent to which holography-inspired approaches to memory really model what is going on in the brain has yet to be determined. But one perplexing aspect they capture nicely is the *contextual* nature of the retrieval process. Much as light coming from one angle elicits one 'version' of what is stored in the holograph and light coming from another angle elicits another, in the context of fine cuisine, 'pot' evokes something you cook in, but in the context of drugs it evokes marijuana. The fact that the meanings of concepts are not rigid or static but shift fluidly depending on context increases dramatically the mind's potential to both inform and be informed by the world, and retrieval from memory is actually a

[8] Thus each memory location is particularly responsive to some stimulus property, and more likely to become involved in the storage and retrieval of experiences with this property. This stimulus property may be something we would recognize, such as size or colour, or it may not be something we are likely to ever consciously conceive of, such as, perhaps, the property of being a word that simultaneously evokes feelings of joy and reminds one of snow. Thus there will be overlap in the properties of memory traces that get stored in a particular location, though for any given location, the different memories stored there will also have *other* properties that are quite different. So if one specifies the memory location of interest, one cannot specify the particular memory trace, because many memory traces are stored in each location. However if one specifies the particular memory trace of interest, one cannot specify the particular memory location involved, because the memory trace will be stored, in a graded fashion, in all the locations responsive to its various salient properties. To make this more concrete, say you hear a meow sound coming from a box but you don't know which cat it is, Glimmer the white one, or Inkling the black one. (This in fact happened.) If you want to specify a particular memory location that gets activated at this instant, say one that is particularly responsive to meow sounds, you find that there are many different memories stored there, some that involve meow sounds made by Glimmer, others that involve meow sounds made by Inkling, and still others that involve meow sounds made by other cats, or by people pretending to be cats, and so forth. On the other hand, if you want to specify what memory will be evoked by this experience then many different locations will be involved. If this experience of the meow sound coming from the box reminds you of an event wherein you heard a meow sound in the next room and it turned out to be Glimmer, this particular memory trace is stored in and gets evoked through the activation of *many* different memory locations. In other words, there is a kind of uncertainty relationship between the property of interest and the location in memory. (Of course, to be completely certain of this, we might want to do a 'cat' scan . . . !)

matter of *reconstruction*. As Edelman and Tonini (2000, p. 101) put it: ‘Every act of perception is, to some extent, an act of creation, and every act of memory is, to some degree, an act of imagination.’

In quantum mechanics it was these very phenomena — the presence of contextuality and the emergence of genuinely new states — that forced the development of new mathematical formalisms. The mind often faces situations that are contextual in the same way that situations in quantum mechanics are; situations in which the state of the entity (in our case, a cognitive state) *actualizes in the process of being measured* or tested. As an example, say that Natalie is driving to the election booth, in a state of indecision as to whom she will vote for. We refer to the state of her mind as $p(t_0)$. In a first scenario, a reckless driver almost runs into her, and her state of mind collapses to the decision to vote for Cameron, the candidate who has taken a strong stance to make the roads safer. In a second scenario, she is stopped and given a speeding ticket, and her state of mind collapses to the decision to vote for the alternative candidate, Conner, who has argued that too much money is spent on the police force. Now we cannot say that the state $p(t_0)$ her mind was in when she began the trip to the election booth had as a true property ‘I will vote for Cameron’. However, we also cannot say that it had as a true property ‘I will *not* vote for Cameron’. Because of the contextual nature of the situation, both were *potentially* true, and both were *potentially* false. As another example if someone were to ask you in a hostile tone of voice ‘Are you angry?’ you might answer (and believe) ‘yes!’, whereas if asked the same question in a sympathetic voice you might answer (and believe) ‘no’. Again, before the question was asked, both were potentially true, and both potentially false; it isn’t a ‘one or the other is true’ situation. The opinion was contextual; it actualized *in the process* of being measured. Because of the contextual manner in which concepts are formed, evoked from memory, and often merged together to arrive at a decision or understanding of a situation, generalizations of the formalisms originally developed for quantum mechanics are turning out to be transferable to the cognitive process (this is gone into in detail in Aerts & Aerts, 1997; Gabora & Aerts, 2002a,b). So as a new thought takes shape, a new conscious experience is being had, and we have a formalism with which we can begin to capture that, contextual though it may be.

IV: Why Assessments About Consciousness Are Warped

Let us now address the second challenge to a fundamental theory: if consciousness is a primitive feature of the universe, why does it seem that only humans, and perhaps some animals, are conscious?

Although it would be hard to prove that consciousness is fundamental and ubiquitous, it is relatively easy to see why, if this is the case, we would be unaware of it. We saw that because organisms are autocatalytically closed, the direction of informing is inward-biased. Thus, if the double aspect theory is true, phenomenality is biased toward involving one part of the self experiencing another part of the self, and away from experiencing the exterior world. This is

even more the case for those with conceptually closed models of the world (including the self and its relation to that world), i.e. modern humans.

Moreover, it would seem that some degree of subjective isolation is necessary. An organism eats other living things to survive; it must value — or behave as if it values — its own subjective experience over that of those it consumes. For example, a wolf that is prone to experience life from the rabbit's point of view might well be less inclined to eat the rabbit than a wolf that is not. As another example, a meat-eater may *believe* that a cow is conscious, but be disinclined to empathically experience this consciousness in a first person sort of way. The better an organism is at shielding itself from the consciousness of non-self parts of the world, the more readily it will view a sufficient proportion of the stimuli it encounters as potential foodstuffs, consume them, and survive. This does not mean that we can be expected to deny the possibility that other entities are conscious. It is in the interest of our inclusive fitness to treat living things that are genetically similar to us as if they were conscious.⁹ We share our genetic makeup, to varying degrees, with all living things; accordingly, we gauge the degree to which other entities are conscious via the extent to which the phenotypic expression of their genetic makeup, as manifested through appearance and behavior, mirror our own. However, there is no *a priori* reason to believe that entities that do not express their conscious experience in ways we can mentally simulate, or readily empathize with, are not consciousness. *Being* conscious is not equivalent to *manifesting* it in characteristically human ways. So the feasibility of consciousness in entities quite unlike ourselves cannot be ruled out.

The bottom line is that since even if the world is permeated with a primitive form of consciousness we would be largely shielded from consciousness external to ourselves, we are in no position objectively to assess the extent to which entities other than ourselves are conscious. It follows that the seeming paucity of consciousness may be an illusion.

V: Light of Consciousness, My Love, the Light of my Life

There is a web called Indra's Net, made of threads of light. It stretches horizontally through space, and vertically through time. At every intersection dwells an individual, and at every individual lies a crystal bead of light. — *Buddhist allegory*

We have looked at various transitions from simple physical systems, to living systems, to the modern human mind, and seen how information got amplified. If information has a phenomenal aspect, then consciousness gets amplified along the way. At this point I will go out on a limb and suggest that to the extent that lovers (and also perhaps, enemies) put one another into such a state that they notice and respond to the subtleties of one another's informative actions, attitudes, and so forth, the state of being in love perhaps constitutes a yet more

[9] There may be a trade-off between acting as if other species *are* conscious (which aids the survival of the genes we share with them), and acting as if they are *not* (so we will eat them to survive ourselves). This might help explain our ambivalent attitudes about animal consciousness.

conscious state. Also, the loss of love can leave one feeling zombie-like, suggesting that the converse may also hold. It is interesting that the substance with the highest refractive index, and therefore most effective at trapping (potentially phenomenal) light, is the substance we associate with love: diamond.

As Zajonc (1993) points out, light has been used as a metaphor for love, wisdom, and insight since the dawn of civilization. Religious history is replete with accounts of something not just vaguely light-like but an actual experience of a rarefied light that is felt rather than seen, and seems to burn from within. Many cultures have a name for this inner light. The Eskimo shaman calls it *qaumaneq*. Vedanta Hinduists call it *Atma*. The *Tibetan Book of the Dead* refers to it as the clear light of Buddha-nature. The metaphor permeates our language, as in: enlightenment, moment of illumination, he beamed, her face lit up, to glow with enthusiasm, flash of brilliance, ray of hope, it is clear that, dim-witted person, light of my life, show me the light, dark night of the soul, etc. Even pictorial forms of communication have this property. Everyone knows what it means when lightning flashes from Mary Worth's eyes, or a lightbulb appears above Charlie Brown's head. Whether or not you refer to it as the 'light of consciousness', as Wolken (1986) shows in some detail, *all* organic and cognitive information processes originate with and are made possible by the harnessing of light through processes such as photosynthesis.

VI: Summary and Conclusions

Fundamental approaches bypass the problem of how to get consciousness out of non-conscious components by positing that a primitive form of consciousness is in the very building blocks, starting from the lowest level. There is a trivial sense in which everything is at least *potentially* conscious. Any particle can convert to a photon which can undergo photosynthesis and get incorporated into, say, a lettuce leaf, and end up in my brain, which I know is conscious. So even if we don't obviously live in a world in which everything *is* conscious, we live in a world where everything is *potentially* conscious. It is admittedly a step further to say that everything *is* conscious, or that consciousness is a fundamental feature of the universe. Despite the counterintuitiveness of this thesis, it has held sway for a long time, and there are various ways in which it has been developed. In a bold move, but bolstered with persuasive and influential arguments, Chalmers (1996) proposed that all information has a conscious, or phenomenal aspect. Two questions this proposal leaves us with are first, how do you get from phenomenal information to the real McCoy, human consciousness? And second, why does it seem to us that only humans, and perhaps some animals, are conscious?

Following up on Chalmers' proposal, it seems reasonable to suggest that the degree to which an entity is conscious is a function of the degree to which it amplifies information. We began by looking at some simple principles that demonstrate that local amplification is feasible. These principles were exemplified using a sphere composed of a material with a high refractive index such that incoming light is repeatedly reflected back into the sphere and thus trapped. To

stay informative, the sphere maintained its dynamics at the edge of chaos through simultaneous processes of divergence and convergence. We saw that a signal can inform an object by matching the natural frequency distribution of that object and thereby resonating with it. Divergence was facilitated through differential resonant frequencies at the surface of the sphere. The range of resonant frequencies increased as we penetrated the sphere, enabling convergence through superposition, including constructive and destructive interference.

The sorts of mechanisms through which conscious minds have come into existence and through which they amplify potentially phenomenal information have some similarities to principles involved in the amplification of light in our simple structure. We looked at several transitions from inorganic matter to conscious humans that profoundly increased the degree to which information is amplified. The first was the origin of life, which likely occurred through autocatalytic closure. Another significant transition took place in the structure of the human mind when memories and abstractions became interconnected to yield what from a first-person perspective was experienced as a relationally structured, modifiable mental model of the world, or worldview. We saw how this could have happened through conceptual closure. Yet another transition occurred when humans acquired the ability to vary the degree of focus according to the relative utility of correlation-based versus causation-based thinking. If the double aspect theory of information is correct, with each transition, not only were there more possible states for the entity to be in, but the degrees of freedom of what could be *experienced* increased.

It is interesting to look at how closure amplifies information. As separate elements transform into a closure structure, they generate not just new information-processing components, but exactly those that can be exploited by what is already in place such that they can collectively function as an integrated whole. Once a potentially informative agent (such as, say, food, or a sensory stimulus) becomes part of the system, it has more opportunities to interact with other parts of the system than with that which is not part of the system. The direction of information processing is nonsymmetrically inward-biased as it was in the sphere. Moreover, because this structure is able to replicate itself with endless variety and increasing complexity, this process of locally amplifying information is self-perpetuating.

Dynamical and holographic models of memory suggest that perceptual and cognitive information is temporally coded, and combines through superposition, making use of constructive and destructive interference. In this way, individual streams of phenomenally-endowed information could combine their subjectivities into a single, concentrated subjectivity, bringing about an unprecedented transition in the degree to which consciousness is locally amplified. In these models, significant changes in phase relations are registered and potentially stored as memories. Different memory locations have different natural frequency distributions, so the distribution of memory locations activated (determined by the size and shape of the activation function) depends on which locations resonate with the signal. Whether these models turn out to be realistic is

not yet clear, but they do capture the contextual, reconstructive nature of memory, and support the idea that generalizations of formalisms developed for quantum mechanics (developed because of the need to handle problems arising also in cognition: context and the appearance of genuinely new states) may prove useful to formal theories of cognition and consciousness.

Note that since closure fences an entity off, to some extent, from its environment, it might not only amplify the entity's own consciousness, but tend to make the entity underestimate the degree to which *others* are conscious. Moreover, to the extent that we must eat non-self parts of the world to survive, there may be an evolutionary advantage not to be overly distracted by the consciousness of the world outside the self. Thus there is reason to believe that we tend to underestimate how widespread consciousness is.

If we reject fundamental approaches to consciousness out of hand, we are faced with the daunting task of explaining how the qualia of first person experience could suddenly emerge. Either that, or conclude that one's personal experience of consciousness is an illusion (Dennett, 1991). The line of reasoning presented in this paper suggests that it is the relative paucity of consciousness in the universe that is an illusion. The miracle of consciousness may be, not that you or I have conscious experience, but that the consciousness everything else is experiencing doesn't usually interfere.

Acknowledgements

I would like to thank Peter Cariani, David Chalmers, Allan Combs, Anthony Freeman, Ben Goertzel, and Glenn Sincerbox for discussion and comments. I am grateful to Gary Waldman whose book *Introduction to Light: The Physics of Light, Vision, and Color* (1983) I used a reference for the material on light. This work was supported by Flanders AWI-grant Caw96/54a.

References

- Abeles, M. and Bergman, H. (1993), 'Spatiotemporal firing patterns in the frontal cortex of behaving monkeys', *Journal of Neurophysiology*, **70** (4), pp. 1629–38.
- Aerts, D. & Aerts, S. (1997), 'Application of quantum statistics in psychological studies of decision processes', in *Topics in the Foundation of Statistics*, ed. B. Van Fraassen (Dordrecht: Kluwer Academic).
- Aiello, L. and Dunbar, R. (1993), 'Neocortex size, group size, and the evolution of language', *Current Anthropology*, **34**, pp. 184–93.
- Alkon, D.L. (1989), 'Memory storage and neural systems', *Scientific American*, July, pp. 26–34.
- Barron, F. (1963), *Creativity and Psychological Health* (Hong Kong: Van Nostrand).
- Bateson, G. (1972), *Steps to an Ecology of Mind* (San Francisco, CA: Chandler).
- Bickerton, D. (1990), *Language and Species* (Chicago, IL: Chicago University Press).
- Bickerton, D. (1996), *Language and Human Behavior* (London: UCL Press).
- Boden, M.A. (1990/1992), *The Creative Mind: Myths and Mechanisms* (London: Weidenfeld & Nicolson, revised edition, Cardinal).
- Bohm, D. (1980), *Wholeness and the Implicate Order* (London: Routledge & Kegan Paul).
- Campbell, F.W. and Robson, J.G. (1968), 'Application of Fourier analysis to the visibility of gratings', *Journal of Physiology*, **197**, pp. 551–66.
- Cariani, P. (1995), 'As if time really mattered: Temporal strategies for neural coding of sensory information', in *Origins: Brain and Self-organization*, ed. K. Pribram (Hillsdale, NJ: Erlbaum).
- Cariani, P. (1997), 'Temporal coding of sensory information', in *Computational Neuroscience: Trends in Research 1997*, ed. J.M. Bower (Dordrecht, Netherlands: Plenum).
- Carstairs-McCarthy, A. (1999), *The Origins of Complex Language* (Cambridge: Cambridge University Press).
- Chalmers, D.J. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press).

- Cemin, S.C. and Smolin, L. (1997), 'Coevolution of membranes and channels: A possible step in the origin of life', *Journal of Theoretical Biology* (October issue) [adap-org/9709004].
- Combs, A. (1996), *The Radiance of Being* (St. Paul, MN: Paragon House).
- Dartnell, T. (1993), 'Artificial intelligence and creativity: An introduction', *Artificial Intelligence and the Simulation of Intelligence Quarterly*, **85**.
- Davidson, I. (1991), 'The archeology of language origins: A review', *Antiquity*, **65**, pp. 39–48.
- Deacon, T.W. (1997), *The Symbolic Species: The Co-Evolution of Language and the Brain* (New York: W.W. Norton & Company).
- De Duve, C. (1995), *Vital Dust: Life as a Cosmic Imperative* (New York: Basic Books).
- Dennett, D. (1978), *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, MA: Bradford/MIT Press).
- Dennett, D. (1991), *Consciousness Explained* (Boston, MA: Little, Brown and Company).
- De Valois, R.L. and De Valois, K.K. (1988), *Spatial Vision* (Oxford: Oxford University Press).
- Donald, M. (1991), *Origins of the Modern Mind* (Cambridge, MA: Harvard University Press).
- Donald, M. (2001), *A Mind So Rare* (New York: Norton Press).
- Dunbar, R. (1993), 'Coevolution of neocortical size, group size, and language in humans', *Behavioral and Brain Sciences*, **16** (4), pp. 681–735.
<http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.dunbar.html>
- Dunbar, R. (1996), *Grooming, Gossip, and the Evolution of Language* (London: Faber & Faber).
- Dyson, F. (1979), *Disturbing the Universe* (New York: Harper & Row).
- Edelman, G.M. (1987), *Neural Darwinism: The Theory of Neuronal Group* (New York: Basic Books).
- Edelman, G.M. (1992), *Bright Air, Brilliant Fire* (New York: Basic Books).
- Edelman, G.M. and Tononi, G. (2000), *A Universe of Consciousness* (New York: Basic Books).
- Emmers, R. (1981), *Pain: A Spike-Interval Coded Message in the Brain* (Philadelphia, PA: Raven Press).
- Erdos, P. and Renyi, A. (1959), 'On random graphs, I', *Publicationes Mathematicae*, **6**, pp. 290–7.
- Erdos, P. and Renyi, A. (1960), 'On the evolution of random graphs', *Institute of Mathematics, Hungarian Academy of Sciences*, **5**, pp. 17–61.
- Eysenck, H.J. (1995), *Genius: The Natural History of Creativity* (Cambridge: Cambridge University Press).
- Farmer, J.D. Kauffman, S.A. and Packard, N.H. (1987), 'Autocatalytic replication of polymers', *Physica D*, **22** (50).
- Feigl, H. (1958/1979), 'The "mental" and the "physical"', *Minnesota Studies in the Philosophy of Science*, **2**, pp. 370–497 (Reprinted by University of Minnesota Press).
- Feist, G.J. (1999), 'The influence of personality on artistic and scientific creativity', in *Handbook of Creativity*, ed. R.J. Sternberg, (Cambridge: Cambridge University Press).
- Feldman, J.A. and Ballard, D.H. (1982), 'Connectionist models and their properties', *Cognitive Science*, **6**, pp. 205–54.
- Fodor, E.M. (1995), 'Subclinical manifestations of psychosis-proneness, ego-strength, and creativity', *Personality and Individual Differences*, **18**, pp. 635–42.
- Foster, J. (1989), 'A defense of dualism', in *The Case for Dualism*, ed. J. Smith & J. Beloff (Charlottesville, VA: University of Virginia Press).
- Gabora, L. (1998), 'Autocatalytic closure in a cognitive system: A tentative scenario for the origin of culture', *Psychology*, **9** (67). <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?9.67>.
- Gabora, L. (2000), 'Conceptual closure: Weaving memories into an interconnected worldview', in *Closure: Emergent Organizations and their Dynamics*, ed. G. Van de Vijver & J. Chandler, Annals of the New York Academy of Sciences 901.
- Gabora, L. (2002), 'Cognitive mechanisms underlying the creative process', *Proceedings of the Fourth International Conference on Creativity and Cognition*, October 13–16, Loughborough University, UK. <http://www.vub.ac.be/CLEA/liane/papers/CandC/CandC.htm>.
- Gabora, L. (in press), 'Cultural entities are not replicators but autocatalytic networks of them are', *Biology and Philosophy*.
- Gabora, L. (submitted), 'Origin of the modern mind through conceptual closure'.
<http://www.vub.ac.be/CLEA/liane/papers/ommcc/ommcc.htm>
- Gabora, L. and Aerts, D. (2002a), 'Contextualizing concepts', *Proceedings of the 15th International FLAIRS Conference*, Pensacola FL, May 14–17, sponsored by the American Association for Artificial Intelligence, pp. 148–52. <http://www.vub.ac.be/CLEA/liane/papers/flairs.html>.
- Gabora, L. and Aerts, D. (2002b), 'Contextualizing concepts using a mathematical generalization of the quantum formalism', invited to appear in special issue of *Journal of Experimental and Theoretical Artificial Intelligence* on 'Categorization and Concept Representation: Models and Implications'.
<http://www.vub.ac.be/CLEA/liane/papers/jetai.pdf>.
- Ghose, A. and Aurobindo, S. (1998), *Integral Yoga: Sri Aurobindo's Teaching & Method of Practice* (Twin Lakes, WI: Lotus Press).
- Goertzel, B. (1993a), *The Structure of Intelligence* (Berlin: Springer-Verlag).
- Goertzel, B. (1993b), *The Evolving Mind* (London: Gordon and Breach).

- Goertzel, B. (1995), 'Chance and consciousness', *Dynamical Psychology*.
<http://www.goertzel.org/dynapsyc/1995/GOERTZEL.html>.
- Griffin, D.R. (1998), *Unsnarling the World-Knot: Consciousness, Freedom, and the Mind-Body Problem* (Berkeley, CA: University of California Press).
- Hancock, P.J.B., Smith, L.S. and Phillips, W.A. (1991), 'A biologically supported error-correcting learning rule', *Neural Computation*, **3** (2), pp. 201–12.
- Hartshorne, C. (1968), *Beyond Humanism: Essays on the Philosophy of Nature* (Lincoln, NE: University of Nebraska Press).
- Hebb, D.O. (1949), *The Organization of Behavior* (New York: Wiley).
- Hinton, G.E. (1990), 'Implementing semantic networks in parallel hardware', in *Parallel Models of Associative Memory*, ed. G.E. Hinton & J.A. Anderson (Mahwah, NJ: Lawrence Erlbaum Associates).
- Holden, S.B. & Niranjan, M. (1997), 'Average-case learning curves for radial basis function networks', *Neural Computation*, **9** (2), pp. 441–60.
- Hopfield, J.J. (1982), 'Neural networks and physical systems with emergent collective computational abilities', *Proceedings of the National Academy of Sciences*, **79**, pp. 2554–8.
- Hopfield, J.J., Feinstein, D.L. and Palmer, R.G. (1983), "'Unlearning" has a stabilizing effect in collective memories', *Nature*, **304**, pp. 158–9.
- James, W. (1890/1950), *The Principles of Psychology* (New York: Dover).
- Johnson-Laird, P.N. (1983), *Mental Models* (Cambridge, MA: Harvard University Press).
- Karmiloff-Smith, A. (1992), *Beyond Modularity: A Developmental Perspective on Cognitive Science* (Cambridge, MA: MIT Press).
- Kauffman, S.A. (1993), *Origins of Order* (Oxford: Oxford University Press).
- Kauffman, S.A. (1999), 'Darwinism, neoDarwinism, and the autocatalytic model of culture: Commentary on *Origin of Culture*', *Psycoloquy*, **10** (22). <http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?10.022>
- Kurak, M. (2001), 'Buddhism and brain science', *Journal of Consciousness Studies*, **8** (11), pp. 17–26.
- Landfield, P.W. (1976), 'Synchronous EEG rhythms: Their nature and their possible functions in memory, information transmission and behavior', in *Molecular and Functional Neurobiology*, ed. E.H. Gispen (New York: Elsevier).
- Langton, C.G. (1992), 'Life at the edge of chaos', in *Artificial life II*, ed. C. G. Langton, C. Taylor, J.D. Farmer & S. Rasmussen (Boston, MA: Addison-Wesley).
- Lestienne, R. (1996), 'Determination of the precision of spike timing in the visual cortex of anesthetized cats', *Biology and Cybernetics*, **74**, pp. 55–61.
- Lestienne, R. and Strehler, B.L. (1987), 'Time structure and stimulus dependence of precise replicating patterns present in monkey cortical neuron spike trains', *Neuroscience*, April issue.
- Lockwood, M. (1989), *Mind, Brain, and the Quantum* (Oxford: Oxford University Press).
- Lu, Y.W., Sundararajan, N. and Saratchandran, P. (1997), 'A sequential learning scheme for function approximation using minimal radial basis function neural networks', *Neural Computation*, **9** (2), pp. 461–78.
- Mackay, D.M. (1986), 'Vision: The capture of optical covariation', in *Visual Neuroscience*, ed. J.D. Pettigrew, K.J. Sanderson and W.R. Levick (Cambridge: Cambridge University Press).
- Marr, D. (1969), 'A theory of the cerebellar cortex', *Journal of Physiology*, **202**, pp. 437–70.
- Maturana, H.R. and Varela, F.J. (1987), *The Tree of Knowledge: The Biological Roots of Human Understanding* (Boston, MA: Shambhala).
- Maynard-Smith, J. and Szathmary, E. (1994), *The Major Transitions in Evolution* (Oxford: Oxford University Press).
- Metzinger, T. (1995), 'Faster than thought: Holism, homogeneity, and temporal coding', in *Conscious Experience*, ed. T. Metzinger (Paderborn/Thorverton: Schöningh/Imprint Academic).
- Mishkin, M. and Appenzeller, T. (1987), 'The anatomy of memory', *Scientific American*, June, pp. 62–71.
- Mithen, S. (1996), *The Prehistory of the Mind: A Search for the Origins of Art, Science, and Religion* (London: Thames & Hudson).
- Montero, B. (2001), 'Post-physicalism', *Journal of Consciousness Studies*, **8** (2), pp. 61–80.
- Morowitz, H. (1992), *The Beginnings of Cellular Life* (New Haven, CT: Yale University Press).
- Mountcastle, V. (1993), 'Temporal order determinants in a somatosensory frequency discrimination: Sequential order coding', *Annals of the New York Academy of Science*, **682**, pp. 151–70.
- Nagel, T. (1979), *Mortal Questions* (Cambridge: Cambridge University Press).
- Neisser, U. (1963), 'The multiplicity of thought', *British Journal of Psychology*, **54**, pp. 1–14.
- O'Keefe, J. (1986), 'Is consciousness the gateway to the hippocampal cognitive map? A speculative essay on the neural basis of mind', *Brain and Mind*, **10**, pp. 573–90.
- Oparin, A.I. (1971), 'Routes for the origin of the first forms of life', *Subcellular and Cellular Biochemistry*, **1** (75).
- Orsucci, F. (1998), *The Complex Matters of the Mind* (Singapore: World Scientific).
- Perkell, D.H. and Bullock, T.H. (1968), 'Neural coding', *Neurosciences Research Program Bulletin*, **6** (3), pp. 221–348.
- Piaget, J. (1926), *The Language and Thought of the Child* (London: Routledge & Kegan Paul).

- Pribram, K.H. (1991), *Brain and Perception: Holonomy and Structure in Figureal Processing* (Hillsdale, NJ: Erlbaum).
- Pribram, K.H. (1997), 'The deep and surface structure of memory and conscious learning: Toward a 21st century model', in *Mind and Brain Sciences in the 21st Century*, ed. R.L. Solso (Boston, MA: MIT Press).
- Pribram, K. and Bradley, R. (1998), 'The brain, the me, and the I', in *Self-awareness: Its Nature and Development*, ed. M. Ferrari and R. Sternberg (New York: Guilford Press).
- Prigogine, I. and Stengers, I. (1984), *Order out of Chaos* (New York: Bantam Books).
- Psaltis, D. and Mok, F. (1995), 'Holographic memories', *Scientific American*, **273** (5), pp. 70–6.
- Richards, R.L., Kinney, D.K., Lunde, I., Benet, M. & Merzel, A. (1988), 'Creativity in manic depressives, cyclothymes, their normal relatives, and control subjects', *Journal of Abnormal Psychology*, **97**, pp. 281–9.
- Rieke, F. and Warland, D. (1997), *Spikes: Exploring the Neural Code*. (Cambridge, MA: MIT Press).
- Rips, L.J. (2001), 'Necessity and natural categories', *Psychological Bulletin*, **127** (6), pp. 827–52.
- Rosch, E. (1978), 'Principles of categorization', in *Cognition and Categorization*, ed. E. Rosch and B.B. Lloyd (Hillsdale, NJ: Lawrence Erlbaum).
- Russ, S.W. (1993), *Affect and Creativity* (Hillsdale, NJ: Erlbaum).
- Russell, B. (1926), *The Analysis of Matter* (London: Kegan Paul).
- Seager, W. (1995), 'Consciousness, information, and panpsychism', *Journal of Consciousness Studies*, **2** (3), pp. 272–88.
- Shannon, C.E. (1948), 'A mathematical theory of communication', *Bell Systems Technical Journal*, **27**, pp. 379–423. Reprinted in *The Mathematical Theory of Communication*, C.E. Shannon and W. Weaver (Champaign, IL: University of Illinois Press).
- Sincerbox, G. (1999), 'Holographic storage: Are we there yet?', Unpublished MS.
<http://www.optics.arizona.edu/Glenn/holograp1.htm>
- Slovan, S. (1996), 'The empirical case for two systems of reasoning', *Psychological Bulletin*, **9** (1), pp. 3–22.
- Sporns, O., Gally, J.A., Reeke, G.N. Jr. and Edelman, G.M. (1989), 'Reentrant signalling among simulated neuronal groups leads to coherency in their oscillatory activity', *Proceedings of the National Academy of Sciences USA*, **86**, pp. 7265–9.
- Sporns, O., Tonini, G. and Edelman, G.M. (1991), 'Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections', *Proceedings of the National Academy of Sciences USA*, **88**, pp. 129–33.
- Stoljar, D. (2001), 'Two conceptions of the physical', *Philosophy and Phenomenological Research*, **62**, pp. 253–81.
- Strawson, G. (2000), 'Realistic materialist monism', in *Toward a Science of Consciousness III*, ed. S. Hameroff, A. Kasniak & D. Chalmers (Cambridge, MA: MIT Press).
- Stumpf, C. (1965), 'Drug action on the electrical activity of the hippocampus', *International Review of Neurobiology*, **8**, pp. 77–138.
- Taylor, K. (2001), 'Applying continuous modelling to consciousness', *Journal of Consciousness Studies*, **8** (2), pp. 45–60.
- Von der Malsburg, C. (1986), 'Am I thinking assemblies?', in *Proceedings of the 1984 Trieste Meeting on Brain Theory*, ed. G. Palm & A. Aertsen (Berlin: Springer-Verlag).
- Waldman, G. (1983), *Introduction to Light: The Physics of Light, Vision, and Color* (Englewood Cliffs, NJ: Prentice-Hall).
- Weisberg, R.W. (1986), *Creativity: Genius and Other Myths* (New York: W.H. Freeman).
- Whitehead, A.N. (1929), *Process and Reality* (New York: Macmillan).
- Willshaw, D. and Dayan, P. (1990), 'Optimal plasticity from matrix memories: What goes up must come down', *Neural Computation*, **2** (1), pp. 85–93.
- Wolken, J. (1986), *Light and Life Processes* (New York: Van Nostrand Reinhold).
- Zajonc, A. (1993), *Catching the Light: The Entwined History of Light and Mind* (Oxford: Oxford University Press).

Paper received November 1999; revised February 2002