Philosophica 57 (1996, 1) pp. 53-90

A DAY IN THE LIFE OF A MEME

Liane Gabora

ABSTRACT

Like the information patterns that evolve through biological processes, mental representations or memes evolve through adaptive exploration and transformation of an information space through variation, selection, and transmission. However since memes do not contain instructions for their replication our brains do it for them, strategically, guided by a fitness landscape that reflects both internal drives and a worldview that forms through meme assimilation. This paper presents a tentative model for how an individual becomes a meme-evolving agent via the emergence of an autocatalytic network of sparse, distributed memories, and discusses implications for complex creative thought processes and why they are unique to humans. A hypothetical scenario for the evolutionary dynamics of a given meme in a society of interacting individuals is presented.

1. *Introduction: a second form of evolution*

While some ideas or concepts instantly fade into obscurity, others spread through a society, getting progressively refined and embellished, forging connections to established conceptual frameworks along the way. Thus the mental representations that underlie the content and expression of ideas, like the strands of DNA that encode instructions for building and maintaining living organisms, seem to evolve. Accordingly there has been a slow but steady effort to map the concept of evolution onto the dynamics of culture. Popper [1963] and Campbell [1987] alerted us to the evolutionary flavour of epistemology. Dawkins [1976] introduced the notion of a meme -a replicator of cultural information analogous to the gene. In his words: 'Just as genes propagate themselves in the gene pool by leaping from body to body via sperm or eggs, so do memes propagate themselves in the meme pool by leaping from brain to brain'. Others

have drawn from mathematical models of population genetics and epidemiology to model the spread of ideas [Cavalli-Sforza & Feldman 1981; Lumsden & Wilson 1981; Schuster & Sigmund 1983; Boyd & Richerson 1985; Hofbauer & Sigmund 1988]. These works point toward the possibility that memetic or cultural evolution constitutes a second form of evolution, distinct from yet intertwined with biological evolution, with the potential to provide the kind of overarching framework for the social and cognitive sciences that the first form provides for the biological sciences. However, these works have not brought about consensual understanding that cultural evolution can tell us much about human thought and behaviour, a situation that seems unfortunate given the success of the biological precedent. Although much was known about living things before Darwin, his theory of how life evolves through natural selection united previously disparate phenomena and paved the way for further biological inquiry.

This paper outlines a theory of how memes evolve, and illustrates how a memetic perspective provides not only not only a foundation for research into the dynamics of concepts and artifacts at the societal level, but a synthetic framework for understanding how mental representations are generated, organized, stored, retrieved, and implemented at the level of the individual. It also discusses obstacles to this kind of synthesis, such as our predisposition to focus on the individual as the basic unit, even when the individual is not the object of the evolutionary process under consideration, and concludes with an example of what Dennett [1995] calls a 'meme's eye view'.

Sceptics may wonder how we can hope to develop a theory of cultural evolution when we do not yet know the physiological details of how memes are instantiated in the brain. This situation has a precedent: Darwin came up with the theory of biological evolution through natural selection before the discovery of genes. It turned out that genes are laid out in a fairly straightforward way in physical space, which does not appear to be the case with memes. This does not mean they can't evolve, so long as there is a way of retrieving the components of a meme so they can work together as a unit. We may not know exactly how the information manifested in, say, a handshake between two individuals - with its unique arrangement of contact points, applied forces, and trajectory - can be traced back to these individuals' mental representations of handshakes, each other, and the situation they are in. But let us proceed with the confidence that a solution exists and can be found.

1.1 Components of an evolutionary system

In order for evolution to happen there must be:
1. A pattern of information (a state within a space of possible states).
2. A way to generate variations of the pattern (explore or transform the space).
3. A rationale for selecting variations that are adaptive - tend to give better performance than their predecessors in the context of some problem or set of constraints (a fitness landscape applied to the space).
4. A way of replicating and transmitting (or amplifying, as molecular biologists refer to it) the selected variations.

In biological evolution the evolving patterns of information are genes encoded as sequences of nucleotides. Variations arise through mutation and recombination, and natural selection weeds out those that are maladaptive. Replication takes place at the level of the genotype. In cultural evolution, the evolving patterns of information are memes - mental representations of ideas, behaviours, or other theoretical or imagined constructs, perhaps encoded as patterns of neuron activation. Variations are created by combining, perturbing, and reorganizing representations, consciously or unconsciously, or through errors in transmission. Replication is phenotypically mediated; it occurs when representations are transformed into action or language, transmitted through processes such as imitation, and reproduced, more or less, in another brain. Incorporation of these new information patterns into the society alters the selective pressures and constraints exerted by the social environment, which in turn leads to the generation of yet more patterns. Thus mental representations, like DNA, comprise a self-sustained system for the relentless exploration and
transformation of a space of possible patterns.

1.2 Disentangling cultural evolution from biological evolution

The line of reasoning presented here can be succinctly conveyed in terms of information, which is related to the number of differences required to specify the state of a system [Shannon 1963; Bateson 1972]. States have not only a structure of difference relations between them, but also a combinatorial structure: each state can itself be an information space, so that complex information can be built up from simple information. We

can rate each state in a space of possible states against some performance measure or fitness criterion, and the result is referred to as a fitness landscape. The world can be viewed as a vast computation where information is created, transformed, and destroyed. The information we encounter exhibits pattern, or statistical regularities that can be expressed mathematically.

After seeing many shadows cast by the same object we can develop an internal model of what that object looks like without having ever seen it, and if there is more than one object casting shadows we can learn to tell which object is casting any particular shadow. Similarly, by viewing every pattern we encounter as a shadow or footprint of one or more broad causal principles[2], we can eventually circumscribe the causal principles to which all pattern can be traced. We now take a step in this direction, the ultimate goal being to disentangle cultural information from biological information.

If you were to go back to some time during the first billion years of Earth's history, the only causal principle you would need to invoke to explain pattern in the information present (with the exception of yourself) would be the physical constraints and self-organizing properties of matter.

If you were to go back to some time after the origin of life, approximately three billion years ago, this would no longer be the case. Not that life doesn't exhibit the properties of matter. But it would be virtually impossible for, say, a giraffe to appear in an information space not acted upon by natural selection. Another causal principle - biological evolution - would have to be invoked from this point on.

Today the Earth is embedded with artifacts like computer networks and circuses that cannot be accounted for by appeal to either the properties of matter or biological evolution. That is, biological evolution does not provide us with adequate explanatory power to account for the existence of computers any more than the properties of matter can explain the existence of giraffes. Computers are manifestations of yet another causal principle: the evolution of culture.

Thus pattern in the structure and dynamics of information we encounter in the everyday world can be traced to three broad causal principles - the physical constraints and self-organizing properties of matter, biological evolution, and cultural evolution. This classification scheme, like all classification schemes, is somewhat arbitrary. There may be subclasses of these principles that deserve to be considered principles unto

themselves[2], or one could argue that evolution is a self-organizing property of matter, albeit a spectacular one[3]. The point is: culture is the only process that has arisen since the origin of life that relentlessly exploits the combinatorial potential of information.

Since the machinery that renders cultural evolution - the human brain- is a product of biological evolution, it is easy to confuse these two forms of evolution. To make things slipperier, much of what is 'out there' can not be cleanly traced to a biological or cultural origin. We will discuss how biology constrains culture through the preferential spread of memes that satisfy biologically-derived needs. It goes the other way too; culture not only affects biological fitness via behaviour - the Baldwin Effect - but it dramatically modifies the biological world. Some of the ways in which biological information gets tainted with cultural information seem to be relatively inconsequential, such as the trimming of hedges, whereas others, such as dog-breeding, have a long-lasting effect. In fact, one could view dogs as the consequence of a memetic trajectory that was launched by the need to protect property. At any rate the bottom line is: despite the fact that culture is grounded in biology (like biology is grounded in the physical constraints and self-organizing properties of matter), the probability of computers arising spontaneously in an information space not acted upon by cultural evolution (like the probability of giraffes arising spontaneously in an information space not acted upon by biological evolution) is vanishingly small. Thus it is inappropriate to dismiss culture as a predictable extension of biological evolution. It is qualitatively different from anything else
biology has produced.

1.3 Conceptual linkage disequilibrium

Arguments against a theory of cultural evolution generally consist of a series of statements as to how the cultural situation differs from that of biology [e.g. Gould 1991; Thagard 1980]. These arguments, however, do not constitute a viable reason to discard the idea that culture is an evolutionary process. Imagine that 100 years before Darwin proposed the theory of biological evolution through natural selection, another scientist had discovered another system whereby patterns of information evolved, say in a test tube. Given this scenario, would it have made sense for Darwin to dismiss out of hand or downplay the importance of a theory

of biological evolution simply because the evolution of biological organisms proceeds through different mechanisms from the originally-discovered test tube form of evolution? This would obviously have been foolish. It would have robbed humanity of not only the unifying power of a theory of biological evolution, but the opportunity to use knowledge of how evolution works under one set of constraints and affordances as a scaffold to direct the study of how it works under a different set of constraints and affordances. Nevertheless time and again it is implied that a theory of cultural evolution is doomed simply because it would have to work through different mechanisms from those of biological evolution.

Ironically this situation in itself provides us with a nice example of how knowledge of evolution acquired in the realm of biology can help unravel analogous situations in the realm of culture. The biasing effect of historical association is an important theme in population genetics. Alleles of linked genes, such as the those that code for red hair and freckles, continue to co-occur more often than chance even after individuals in the lineage from which these alleles originated begin mating randomly with individuals from other lineages that did not have these alleles. One can theoretically measure the number of generations necessary for these genes to achieve a state of random association or linkage equilibrium, and this process can be modeled computationally. Similarly, psychologists speak of mental set, wherein there is difficulty applying an idea or problem-solving technique to situations other than the one in which it was originally encountered, or exposure to one problem-solving technique interferes with the ability to solve a problem using another technique [e.g. Luchins 1942]. We could view mental set as a state of conceptual linkage disequilibrium. In the present example, achieving conceptual linkage equilibrium amounts to abstraction of the concept of evolution from its biological manifestation so that it can be applied with ease to the case of culture. One could argue that it would make sense for cultural evolution to be the default form of evolution in disciplines outside of biology, much as in tropical climates the default form of skiing is water-skiing is rather than snow-skiing.

2. *Meme, not individual, as a basic unit of cultural evolution*

The memetic approach to cognition is not incompatible with approaches

that stress the role of innate mechanisms [e.g. Pinker 1995]. Rather, as Lumsden and Wilson [1981] point out, it builds on this framework, adding that the study of cognition will flounder until we admit that the role of transmission is equally undeniable. The memetic approach involves relinquishing our focus on the individual as the unit of interest, and concentrating instead on the meme as the object of a second evolutionary process that makes cognition possible. This perspective discloses population-level phenomena that are easily overlooked because they are not readily detected through introspection. In that regard folk psychology errs through omission. Our anthropocentric tendency to focus on the individual is probably exacerbated by cultural linkage disequilibrium in that in biology the individual is the object of the relevant evolutionary process, or more specifically, the phenotypic expression of the information undergoing evolution.

2.1 The distinction between a meme and its phenotypic implementation

Durham [1991] defines a meme as 'any kind, amount, and configuration of information in culture that shows both variation and coherent transmission'. Problems with this definition arise because it does not distinguish between cultural information as mental representation and cultural information as implemented behaviour or artifact.

The genotype-phenotype distinction is useful here. The cultural analog of a genotype is the mental representation of a meme, and the analog of a phenotype is its implementation, or the form it takes if it gets expressed or communicated, typically as action or vocalization. Implementation transforms a meme, incorporating syntactic features characteristic of the channel through which it is conveyed [Brooks 1986]. Thus, for example, a dance step looks different with each individual who performs it. There can be nonlinear (epistatic) relations amongst the features of a meme, or between a meme and its implementation.

2.2 The interconnectedness of memes

Biologists use the term 'allele' to capture the notion of alternative heritable versions of a gene, and Durham [1991] accordingly adopted the term 'allomeme' to refer to alternative versions of a meme. This basic concept was tailored to meet the constraints of biology; we all have the

same number of genes, and two alleles of each gene (one from each parent). The cultural analog may be too clumsy to capture the subtle relationships between memes. Memes often appear to be stored in a distributed, network-like fashion, connected through webs of association [Hebb 1949; Quillian 1968; Pribram 1974]; there is not necessarily a definitive rationale for saying where one stops and another begins, in semantic space let alone physical space. For example would we consider 'My mother looks good in blue' and 'My mother looks good under a blue umbrella' to be allomemes of the same meme, or different memes?

This kind of difficulty is circumvented by avoiding the notion of alternate versions altogether and using the term 'feature' to refer to a component of a meme. Thus related memes share features. In this paper, 'feature' can refer to a component with any degree of granularity below that of the meme in question; thus the scope of what might be considered a feature could range from an entire array of visual information depicting every perceived quality of a particular umbrella (such as might occur early on in perception) to one bit of information indicating the presence or absence of an umbrella (such as might occur at an advanced stage of cognitive processing).

2.3 Meme as pattern of information encoded in the focus

The concept of a meme can be clarified further by invoking Kanerva's [1988] notion of the focus - that part of the mind in which sensation (either external or internal e.g. hunger) and stored memory interact to produce a stream of experience. The states of the neurons that comprise the focus determine the content and quality of an individual's awareness. One can think of a meme as a pattern of information that is or has been encoded in an individual's focus. It can be subjectively experienced as a sensation, idea, attitude, emotion, or combination of these, and it can direct implementation by the motor apparatus.

Frequently many memes get integrated into one through a process of 'chunking' [Miller 1956]. This process involves forming associations amongst previously- learned memes and establishing this constellation of associations as a new meme in long term memory; it is analogous to the formation of coadapted genes, or schemata [Holland 1975]. Whereas chunking generally refers to the binding of semantically unrelated memes (as in the memorization of an arbitrary string of numbers), categorization

involves the recognition of semantic relationships. Categorization and the resulting hierarchical structure of knowledge is dealt with by others in this volume and elsewhere [e.g. Van Loocke 1991]. The topic is not addressed in great depth here, tough it is of relevance to point out that as a consequence of chunking/categorization, the complexity of what can be held in the focus (and thus of what constitutes a meme, and thus a feature) will differ amongst individuals, and within an individual over time.

2.4 Core, enabler and hitchhiker features

A first step toward a science of memetics is to decompose memes into features or feature schemata according to how they relate to fitness. Here we will distinguish the following categories: (1) core features, that contribute directly to the fitness of a meme, (2) enabler features, that enable or facilitate the implementation or expression of core features, and (3) hitchhiker features, which exist in the meme due to arbitrary or accidental historical associations to features of the first two kinds. Core features tend to convey semantic information, and enabler features syntactic information, though one can think of situations in which some semantic information serves simply to facilitate expression of other semantic information i.e. functions as an enabler. The first two categories are vaguely analogous to the categorization of genes as structural or regulatory, and the last category is inspired by the phenomenon of genetic hitchhiking [Kojima & Schaffer 1967]. The closer together genes are on a chromosome the less likely they will be separated by crossover, so the more tightly linked they are said to be. Hitchhiker alleles confer no fitness advantage, but endure because they are linked to alleles that are important for survival. In both genetic hitchhiking and its cultural analog there is indirect selection for useless (or even detrimental) patterns through their association with beneficial ones. The concept of hitchhiking is closely related to that of exaptation -the evolution of organs or traits not evolved through natural selection for their current use [Gould 1982].

3. *Selection and the memetic fitness landscape*

The next few sections examine in some detail how each of the three phases of evolution - selection, variation, and transmission - map onto the

case of culture. Though these phases are discussed one at a time, it is worthwhile to keep in mind that in culture they are less spatiotemporally distinct than in biology. Selection can be coupled to either the generation of variation, or replication, or all three can occur simultaneously (for example when paraphrasing).

3.1 Memes rely on brains to select, vary and replicate them

Von Neumann [1966] postulated that any self-replicating system consists of two parts: (1) uninterpreted information - a self-description that is passively copied to offspring, and (2) interpreted information -instructions for how to construct offspring. This turned out to be true of the genetic code, but unlike genes, memes do not come with instructions for their reproduction. They rely on us, their hosts, to create, select, and replicate them. Since we preferentially spread ideas that satisfy needs, our needs define viable niches for memes to evolve toward. As infants we might cry and kick no matter what need is most pressing, but as children we acquire and continually refine a repertoire of memes that, when implemented, satisfy various needs. We learn that reaching into the cookie jar satisfies one need, shouting 'help' satisfies another, et cetera. Our memes, and the behaviour they elicit, slide into need-defined attractors (regions of stability) in the memetic fitness landscape.

3.2 Brains select memes that satisfy biological and cultural needs

Since many of our needs have a biological basis- e.g. the need for food, shelter, et cetera - meme generation is largely constrained by our heritage as products of biological evolution. Thus the topology of the memetic fitness landscape largely echoes that of the biological fitness landscape. In the short term the biological fitness landscape, and thus the memetic fitness landscape, fluctuates continuously as one need is satisfied and others take precedence [Hull 1943; McFarland & Sibly 1975; Gabora & Colgan 1988; Maes 1991]. For example, after eating, ideas that pertain to finding food are less likely. However over the lifetime of an individual the set of biologically-based needs remains relatively constant. The trajectory of survival-motivated thought can be described as a limit cycle (periodic attractor) that moves through the set of stable memes whose implementations satisfy the various biological needs.

Variation-inducing operations restructure conceptual space and thus affect the memetic fitness landscape. Much as the evolution of rabbits created ecological niches for species that eat them and parasitize them, the invention of cars created cultural niches for gas stations, seat belts, and garage door openers. As one progresses from infanthood to maturity, and simple needs give way to increasingly complex needs, the stream of thought acquires the properties of a chaotic or strange attractor, which can be viewed as the formation of crevices in the original limit cycle. The landscape is fractal (i.e. there is statistical similarity under change of scale) in that the satisfaction of one need creates other needs - every crevice when examined closely reveals more crevices. This is analogous to the fractal distributions of species and vegetation patterns described by ecologists [Mandelbrot 1982; Palmer 1992; Scheuring & Riedi 1994]. An endpoint of a cultural evolution trajectory turns out to be not just a point in multidimensional space, but a set of points with their own fitness metric - a micro-landscape 'in its own right'. So although the memetic fitness landscape loosely follows the biological fitness landscape, there are places where it deviates, and this effect probably becomes more pronounced throughout an individual's lifetime. This means that the potential for meme diversity, though constrained by host need, is open-ended.

It can be useful to think in terms of not only an individual's memetic fitness landscape, but also a societal memetic fitness landscape, wherein minute-to-minute fluctuations (such as the urgent but temporary need to find a bathroom) are averaged out, and all frontiers of human endeavour are incorporated.

3.3 The landscape is sculpted by the need for worldview coherence

A need that seems to surface to the forefront (have a large impact on the focus) when other needs are not pressing is the need to connect fragmented representations of the world into a logically-consistent worldview. Since our ability to make predictions and evaluate possible plans of action hangs on the accuracy of this worldview, the survival value of such a tendency is clear. McCulloch and Pitts [1943] showed that networks made of neuron-like components that perform the logical operations AND, OR, and NOT are theoretically capable of computing any Turing machine-computable function. In connectionist-type systems, logical relations are

represented implicitly as constraints on the possible states of a system, and computation proceeds through settling into a solution that satisfies many constraints rather than explicitly calculating a function [Rumelhart and McClelland 1986]. This is accomplished through modification of association strengths amongst the components of the system, and the process is referred to as annealing or relaxation.

3.4 Hard-wired selection

To the extent that the memetic fitness landscape echoes the shape of the biological fitness landscape, to which we have been adapting since life began, cultural selection is built right into our architecture. Our perceptual and cognitive systems are wired up such that they are primed to focus on and highlight those aspects of external reality that are relevant to our survival (or were in the past). The mental representations we form reflect that bias [e.g. Hubel & Wiesel 1979; Marr 1982]. Second, the associative organization of memory constrains variation-generating operations. So selection is built right into our hardware.

3.5 Malleable forms of selection

In order to create, or even just understand, a new meme, there has to be a conceptual framework from within which it will make sense, and a need, or niche, for it. Therefore, any relevant 'precursor memes' must first be assimilated [Wallas 1926]. This constraint amounts to a malleable, or plastic, form of selection on new memes. Selection can also occur after a representation has been internalized but prior to being phenotypically expressed. For example, mentally simulating what would happen if an idea were implemented can weed out unworthy ideas [Nersessian 1993]. The success of mental simulation varies with the accuracy of ones' internalized model of the world, but it provides at least a rudimentary form of selection. Finally, selection can operate through biased transmission; that is we choose to imitate certain individuals and not others [Boyd & Richerson 1985].

4. *The generation of variation in culture*

4.1 Strategy guides trajectories through the memetic fitness landscape

The existence of an open niche does not guarantee that the niche will ever be found. In biology the process by which this happens is largely random. Though most mutations and recombinations are detrimental, so many variants are generated that it is not necessary to be clever about how they are generated. We could say that biological evolution is a more breadth-first search algorithm than cultural evolution because it relies primarily on massive parallelism rather than strategy.

In culture, on the other hand, variants are generated strategically. We could say that cultural evolution is a more depth-first approach to searching a space of possibilities. The trajectory of a stream of thought is constrained by connections between representations that are similar or spatiotemporally related [Schank 1983], which increases the probability that an advantageous variant is found. For example, when considering the problem of having to get out of your car every day to open the garage door, you would not think about doilies or existentialism, but concepts related to the problem - electricity, human laziness, and various openers you have encountered before. During creative thought, memes evoke or activate one another, altering or strategically (though not necessarily consciously) manipulating them, a process that is said to involve pattern completion, constraint satisfaction [Rumelhart & McClelland 1986], and the tweaking, blending, redescription, abstraction, and recoding of representations [Hofstadter 1985; Holland et al. 1986; Karmiloff-Smith 1986; 1992; Ram 1993; Clark & Thornton in press]. Neurophysiological evidence suggests that creating new contexts for representations, that is manipulating them, involves hippocampal binding or linking [Squire 1992], and synchronization [Klimesch 1995], of features encoded by distributed cortical cell assemblies.

To sum up: fuelled by need and constrained by association we carve out trajectories through meme space, and because the fitness landscape that guides this process is fractal, every time that landscape steers the production of a new meme (or even just a slight variant of a preceding meme), the new meme in turn redefines the landscape, and so on, recursively.

4.2 Sparse distributed memory as a platform for generating variation

Sparse, distributed memory [Kanerva 1988], or SDM, is a mathematical model of the mechanics underlying the storage and retrieval of memories. It was motivated by the desire to understand how memory provides conscious experience with a thread of continuity via the spontaneous sequential activation of concepts or experiences that are related to one another, sometimes superficially, and other times through resemblances that are highly abstract or metaphorical.

Kanerva draws an analogy between the focus and a combined address-datum register in a computer; they both contain data and serve as a pointer to memory, and can both read from and write to memory. An instant of experience is encoded in the focus by a high-dimensional vector of difference relations, or bits, that represent the presence or absence of some feature, and the mathematics generalizes such that a pattern of bits can represent a value along some dimension. The Hamming distance between two memes is the number of bits that differ. (So the Hamming distance between 11111 and 11100 is two.) Since each meme has an antipode (for example, the antipode of 11111 is 00000), the space of all possible memes can be visualized as a sphere. The address of a meme is the information pattern that specifies where the meme is stored.

If L is the number of possible features in a meme, the number of possible memes is 2L. Assuming L is large the size this space is enormous, so the memory is sparse in that it stores only a small fraction of the set of all possible memes. For example, to construct a SDM with L=1,000, then out of the 21,000 possible addresses, a workable number of them, say 1,000,000, are chosen at random to be actual storage locations. The number of memes at Hamming distance k away from any given meme is equal to the binomial coefficient of L and k, which is well approximated by a Gaussian or normal curve. If meme X is 111...1 and its antipode 000...0, and we consider meme X and its antipode to be the 'poles' of the hypersphere, then approximately 68% of the other memes lie within one standard deviation (sqrt[L]) of the 'equator' region between these two extremes. As we move through Hamming space away from the equator toward either Meme X or its antipode, the probability of encountering a meme falls off sharply by the proportion sqrt[L]/L. In our example, the median distance from one location to another is 424 bits, and 99.8% of stored memes lie between 451 and 549 bits of any given

location.

A computer reads from memory by simply looking at the address in the address register and retrieving the item at the location specified by that address. The sparseness of the SDM prohibits this kind of one-to-one correspondence, but it has two tricks up its sleeve for getting around this problem.

First, it feigns content addressability, as follows. The particular pattern of 1s and 0s that constitutes a meme causes some of the synapses leading out from the focus to be excited and others to be inhibited. The locations where memes get stored are memory neurons, and the address of a neuron amounts to the pattern of excitatory and inhibitory synapses from focus to memory that make that neuron fire. Activation of a memory neuron causes the meme to get written into it. Thus there is a systematic relationship between the memes' information content and the locations they activate.

Second, since the probability that the ideal address for storing a meme corresponds to an actual location in memory is vanishingly small, storage of the meme is distributed across those locations whose addresses lie within a sphere (or more accurately, hypersphere) of possible addresses surrounding the ideal address. The radius (in Hamming metric) of this sphere is determined by the neuron activation threshold. Each location participates in the storage of many memes. In this example we assume that 10,000 memes have been stored in memory. Each meme is stored in 1,000 (of the 1,000,000 possible) locations, so there are approximately 10 memes per location. The storage process works by updating each of the L counters in each location; to store a 1 the counter is incremented by 1, and to store a 0 it is decremented by 1. These nearly one million operations occur in parallel.

If after a meme, say meme X, is stored, the individual's attention is directed toward external stimuli, then nothing is retrieved from memory. But to the extent that memory contributes to the next instant of awareness, the storage of X activates retrieval of not only X itself but all the other memes that have been stored in the same locations. The next meme to be encoded in the focus, X', is found by determining the best match; that is, by averaging the contributions of all retrieved memes feature-by-feature. Whereas the 1,000 retrieved copies of X (and memes similar to X) reinforce one another, the roughly 10,000 other retrieved memes are statistically likely to cancel one another out, so that X' ends up being

similar to X. Though X' is a reconstructed blend of many memes it can still be said to have been retrieved from memory. X' can now be used to address the memory, and this process can be reiterated until it converges on meme Y that satisfies a current need. The closer Y is to X, the faster the convergence. In our example, assuming r = 425, if X and Y are more than 200 bits apart Y is unlikely to be retrieved, but if they are 170 bits apart Y will be retrieved in about four iterations.

Keeler [1988] has shown that SDM is a superset of Hopfield-type and connectionist models of autoassociative or heteroassociative memory. SDM is used here because its formulation lends itself to an understanding of the mechanics of phenomena we are interested in. Because of how the dynamics emerges from the statistics, rather than from a central executive, it can cope with creative and seemingly unmechanical cognitive phenomena such as wordplay or slips of the tongue. Moreover it is ideally suited to handle the problem of sequential access, which will become relevant when we look at how an infant establishes a train of thought To model the recollection of a sentence, meme X is simply used as the address to write Y, Y as the address to write Z, and so on. Working memory can be viewed as the memes that lie within a given Hamming distance of the meme in the focus such that they are retrievable within a certain number of iterations. Categorization could involve the identification of a feature schema, and readdressing memes that contain this schema so that their new addresses put them within working memory reach of one another. Kanerva shows that the architecture of common neural components and circuits in the brain are ideally suited to implement a SDM.

In SDM, associations between memes are not explicitly represented as connection strengths but as proximity in multidimensional space. However in the end they amount to the same thing. The smaller the Hamming distance between two memes, the higher the probability that they will be retrieved simultaneously and blended together in the focus (or one after the other in a chain of related thoughts). What allows the memes to be retrieved simultaneously, however, is that they are either stored in the same neurons or in neurons with nearby addresses, which in turn reflects the neurons' connectivity. Thus factors that affect the storage of a meme will affect the retrieval of that meme; the two processes are intimately connected.

5. *The replication and transmission of memes*

Transmission links the memetic processing within an individual to not only memetic processing in other directly-encountered individuals, but processing in individuals they encounter, and so on. The ideas and inventions any one individual produces build on the ideas and inventions of others. This phenomenon is known as the ratchet effect, and its significance is demonstrated in the following example. If you were suddenly dropped into the Australian desert, you probably would not survive for long. However if you were to run into an aborigine who grew up learning desert survival skills from her family and community that had been passed on and improved upon for generations (such as how to find water in obscure places) you might survive for some time[4].

5.1 Internal meme replication via implicit pointers to memory

We saw how, unlike genetic material, memes do not contain instructions for how to make copies of themselves; they replicate when their hosts teach or imitate one another. The memes in a SDM-like memory, however, have a self-replication capacity in the following sense. The pattern of information that constitutes a meme determines which of the synapses leading out from the focus are excited, and which are inhibited -it determines how a pattern of activation flows through the memory network - which in turn determines the locations where the meme is stored. Thus embedded in the neural environment that supports their replication, memes act as implicit pointers to memory. These pointers prompt the dynamic reconstruction of the next meme to be subjectively experienced, which is a variation of (statistically similar to) the one that prompted it. It is in that sense that they self-replicate.

5.2 Transmission is Lamarkian and phenotypically mediated

Internal replication (with variation) makes cultural transmission Lamarkian - modifications acquired since the acquisition of a meme can be passed on to others [Dawkins 1976]. The related point that transmission is phenotypically mediated, as Dennett [1995] points out, makes a 'science of memetics' less daunting. It means that, unlike biologists, we don't have to fully understand the nature of mental representation to study

transmission.

5.3 Any experience can affect transmission

While biological needs affect the focus from the inside, environmental stimuli impact it from the outside. The information-based orientation supports a broader conceptualization of the transmission process than is generally taken. For the purpose of understanding the evolutionary mechanics underlying culture, any interaction between an organism and its environment that impacts the focus is part of this process. It often occurs through imitation of conspecifics [Smith 1977; Bonner 1980; Robert 1990], or guided instruction [Vygotsky 1978; Tomasello et al. 1993], but not necessarily. For example, does it matter whether a child learns to peel a banana by watching her mother, or a monkey, or a cartoon character on TV? What matters is that the child has a mental representation of how to peel a banana. All kinds of interaction with the environment provide us with new representations or alter existing ones, and therefore have the potential to affect the interplay of ideas and emotions that are culturally transmitted.

6. *A tentative scenario for memetic evolution*

We have discussed how memes evolve through selection, variation, replication and transmission. We turn now to how the evolutionary perspective can shed light on the dynamics of mental representations at both the individual and societal levels.

6.1 The origin of life and its cultural analog

The origin of life poses the following paradox: how could something as complex as a self-replicating molecule arise spontaneously? Traditional attempts to explain this entail the synchronization of a large number of vastly-improbable events. Proponents argue that the improbability of the mechanism they propose does not invalidate it because it only had to happen once; as soon as there was one self-replicating molecule, the rest could be copied from this template. However Kauffman [1993] proposes an alternative scenario that does not entail the synchronization of nume-

rous improbable events. He suggests that life arose through the self-organization of a set of autocatalytic polymers. When catalytic polymers interact with one another their average length increases. As their length increases, the number of reactions by which polymers can interconvert increases faster than the number of polymers. Therefore a set of interacting molecules under conditions such as are likely to have existed at the time life began would inevitably reach a critical point where there is a catalytic pathway to every polymer present. Jointly they form a self-reproducing metabolism.

We now ask: What is the cultural analog to the origin of life? One could say it is the point in history when organisms acquired the capacity for social transmission, but as many authors [e.g. Darwin 1871; Plotkin 1988] have pointed out, although transmission is wide-spread throughout the animal kingdom, no other species has anything remotely approaching the complexity of human culture. Donald [1991] argues convincingly that the bottleneck in cultural evolution is the capacity for innovation. Innovation requires more than a kind of awareness that integrates survival needs with environmental affordances, and draws upon memory only to interpret stimuli, or consult a mental map, or recall how some drive was satisfied in the past. It requires an ongoing train of representational redescription. This suggests that the cultural analog to the origin of life was the origin of the first self-perpetuated, potentially-creative stream of thought in an individual's brain.

When an infant has its first experience, there is nothing in memory to draw upon to contribute to that experience; the first meme to occupy its focus does not remind it of anything. Therefore experience is initially driven only by external or internal stimuli, not by memory. Thus the evolution of culture poses a paradox analogous to that of the origin of the self-replicating molecule - how does an infant develop the capacity for a self-sustained train of thought that creatively integrates new experiences with previous ones? Consistent with Kauffman's assertion that the bootstrapping of an evolutionary process is not an inherently improbable event, the 'it only had to happen once' argument does not hold water here because the cultural analog to the origin of life takes place in the brain of every infant.

6.2 Establishing an autocatalytic set of sparse, distributed memories

This section outlines how a SDM-like stream of thought might get established. Let us say that the first meme to occupy an infant's focus and then get stored in memory is a visual experience of its mother in a blue coat. The next is the sound of a dog barking. The Hamming distance between these memes exceeds the maximum for one meme to evoke the memory of another, so the barking does not remind the infant of its mother. Later the infant sees its mother in a red coat. This meme evokes or 'catalyzes' the memory of its mother in a blue coat. To avoid getting stuck in an endless loop wherein 'mother in blue coat' then evokes 'mother in red coat' et cetera, it may form the category 'mother'. However that meme does not remind it of anything, so this stream of thought dies off quickly.

As the infant accumulates memes, the statistical probability that a meme in the focus will activate a meme from storage increases, so the streams of remindings get longer. Eventually the memory becomes so densely packed that any meme that comes to occupy the focus is bound to be close enough in Hamming distance to some previously-stored meme(s) to activate a variant of itself. This marks a phase transition to a state in which, just as with the origin of life, the sequential activation of self-similar patterns is self-propelled; the memes now form an autocatalytic set. The focus is no longer just a spot for coordinating stimuli with action but a forum for the variation-producing operations that emerge naturally through the dynamics of iterative retrieval. The resultant memes evolve along different trajectories toward different basins of attraction, 'specializing' in the fulfilment of one need or another. Those that satisfy the same need compete until one becomes habitual, while those that fulfil different needs are able to coexist within the same host. As with biological speciation, small differences are amplified through positive feedback leading to transformation of the space of viable niches for the evolution of information patterns.

Note that in this example the 'mother' meme is the infant's first category. A simple way of describing this situation is: if the 'mother in blue coat' meme is represented as 111, and the 'mother in red coat' meme is represented as 110, the 'mother' meme can be represented as 11*, where * means either 1 or 0. It is also the infant's first derived meme. That is, it is the first information pattern to enter the focus not

purely by way of external stimuli but through the necessity of a logical operation on previously-stored memes 'in this case an OR gate' which could be realized in the brain via adjustment of connection strengths. The act of categorization projects the original information space, which had n relevant dimensions, onto a new space that has n-1 dimensions (for example, here coat colour is no longer relevant); it effectively makes the space denser. This increases susceptibility to the autocatalytic state. On the other hand, creating new memes by combining previously-stored memes could interfere with the establishment of a sustained stream of thought by increasing the dimensionality of the space, thereby decreasing density. If indeed cross-category blending disrupts conceptual autocatalysis, one might expect it to be less evident in young children than in older children, and this expectation is born out experimentally [Karmiloff-Smith 1990].

Note also that the density of memes necessary to reach and maintain this autocatalytic state will depend on the neuron activation threshold. If the threshold is too high (the hypersphere of potentially activated memes is too small) even very similar memes can not evoke one another, so a stream of remindings, if it happens at all, dies off readily. If the threshold is too low (the hypersphere too large), then any meme will evoke a multitude of others not necessarily meaningfully related to it. Successive patterns in the focus will have little or no resemblance to one another; the system may be catalytic but it is not autocatalytic. The free-association of the schizophrenic [see Weisberg 1986] seems to correspond to what one might expect of a system like this. For memory to produce a steady stream of meaningfully-related yet potentially creative remindings, the threshold must fall within a narrow intermediate range. This is consistent with Langton's [1992] finding that the information-carrying capacity of a system is maximized when its interconnectedness falls within a narrow regime between order and chaos. The situation may turn out to be slightly more complicated; sustaining a creative train of thought may involve not only keeping the activation threshold within a narrowly-prescribed range but dynamically tuning it in response to the situation at hand. This is particularly likely if the memory is not uniformly dense (i.e. clusters of highly-correlated memes) or if different kinds or stages of thought require different degrees of conceptual blending.

Thus we have a plausible scenario for how cultural evolution, like biological evolution, could have originated in a phase transition to a self-

organized web of catalytic relations between patterns.

7. *Viewing psychological phenomena within a cultural evolution framework*

7.1 Mental censors, worldview cohesion and the unconscious

Initially an infant is unselective about meme acquisition, since (1) it doesn't know much about the world yet, so it has no basis for choosing, and (2) its parents have lived long enough to reproduce, so they must be doing something right. However just as importing foreign plants can bring ecological disaster, acquisition of a foreign meme can disrupt the established network of relationships amongst existing memes. Therefore the infant develops mental censors that ward off internalization of potentially disruptive memes. Censors might also be erected when a meme is found to be embarrassing or disturbing or threatening to the self-image [Minsky, 1975]. In the architecture described here this could be accomplished by increasing the activation threshold so as to prevent the content of the focus from assimilating with stored memes. This amounts to a premature termination of the relaxation process.

On the other hand, when the cost of the disruption is outweighed by the potential benefit accrued by a world model that can accommodate the new meme, the threshold would be lowered. Most thoughts seem to have little effect on our understanding of the world at large, but once in a while we experience a meme that significantly modifies our world view. The situation is reminiscent of superconductivity; lowered resistance increases correlation distance, and thus a perturbation to any one pattern can percolate through the system and affect even distantly-related patterns. It would be interesting to determine experimentally whether the 'inductiveness' of our memes, like other self-organizing systems, exhibits the ubiquitous inverse power law [Bak, Tang & Weisenfeld 1988]. Just as in a sand pile perched at the edge of chaos once in a while a collision between two grains will lead to another in just the right chain reaction to generate a large avalanche, occasionally one thought will trigger a chain reaction of others in a way that reconfigures the conceptual network.

The concept of the unconscious has been influential and useful despite the obvious incongruity: how is it that we can consciously discuss

something that is unconscious? What has been called the unconscious may be the fleeting experience of memes that are dynamically reconstructed asin a SDM but which do not readily assimilate with other memes and so get discarded from the focus. In other words, the need for worldview consistency prohibits further computational resources from being spent on trying to integrate what appears to be a nonsensical construction into the memory. Of course there is no reason why a meme that is not immediately integrated into the memory might nevertheless affect the memory; the very process of determining whether it can be assimilated or not might itself have effects that infiltrate the system. This possibility is supported by the finding that subjects' behaviour can be affected by priming material of which they have no recall [e.g. Dunbar & Schunn 1990; Fehrer & Raab 1962]. Subconscious processing of this sort could, in fact, resculpt the memetic fitness landscape in such a way that a previously-discarded meme is more readily assimilated the next time it is encountered.

7.2 Cultural momentum

Despite being derived, directly or indirectly, from human need, memes do not always promote our survival [Greene 1978; Alexander 1980]. As Dawkins [1982] points out, 'It is true that the relative survival success of a meme will depend critically on the social and biological climate in which it finds itself, and this climate will certainly be influenced by the genetic make-up of the population. But it will also depend on the memes that are already numerous in the meme-pool.' Much like runaway selection in biology, once a meme can replicate with variation on the basis of some selection criterion, it can evolve out of the orbit of the need that originated it. We can't help but engage in a stream of thought, spontaneously generating new memes like 'if only such and such had been different...', any more than biological evolution can help but generate new species. This cultural momentum could explain why, despite the intuition that individuals control their streams of thought, creators often express surprise at the sudden appearance of an idea, and deny active effort in its immediate creation [Bowers et al 1990; Guilford 1979; Kubose et al 1980; Wallas 1926]. We seem to control the birth of 'our' ideas only to the extent that we provide a fertile ground for them to be fruitful and multiply' by internalizing relevant background knowledge, identifying new needs, and exposing ourselves to stimuli that help trigger

ideas that fulfil those needs (if you don't like this idea, don't blame me.)

Spurious basins of attraction sometimes arise in recurrent neural networks through the compositional interaction of explicitly-trained attractors [Hopfield 1982]. Cultural momentum may boil down to a phenomenon of this sort. Just because the memetic fitness landscape largely echoes the biological fitness landscape, that doesn't mean that behaviour elicited by memes in spurious basins of attraction arising through representational redescription need always be conducive to survival. Nevertheless a stream of thought could be censored before it elicits harmful behaviour. Streams of thought probably get sidetracked on a regular basis, not just by censors, but by minute-to-minute undulations in the hyperdimensional fitness landscape, that is, change in the relative urgency of the multitude of survival-related or derived needs impacting the focus.

The concept of cultural momentum sheds light on the issue of free will. Those who argue for the existence of a central executive in memory may come to be viewed as the creationists of philosophy and cognitive science. Human will can instead be viewed as the emergent orchestration of needs, stimuli, and retrieved memories impacting the focus, which is subject to cultural momentum and therefore, in a sense, beyond our control.

7.3 The birth of creative ideas

The biologically-inspired model developed here supports a variant of the combination theory of creativity - that new ideas arise through combinations and transformations of old ones [Boden 1991; Koestler 1964]. The aspect of this theory that does not ring true is that it neglects the role of emotions. Here we consider emotions, as well as ideas, to be encoded as information in memes; some components of a meme are simply interpreted by a part of the mind that experiences them as emotion, whereas others are interpreted by parts of the mind that experience them as ideas. Much research on analogy deals with how the structure or 'conceptual skeleton' underlying one idea gets abstracted and applied to another [Gentner 1983; Gick & Holyoak 1983]. We can expand on this general idea by suggesting that many forms of creative expression begin with the (unvoiced) question: What would the pattern of information that encoded the emotion I experienced during this particular event look like if ex-

pressed through the constraints of that medium? The existence of inherent limitations on how a pattern could be translated from one domain to another is consistent with the frequent observation that creativity involves both freedom and aesthetic constraint. Thus all creativity is directly or indirectly derived from experience in the world, and since the mathematics underlying this world, the set of all natural functions, is a small subset of all possible functions, the constraints that guide creation are not arbitrary but objective and familiar; for example a drum beat of a song might echo a heartbeat, when the rhythm and chord progression are reminiscent of the sound of someone sobbing we feel sad, and we hear the wrong note even if we have never heard the song before.

It makes sense to expect that a meme or meme complex that has been censored would be vulnerable to being targeted as an area where worldview cohesion could be increased. Since at the time the censored material was experienced it was prohibited from forming associations to obviously-related memes, it in turn can not be retrieved through these sorts of expected or straightforward associations. It can only be retrieved via 'backdoor entrances', that is through associations that reflect structural congruity at an abstract level. Thus a musician may come to habitually funnel patterns encountered in a variety of domains - and particularly censored material - through modules that filter out hitchhiker and enabler features, and adapt the core features (or feature schemata) to the constraints of music. It is in this repackaged format that they are integrated into the memory at large, and it is through this process that the creator establishes a sense of control over memes that were previously 'off-limits'. In an influential paper on the relationship between DNA polymorphism and recombination rates, Begun and Aquadro [1992] suggested that genetic hitchhiking may have significant evolutionary impact:

'This correlation suggests that levels of neutral variation in many of the gene regions for which variation has been measured have been reduced by one or more hitchhiking events. Provided that new selectively favoured mutation goes to fixation before another advantageous mutation arises close to it, each fixation will be surrounded by a 'window' of reduced polymorphism, the relative size of which is proportional to the rate of recombination for that region of the genome.'

The general idea presented here translates nicely to cognition: if a meme goes to fixation in a society due to selective advantage conferred by one or more core features or core-enabler couplings, its hitchhiker

features will also exhibit reduced polymorphism, and the size of the 'window' will vary with the extent to which hitchhiker features are conceptually bound to that meme. One could argue that recreation is the recreation of information patterns in different domains from the ones in which they were originally encountered, thereby filtering out conceptual prejudices that reflect nothing more than mechanical constraints or historical legacies of the original domain. Play, intellectual pursuits, and other creative endeavours are then algorithms for achieving a state of conceptual linkage equilibrium through mental operations that, like genetic recombination, increase polymorphism by reducing fixation through hitchhiking.

The account of creativity proposed here may seem too simple to explain the seemingly limitless human potential for creativity, but it may seem less far-fetched when we consider the variety of species produced by biological evolution, which operates without the benefit of strategy. Furthermore, raw materials for the creative process may be acquired in exceedingly subtle ways. It is conceivable that you might watch a stream flow and without your consciously thinking, 'It flows... things can flow... I could even, in some sense, adopt a more flowing approach to life', the experience might be reconfiguring your memetic infrastructure in a way that makes you more easygoing. I am not making any claims about the extent to which experiences of this sort affect us or even whether they occur at all, but rather suggesting that we not prematurely place a lid on the kinds of processes that could affect a network of representations and thereby affect creation and transmission. A theory of mind and society that can account for phenomena like poetry is not easily achieved.

8. *Why are creativity and culture unique to humans?*

Recall that in order for a network of memes to reach an autocatalytic state, the activation threshold must be calibrated to fall within a narrow range to achieve a delicate balance between the capacity for semantic continuity on the one hand and creative association on the other. The penalty for having too low a threshold would be very high; thoughts would not necessarily be meaningfully related to one another, and thinking would be so muddy that survival tasks are not accomplished. Too high a threshold, on the other hand, would not be life-threatening. The

focus would virtually always be impacted with external stimuli or internal drives such as hunger; memory would be pretty much reserved for recalling how some goal was accomplished in the past. A stream of experience that involved the iterative reorganization of stored memes would likely die out before it produced something creative. This may be the situation present in most brains on this planet, and the reason that apes are limited to episodic memory [Donald 1991].

For animals, the benefits of a sustained train of thought would be minimal because they have neither the vocal apparatus nor the manual dexterity and freedom of upper limbs to implement complex ideas. No matter how brilliant their thoughts were it would be difficult to do something useful with them. Moreover, in an evolutionary line there is individual variation, so the lower the average activation threshold, the higher the fraction of individuals for which it is so low that they do not survive. It seems reasonable to suggest that apes are not a priori prohibited from evolving complex cognition, but that there is insufficient evolutionary pressure to keep the activation threshold low enough to sustain a stream of thought, or to establish and refine the necessary feedback mechanisms to dynamically tune the threshold according to the degree of conceptual fluidity needed at any given instant. It may be that humans are the only species for which the benefits of this capacity outweigh the cost.

9. *A memeus-eye view*

In this section we examine a hypothetical and admittedly speculative scenario for how evolutionary concepts borrowed from biology might apply to the dynamics of a specific meme. We then look briefly at a computer model of some of these phenomena.

9.1 The ontogeny of a meme

One day a classroom bully named Tony put his arm around a girl named Memela. Memela felt threatened by Tony's advance. Her first impulse was to get angry but she censored this reaction. The need for worldview cohesion motivated a desire to escape the restrictive power of this censor and find a 'backdoor vent' for her anger. Eventually she whispered to a

classmate: 'Tony Testosteroni made a pass at me.' Thus began the era of 'Memela's meme'.

Memela's meme can be traced back to a number of precursor or 'protomemes', many of which originated in the minds of family and friends and were subsequently transmitted to Memela. These protomemes provided an environmental niche in which the joke could flourish. Another reason that Memela was predisposed to produce the joke is that her brain spontaneously exerts a high degree of control over its activation threshold. Her ability to find unusual associations by increasing hypersphere radius, and subsequently refine a train of thought by decreasing the radius, facilitates word play such as the establishment of epistatic relationships between semantic and syntactic components of a meme. The semantic applicability of the word 'testosterone' to Tony's aggressive behaviour, and the alteration of this word to make it sound Italian and echo the syntax of (i.e. rhyme with) the Italian first name, contribute to the humour of Memela's meme. Its relation to the highly censored subjects of aggression and sexuality may also have added to its appeal.

9.2 A day in the life of memelaus meme

Conceptual epistasis provokes laughter which draws attention to a meme. Memela's meme took full advantage of this. When Memela told the joke to one of her classmates the ensuing laughter attracted a small crowd of other potential hosts, and within an hour MemelaUs meme had reached most of the students in the classroom. Their willingness to invest time acquiring this meme was a smart move; the joke not only provided amusement, but it proved to be a useful precursor to the formulation and understanding of subsequent jokes in this social circle. Some were direct descendants of Memela's meme, such as jokes along the lines of 'What's for lunch - rigatoni a la testosteroni?' Others were more distantly related, such as nicknames for other classmates that arose because Memela's meme had activated the general concept of nicknaming. Across the classroom, ideas that pertained to nicknaming came to constitute a highly active region of conceptual space analogous to the uncharacteristically high level of polymorphism in a small portion of the human genome known as the major histocompatibility complex [Hughes & Nei 1988].

By the time recess ended it had migrated extensively through the school population. There were certain subpopulations of individuals that

it failed to penetrate, such as social outcasts who were excluded from much of the memetic exchange. These individuals exhibited a cultural version of the Founder Effect [Holgate 1966] - reduced variation in a small population due to genetic drift. Memela's meme also failed to reach students who engaged in projects that took them away from the playground at recess. However since these individuals had had less opportunity to witness Tony's behaviour, they had less need to diffuse their fear of him. Thus even if they had heard Memela's meme they did not possess the necessary precursor memes to fully appreciate it; the prerequisite memetic niches were not in place. At any rate despite its failure to reach these subgroups, Memela's meme experienced a high degree of 'memetic fitness'. It migrated far beyond the classroom in which it was originally formulated, reconfiguring networks of representations in ways that affected the subsequent thought and behaviour of a number of individuals. The telling of this meme and its various incarnations constituted an act of memetic altruism between like-minded individuals analogous to the biological altruism that occurs between genetically-similar individuals. It played a small role in an ongoing network of positive reciprocal interactions through which there emerged a memetically-derived social structure wherein individuals that regularly generated pleasurable or powerful memes came to be observed carefully and imitated frequently, while other individuals were ignored. Thus the fate of Memela's meme and its descendants reflected the social and psychological dynamics of an entire society of interacting individuals.

Memes can fool potential hosts into identifying with them or believing they are wanted or needed via the creation or fortuitous exaptation of supporting memes that we already identify with or that represent things we need or want (as advertisers are well aware). Thus the greater the extent to which we identify with or value ourselves in terms of the memes (including those that pertain to the self) and implemented artifacts we possess or lack, the more vulnerable we are to ever-more-seductive forms of persuasion and advertising which tie up time, energy, and resources that could be applied toward other goals. Like the other students in the classroom, Tony was affected by the sound of laughter advertising the presence and amusement value of Memela's meme. However it wasn't until someone told him that it was a joke was about him that he felt willing to do almost anything to hear it.

Upon hearing the nickname Tony felt ridiculed. One way to defend

oneself against painful or manipulative memes is to construct what Dennett [1991] refers to as a 'meme-immunological system'; that is, formulate new memes specifically to deflect 'memetic antigens'. In the case of Memela's meme this could be something along the lines of 'That nickname is silly and stupid.' However constructing 'memetic antibodies' of this sort is time-consuming, and like any immunological response it has to be repeated every time the outside agent evolves a counter-response. Perhaps this explains the purported benefits of 'transcending the ego' [e.g. Walsh & Vaughan 1993], which can be taken to mean getting in touch with that part of yourself that presumably existed before your mind was colonized by memes.

9.3 Meme and variations: a computer model of cultural evolution

Meme and Variations, or MAV [Gabora 1995] is a computer model of a society composed of interacting neural network-based agents. Unlike other such models that combine biological and cultural evolution [e.g. Ackley 1994; Spector & Luke 1996] these agents don't have genomes, and neither die nor have offspring, but they can invent, implement, and imitate memes. MAV successfully evolves patterns of information through cultural implementations of variation, selection, and replication, and exhibits phenomena observed in biological evolution such as: (1) drift (2) epistasis increases time to reach equilibrium, (3) increasing frequency of variation-generating operations increases diversity, and (4) although in the absence of variation-generating operations meme evolution does not occur, increasing variation-generation much beyond the minimum necessary for evolution causes average fitness to decrease. MAV also addresses issues specific to cultural evolution, such as the effects of mental simulation, imitation, and strategy. Perhaps the most interesting finding it yielded was that although for the society as a whole the optimal creation-to-imitation ratio was approximately 2:1, for the agent with the fittest memes, the less it imitated (i.e. the more computational effort reserved for creation) the better.

MAV will hopefully serve as a stepping stone to more advanced models of memetic evolution. Of particular interest will be models that: (1) like Tierra, a model of biological evolution [Ray 1991], harness the power of evolution to explore and transform an open-ended space of possible patterns, but (2) explore the space strategically on the basis of

accumulated knowledge rather than at random[5], and (3) have fitness landscapes that emerge through the needs of the agents within the constraints of their environment [as in Maes 1991], and (4) have agents that must learn for themselves which memes, when implemented, best satisfy each of their various needs. Mathematical models of culture are too minimal to cope with the open-ended diversity of culturally-derived information (variation is generally restricted to trial and error learning or transmission error) let alone address the numerous intra-individual factors that undoubtedly have emergent inter-individual consequences, such as how representations are grounded in experience and how they are stored, retrieved, and implemented. Models of individual intelligence and creativity, on the other hand, lack transmission and replication. Although this research may not explicitly attempt to address group processes it typically focuses not on the sorts of simple inferences and creative acts that a person raised alone in the wild would be capable of, but on complex acts such as story comprehension, that might be unlikely to develop in isolation. With the advent of massively parallel computers it is becoming increasingly feasible to place computational models of individual creativity and problem-solving in a cultural context. This approach could provide insight into not only problems pertaining to representation and culture, but evolution in general, through comparison with biology. For example, the question of why there is so much redundancy in the genetic code has generated much discussion which may also apply to the question of why there are redundant mental maps in the brain; both may reflect constraints on the nature of an information-evolving code.

10. *Summary*

This paper presents a theoretical framework for viewing mental representations as memes, a perspective that emphasizes their evolution through variation and transmission, and de-emphasizes individual 'hosts'. Although the cultural evolution of memes operates through very different mechanisms from those of biology, culture is the only system comparable to biology, because it is the only other system to exhibit the imperative features of evolution - adaptive exploration and transformation of an information space through variation, selection, and transmission. All pattern in the information we encounter can be traced to either (1) the

physical constraints and self-organizing properties of matter, (2) biological evolution, (3) cultural evolution, or (4) interactions between these causal principles.

One important difference between the two forms of evolution is that culture is less random - new patterns have a greater-than-chance probability of being more fit than their predecessors. The reason for this is interesting. Since memes (unlike genomes) do not come packaged with instructions for their replication, they must rely on the pattern-evolving machinery of our brains to do it for them. Ironically, this state of dependence enhances their proliferative potential, because the machinery they depend upon constructs and continually updates mental models of its world - that is, internalizes and weaves together fragments of its fitness landscape - and uses these model-of-the-world patterns to guide the creation, assimilation, and implementation (phenotypic expression) of other patterns. This situation fosters a continuous coevolutionary interplay between pattern and landscape, and this is one reason why culture can evolve faster than biology.

Cultural evolution presents a puzzle analogous to the origin of life: the origin of a self-sustained stream of potentially-creative thought in an infant's brain. The idea that life originated with the self-organization of a set of autocatalytic polymers suggests a possible mechanism for how this comes about. Once a threshold density of assimilated memes is surpassed, any meme that occupies the focus is close enough in Hamming distance to evoke or 'catalyze' the spontaneous retrieval or creative reconstruction of a statistically similar meme, thus the memes form an autocatalytic set. Note that this macroscopic account suggests an explanation for only that aspect of human consciousness that differentiates us from other 'experiencers'; it does not address the mystery of 'raw awareness' that some say characterizes not only our experience but that of a cow or a mosquito or even a thermostat [e.g. Chalmers 1996]. Whether or not this specific theory turns out to be correct, it illustrates how the analogy to biology can focus our study of culture by providing a scaffold around which explanatory theories can be built.

The autocatalytic theory of consciousness suggests tentative explanations for cognitive phenomena such as cultural momentum and creativity, and why they are virtually unique to humans. These are put forth as examples of new perspectives that are achieved by viewing cognition from a cultural-evolution framework. If we are to take seriously the idea

that culture is an evolutionary process, we can look to evolution to provide the kind of overarching framework for the humanities that it provides for the biological sciences. This approach may put us on the road to understanding the pervasiveness, diversity, and adaptive complexity of the cultural debris that surrounds and infests us.

University of California

Acknowledgements: I would like to thank David Chalmers, Pentti Kanerva, Bruce Sawhill, Bill Sethares, and Patrick Tufts for thoughtful comments on the manuscript.

NOTES

1. Note that by 'causal principle' I mean something that generates useful descriptions, rather than a 'law'.
2. Though viruses are unique in the biological world in that they rely on hosts to replicate, we will consider viral evolution an anomalous offshoot of biological evolution, because: (1) the evolving patterns of information are encoded as sequences of nucleotides, (2) variation is through mutation and recombination, and (3) transmission and selection are mediated through genotype.
3. Or one could argue that the 'selection' of matter over antimatter, and its subsequent amplification and variation, constitutes yet another form of evolution.
4. This example is a variation of one transmitted by R. Boyd [pers. com.]
5. MAV has this property to some extent; a more sophisticated example is Copycat, a model of analogy-building, [Mitchell 1993].

REFERENCES

Alexander R.D. (1980), *Darwinism and human affairs*, Pitman.

Ackley D. (1994), A Case for Distributed Lamarkian Evolution. In *Artificial Life III*, Addison-Wesley.

Bak P., Tang C., & Weisenfeld K. (1988), Self-organized criticality. *Phys. Rev.* A 38-364.

Bateson G. (1972), *Steps to an ecology of mind*, Ballantine Books.

Begun D.J. & Aquadro C.F. (1992), Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster.

Nature **356**, pp. 519-520.
Boden M.A. (1991), *The creative mind: Myths and mechanisms*: Cambridge University Press.
Bonner J.T. (1980), *The evolution of culture in animals*: Princeton University Press.
Bowers K.S., Regehr G., Balthazard C. & Parker K. (1990), Intuition in the context of discovery. *Cognitive Psychology* **22**, pp. 72-110.
Boyd R. & Richerson P.J. (1985), *Culture and the evolutionary process*: The University of Chicago Press.
Brooks V. (1986), *The neural basis of motor control*: Oxford University Press.
Campbell D.T. (1987), *Evolutionary epistemology*. In Evolutionary epistemology, rationality, and the sociology of knowledge, eds. G. Radnitzky and W.W. Bartley, pp. 47-89, Open Court.
Cavalli-Sforza, L.L. & Feldman, M.W. (1981), *Cultural transmission and evolution: A quantitative approach*: Princeton University Press.
Chalmers D.J. (1996), The conscious mind: *In search of a fundamental theory*: Oxford University Press.
Clark A. & Thornton C. Trading spaces: Computation, representation, and the limits of uninformed learning. *Forthcoming in Brain and Behavioral Sciences.*
Darwin C. (1871), *The descent of man*: John Murray Publications.
Dawkins R. (1976), *The selfish gene*: Oxford University Press.
Dawkins R. (1982), *The extended phenotype*: W.H. Freeman and Company.
Dennett D.C. (1995), *Darwin's dangerous idead, Little*, Brown and Company.
Donald M. (1991), *Origins of the modern mind*: Harvard University Press.
Dunbar K. & Schunn C.D. (1990), *The temporal nature of scientific discovery*: The roles of priming and analogy. In: *Proceedings of the Twelfth Annual Meeting of the Cognitive Science Society*: pp. 93-100.
Durham W.H. (1991), Coevolution: genes, culture, and human diversity. Stanford University Press.
Fehrer E. & Raab D. (1962), Reaction time to stimuli masked by metacontrast. *Journal of Experimental Psychology* **63**, pp. 143-147.
Gabora L.M. & Colgan P.W. (1990), *A model of the mechanisms underlying exploratory behaviour*. In The simulation of adaptive behaviour, eds. S. Wilson and J.A. Mayer, MIT Press.
Gabora L.M. (1995), Meme and variations: *A computational model of cultural evolution*. In *1993 Lectures in Complex Systems*: Addison-Wessley.
Gentner D. (1983), Structure-mapping: A theoretical framework for analo-

gy. *Cognitive Science* **7**(2).
Gick M.L. & Holyoak K.J. (1983), Schema induction and analogical transfer. *Cognitive Psychology* **15**, pp. 1-38.
Gould S.J. (1991), *Bully for brontosaurus*: reflections in natural history. W.W. Norton & Company, pp. 63-66.
Gould S.J. & Vrba E.S. (1982), Exaptation - a missing term in the science of form. *Paleobiology* **8**(1), pp. 4-15.
Greene P.J. (1978), From genes to memes? *Contemporary Sociology* **7**, pp. 06-709.
Guilford J.P. (1979), Some incubated thoughts on incubation. *Journal of Creative Behaviour* **13**, pp. 1-8.
Hebb D. (1949), *The organization of behaviour*: Wiley and Sons.
Hofbauer J. & Sigmund K. (1988) *The theory of evolution and dynamical systems*: Cambridge University Press.
Hofstadter D.R. (1985), *Variations on a theme as the crux of creativity*. In *Metamagical Themas*, pp. 232-259, Basic Books.
Holgate P. (1966), A mathematical study of the founder principle of evolutionary genetic. J. Appl. *Probl.* **3**, pp. 115-128.
Holland J.K. (1975), *Adaptation in natural and artificial systems*: University of Michigan Press.
Holland J.H., Holyoak K.J., Nisbett R.E. & Thagard P.R. (1986), *Induction*: MIT Press.
Hopfield J.J. (1982), Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* (Biophysics) **79**(8), pp. 2554-2558.
Hubel D.H. & Wiesel T.N. (1979), Brain mechanisms and vision. *Scientific American* **241**(3), pp. 150-62.
Hughes D.M. & Nei M. (1988), Patterns of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, *Nature* **335**, pp. 167-170.
Hull C.L. (1943), *Principles of behaviour*: Appleton-Century-Crofts.
Kanerva P. (1988), *Sparse distributed memory*: MIT Press.
Karmiloff-Smith A. (1986), From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* **23**, pp. 95-147.
Karmiloff-Smith A. (1990), Constraints on representational change: Evidence from children's drawing. *Cognition* **34**, pp. 57-83.
Karmiloff-Smith A. (1992), *Beyond modularity*: A developmental perspective on cognitive science: MIT Press.
Kauffman S. (1993), *Origins of order*: Oxford University Press.
Keeler J.D. (1988), Comparison between Kanerva's SDM and Hopfield-

type neural networks. *Cognitive Science* **12**, pp. 299-329.
Klimesch W. (1995), Memory processes described as brain oscillations. *Psycoloquy*, 6.06.
Koestler A. (1964), *The act of creation*: Picador.
Kojima K. & Schaffer H.E. (1967), Survival processes of linked mutant genes. *Evolution* **21**, pp. 518-531.
Kubose S.K. & Umenoto T. (1980), Creativity and the Zen koan. *Psychologia* **23**(1), pp. 1-9.
Langton C.G. (1992), *Life at the edge of chaos*. In *Artificial life II*. eds. C.G. Langton, C. Taylor, J.D. Farmer & S. Rasmussen, Addison-Wesley.
Luchins A.S. (1942), Mechanization in problem solving. *Psychological Monographs* **54**, No. 248.
Lumsden C. & Wilson E.O. (1981), *Genes, mind, and culture*: Harvard University Press.
Maes P., ed. (1991), *Designing autonomous agents*: Theory and practise from biology to engineering and back: MIT Press.
Mandelbrot B.B. (1982), *The fractal geometry of nature*, W.H. Freeman and Company: San Francisco
Marr D. (1982) *Vision*: Freeman.
McCulloch W.S. & Pitts W. (1943), A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**, pp. 115-133.
McFarland D.J. & Sibly R.M. (1975), The behaviourial final common path. *Philosophical transactions of the London Royal Society* **270B**, pp. 265-93.
Miller G.A. (1956), The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**, pp. 81-97.
Minsky M. (1985), *The Society of Mind*: Simon and Schuster.
Mitchell M. (1993), *Analogy-making as perception*: MIT Press.
Nersessian N. (1993), *In the theoretician's laboratory*: Thought experimenting as mental modelling. In *PSA 1992*, vol. 2, eds. D. Hull, M. Forbes, & K. Okrulik. PSA.
Palmer M.W. (1992), The coexistence of species in fractal landscapes, *American Nature* **139**, pp. 375-397
Pinker S. (1995), *The language instinct*: Harper Perrenial.
Plotkin H.C. (1988), *The role of behaviour in evolution*: MIT Press.
Popper K.R. (1963), *Conjectures and refutations*: Routledge & Kegan.
Pribram K.H. (1974), *How is it that sensing so much we can do so little?* In The neurosciences third study program, eds. F.O. Schmidt & F.G.

Worden: MIT Press.
Quillian M.R. (1968), *Semantic memory*. In Semantic information processing, ed. M. Minsky: MIT Press.
Ram A. (1993), *Creative conceptual change*. Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, pp. 17-26.
Ray T. (1991), *An approach to the synthesis of life*. In *Artificial life II*. eds. C.G. Langton, C. Taylor, J.D. Farmer & S. Rasmussen: Addison-Wesley.
Robert M. (1990), Observational learning in fish, birds, and mammals: A classified bibliography spanning over 100 years of research. *Psych Record* **40**, pp. 289-311.
Rumelhart D.E. & McClelland J.L. eds. (1986), *Parallel distributed processing*: Bradford/MIT Press.
Schank R.C. (1983), *Dynamic memory*: Cambridge University Press.
Scheuring I. & Riedi R.H. (1984), Application of multifractals to the analysis of vegetation pattern, *Journal of Vegetation Science* **5**, pp. 489-496
Schuster P. & Sigmund K. (1983), Replicator dynamics. *Journal of Theoretical Biology* **100**, pp. 533-38.
Shannon C.E. & Weaver W. (1963), *The mathematical theory of communication*: University of Illinois Press.
Smith W.J. (1977), *The behaviour of communicating*: Harvard University Press.
Spector L. & Luke S. (1996), *Culture enhances the evolvability of cognition*. In *Proceedings of the 1996 Cognitive Science Society Meeting*.
Squire L.R. (1992), Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* **99**, pp. 195-231.
Thagard P. (1980), *Against evolutionary epistemology*. PSA 1980, ed. P.D. Asquith & R.N. Giere, pp. 187-96.
Tomasello M. Kruger A.C. & Ratner H.H. (1993), Cultural learning *Behavioral and Brain Sciences* **16**, pp. 495-552.
Van Loocke Ph. (1991), Study of a neural network with a meta-layer, *Connection Science* **4**, pp. 367-379
Von Neumann J. (1966), *Theory of self-reproducing automata*: University of Illinois Press.
Vygotsky L.S. (1978), *Mind in society*: The development of higher psychological processes, eds. M. Cole, V. John-Steiner, S. Scribner & E. Souberman: Harvard University Press.
Wallas G. (1926), *The art of thought*, Harcourt: Brace & World.
Walsh R. & Vaughan F. (1993), *Paths beyond ego*, Jeremoy P.:

Tarcher/Perigee.
Weisberg R.W. (1986), *Creativity: Genius and other myths*: Freeman.