A Predictive Theory of Mental Evolution and Its Consequences

A Predictive Theory of Mental Evolution and Its Consequences

By

David A. Steinberg

Cambridge
Scholars
Publishing

A Predictive Theory of Mental Evolution and Its Consequences

By David A. Steinberg

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-6507-6
ISBN (13): 978-1-5275-6507-4

CONTENTS

CHAPTER I

INTRODUCTION

Introduction

The topics explored in these collected articles follow logically and rigorously from a single observation—the external world can be known only as an internally constructed model. They include the details of sequentially constructing such a model and its resulting expression as a theory of mental evolution with verified predictions. In addition to its integrated description of the mind and its predictions of the archaeology of human phylogeny, the theory has consequences for the nature and temporal origin of consciousness and language, the commonalities of alternate intelligences, the nature of mathematical forms, and the origin and nature of time. It also has implications for the limits of scientific knowledge and modern research in neurology and neuroscience.

The articles present an ordinal and predictive theory of mental evolution and its consequences. The theory is ordinal because the position of each stage of mental development is determined by its relation to its predecessor and successor. Specifically, each stage is an explicit generalization of the one that precedes it, and the theory is ordered by the sequence of these generalizations. Consequently, each level is defined exclusively in terms of levels already defined, so that the resultant precision and clarity allow specific verifiable and falsifiable predictions for each stage of evolution. The theory is unique in many ways, but most importantly, it is scientific. The subject of the mind has traditionally been within the purview of philosophy where the lack of unambiguous definitions prevents recognition of an evolutionary sequence of mental capacities and the possibility of verifiable or falsifiable predictions.

Not only are philosophical ideas such as consciousness undefined but there seems to be little recognition that a definition is useful. Yet without such definitions there can be no coherent discussion about the origin of mental capacities and how they are related. This theory attempts to remedy that situation, and in the process, reclaims territory for science that

previously has been ceded to philosophy. At the very least, it offers axioms and unambiguous definitions with which to disagree.

Investigations at the convergence of science and philosophy can create difficulties from both perspectives. As a general rule, natural scientists dismiss philosophical inquiry and its traditional foci as imprecise and non-heuristic. On the other hand, readers theoretically most interested in this theory of mental evolution likely share a philosophical perspective of the mind and asking them to reject their own frame of reference and adopt a scientific system may prove problematic. There is a natural tendency to dismiss ideas critical of one's own and to be ~~s~~keptical of different approaches to problems at the core of one's discipline. I maintain, however, that when possible, a scientific explanation of a phenomenon is preferred and hope that even those most committed to a purely philosophical view of the mind will at least seriously consider this innovative approach with an open one. This flexibility—scientific and philosophical—may be rewarded with novel solutions to scientifically ignored and philosophically unanswerable questions.

This introduction describes the axioms, definitions, and method for a theory of mental evolution and the rationale and nature of the research that engendered it. It does not include references or significant elaboration, since the arguments with documentation appear many times in the following articles. The repetition will, I hope, gradually increase familiarity with, and understanding of, this novel theory.

Despite the clear and concise definitions, rigorous organization, and logical structure of internal reconstruction and its associated theory of mental evolution, there are at least three elements that may contribute to an initial difficulty in grasping it. First, although the ideas of mental evolution are simple, this does not necessarily mean they will be easily understood on first exposure. The simplicity of the stages of evolution, and their implications, is due to their unambiguous definitions and resultant clarity, but this clarity can be obscured by the competing common, imprecise, and unexamined use of many of the terms.

In addition to a possibly unfamiliar usage of common words, this theory introduces a novel key idea—that of a concept-space—which is simply the place where concepts exist. Since we know of concepts in only one form, as ideas in our minds, it is difficult to separate form and content. However, the precise definition of a concept used here is part of an increasingly sophisticated process of defining categories of objects. As such, it is independent of any representation and can exist in many forms. This, then, is the second difficulty. The increasing complexity of the description of categories and the representation of concepts that form the stages of evolution

involve subtle distinctions that are difficult to perceive because they require an unfamiliar perspective and a precise focus on detail.

The final problem in understanding mental evolution is that we do not experience it personally. As infants we are guided rapidly over the developmental stages by fully-evolved adults, leaving no trace in memory. Only when the theory's structure is viewed phylogenetically, with the stages extended over thousands of years, do the subtle changes that form the evolutionary levels clearly emerge.

The rigor and precision this theory applies to a system of mental evolution is what makes it unique and useful. However, our linguistic preconceptions, the subtle distinctions among novel ideas, and the shift from a personal to a phylogenetic perspective can present a challenge. Although synonymous in common usage, *simple* and *easy* are antonyms when applied to most abstract topics.

Internal Model Versus External World

Contrary to the title, and despite the majority of the topics explored, the assembled articles originated neither in a research project whose goal was a theory of mental evolution, nor in an attempt to engage in discussions about neuroscience or the nature of consciousness. My interest initially was in the form of physical laws, i.e., why do they appear as they do and what exactly do they describe? It seemed reasonable then—and still does—to begin at the beginning, with the seemingly unarguable observation that a physical organism can know the external world solely as an internally constructed model. For this reason, I thought of the investigation as an attempt to create a rigorous description of internal reconstruction. By carefully following this process, it should be possible, I thought, to separate and identify aspects of the internal model that may be consequences of physical structure, thereby clarifying the nature of those aspects that are independent of it.

There are at least two reasons to suspect that the nature of the internal model may be influenced by the mechanism and circumstances of its construction. First, the physical structure of any organism requires that sensory information—the raw data of internal reconstruction—be significantly altered and degraded. The essentially continuous external world must be discretized, sampled through a coarse filter (e.g., the senses), and converted into a common form (e.g., bioelectric impulses) before it can be reassembled. Second, the internal machinery of reconstruction reflects adaptations to idiosyncratic geologic, ecologic, and phylogenetic changes. The resulting internal model would be expected to reflect these unique occurrences.

It seems likely that these two factors may limit the correspondence of the internal model with external realities and may have consequences that can be identified by a careful examination of the reconstruction process. But this limitation itself has a profound astrobiological significance: any rigorous consequences that are derived from physical structure necessarily apply to all sentient species that share those structures. These, then, were my thoughts at the outset of this project and the motivation for its pursuit.

Surprisingly, the question of whether the internal model and the external world may be different is not one that has previously been asked by natural scientists, or as far as I am aware, even recognized. The only explanation must be an implicit assumption that the two do not differ. This could be true, however, only if the external and internal worlds are either equivalent—more rigorously, isomorphic—or identical. Clearly it is *prima facie* impossible that they are identical, so the assumption must be that they are isomorphic. An isomorphism is a mapping between two sets that maintains the relationships among their elements. This means that it maps every element of one set to a unique member of the other, and that every element of the second set is the result of such a mapping.

Because of this "one-to-one and onto" nature, each isomorphic function must have an inverse, and isomorphisms are also known as bijective maps. As such, all elements, structures, and laws of the internal model would be seen as functionally congruent to those of the external world. But this cannot be. Technically, the functional relation between external and internal is a composition of two maps—decomposition and reconstruction. If an isomorphism did exist, by definition, there would be an inverse map from internal to external that would consist of an inverse reconstruction map followed by an inverse decomposition map. But since the latter is not one-to-one it cannot have an inverse, and therefore an isomorphism cannot exist.

Though the internal model and external world are demonstrably not identical or isomorphic, this demonstration does not identify the specific nature and significance of their difference. In theory, there is no constraint on how they may differ, but practical survival necessities impose significant extrinsic limitations on the degree to which the two may diverge. For example, miscategorizing a speeding car as a palm tree will expunge the person's genes from the gene pool. Thus, biological evolution "fine tunes" the model through natural selection. But this process has no influence at an abstract level of description where selective pressures do not directly operate (e.g., a mathematical description of the dynamical system of the person and speeding car). The only way to identify a consequence of reconstruction that affects the structure of the internal model—and therefore

is relevant to the internally reconstructed laws of natural science—is by a careful and detailed analysis.

The focus on internal reconstruction has two immediate consequences. First, the stepwise construction of the model clearly defines the origin and nature of its constituent elements. Since the fundamental components of the theories of natural science all have their analogs in this internal model, these foundational scientific principles, as we know them, must originate as a consequence of internal reconstruction. This positioning of biology as fundamental to physics inverts the traditional order of basic sciences. For example, scientists and philosophers have discussed the nature of time and the integration of its scientific and experiential characteristics for millennia without resolution. Careful examination of the internal origin of time, however, reveals a natural hierarchy that integrates the different types of time and effectively eliminates contradictions.

Second, the structural exigencies of any physical organism result in a necessary congruence of internal models. These similarities establish commonalities of diverse sentient species that previously have been purely speculative, but now become amenable to rigorous investigation.

The Study of Mind

Despite the origin of this investigation in a question about the nature of physical laws, as I learned more about diverse disciplines, it became clear that the investigation of internal reconstruction could be described from the perspective of more familiar fields, thereby enhancing its understanding and also resolving important open questions. The two most significant of these applications are to the foundations of neurology and neuroscience and to the archaeologic record of human phylogeny.

In the rest of this section—based on the origin of neurology and neuroscience as derived by John Hughlings Jackson—I interpret many of the classic issues of the mind and brain using the theoretical structure of internal reconstruction to establish an infrastructure for mental evolution. These include a) the specific types of logic that characterize mind and brain functions; b) an interesting complementarity, or dual structure, demonstrated by the logical characterization; and c) the relationship of mental processes to electrical activity of the brain. I then describe how this analysis differs from a traditional philosophical approach and suggest significant strengths of the former and limitations of the latter.

The Science of Neurology

There are many potential definitions of science. For the present purposes, I will define science as a circumscribed body of knowledge that is organized by a predictive and reproducibly testable theory and includes explicit axioms and methods. By these criteria, the science of neurology began with the pioneering work of John Hughlings Jackson, a Victorian physician. He created a new science of the brain using an evolutionary structure inspired, in part, by the currently discredited ideas of Social Darwinism promulgated by Herbert Spencer. The relevant parts of Hughlings Jackson's formulation use the most general characterizations of evolutionary systems, however, and are not related to the objectionable parts of Spencer's work.

Hughlings Jackson's circumscribed body of knowledge is the human nervous system, his tripartite method is explicit, his axiom is that the nervous system is exclusively sensorimotor, and his reproducibly testable and predictive theory is weighted ordinal representation. His method is explained and referenced in the included papers but is relevant to the current discussion only in that it establishes neurology as a science. His axiom, that all functions of the nervous system are sensorimotor—though, at the highest levels, extraordinarily complex—limits the scope of neurology to clinically observable events. And this restriction, after millennia, finally expunged the variable and irreproducible aspects of the mind and soul that had prevented the emergence of a predictive and reproducible neurological science.

His theory of weighted ordinal representation states that the nervous system is a hierarchy of three functional levels—representing, re-representing, and re-re-representing the physical body. Each element of each level contains an entire copy of the preceding stage—an organization that is, by definition, ordinal.

Two principles underlying this theory are that 1) the stages of evolution demonstrate increasing complexity, increasing definiteness, and increasing interconnections, and 2) higher levels exert an inhibitory control over lower levels. The practical clinical consequence of this evolutionary hierarchy is that pathological states of the nervous system are characterized by two types of observable symptoms—positive and negative. Negative symptoms are due to loss of a higher function and positive symptoms result from release of a less organized lower level from inhibitory control. The increasingly complex, integrated, and interconnected levels mean that the afferent function of the nervous system is the creation of a similarly structured sensory map of the external world.

Although Hughlings Jackson excluded the mind from the purview of neurology, he clearly recognized that mental functions exist. He posited a parallel structure of three levels with a correspondence between the brain

and mind that he termed the doctrine of concomitance in which the brain and mind are completely correlated but causally unrelated, i.e., each brain function correlates with a mental function, but is not its cause.

He did not explicitly discuss the functional relationship between these parallel systems of nervous system and mental evolution. However, his perspective is consistent with my assertion that the afferent function of the nervous system is to create an increasingly complex, ordinally structured sensory map of the external world, and the role of the mind is to interpret this structure.

Hughlings Jackson thought that the highest level of mental function—correlated with the highest level of brain function—was consciousness and predicted the existence of subconscious mental functions concomitant with the two lower levels of nervous system evolution. Being a rigorous and intellectually honest scientist, when he could not find any evidence of lower mental functions in comatose (i.e., non-conscious) patients—the positive symptoms predicted to appear as lower levels emerge from inhibitory control—he rejected his evolutionary structure for the mind. As described in the following articles, this rejection was based on a simple technical error, and when corrected, a suitably adjusted interpretation of his ideas is identical to this theory.

Hughling Jackson's description of the relationship between mind and brain is general and nonspecific, though it is seminal, and helps explain my decision to demonstrate the process of internal reconstruction as one of mental evolution. Still, two additional levels of description are needed to complete his analysis—a logical or functional level and its implied consequence, and a structural or physical level.

Logical/functional observations

The explicit functions of mind and brain emphasize their antipodal differences and dual nature. The brain, as a sensorimotor machine, operates deductively, while the mind operates inductively. In fact, intelligence—as a measure of mental function—can be succinctly defined as the facility of induction. The greater the intelligence, the more rapidly and accurately general conclusions can be drawn from specific instances. As will be seen in this theory of mental evolution, the inductive character of the mind is explicit, as each stage is formed by an equivalence identified in the preceding stage, and identification of an equivalence—or generalization—constitutes induction.

An analogous characterization of the brain as a sensorimotor machine is that it is governed by reflex action throughout. In fact, Hughlings Jackson's

axiom is the culmination of the extended development of an idea known as the Law of Reflex Action. A reflex can be expressed as a syllogism as follows: if "afferent x" occurs then "efferent y" follows; given "x," therefore "y." In classic logic this is known as *modus ponens,* the defining feature of deduction. The nature of the afferent and efferent processes become incredibly complex at higher levels of brain structure, but their nature as sensorimotor, and therefore deductive, is assured by the axiom upon which all neurology and neuroscience is based.

Complementarity

An interesting consequence of the (respectively) inductive and deductive natures of mind and brain is that they are dual structures, or complementary in the rigorous sense of the term—like heads and tails, night and day, or particle and wave. Complementary pairs are mutually exclusive yet reflect a unity at a higher level of abstraction. That is, heads and tails are unified by the structure of the coin, night and day by the rotation of the earth, and particle and wave by the solutions of the wave equation. The complementarity of brain and mind, or deduction and induction, indicates an abstract unity whose exact nature is interesting to contemplate, but one I have not been able to resolve.

Physical/structural observations

An important practical question that immediately arises when discussing a relationship of brain and mind is a structural one, i.e., how do the processes of the mind, whatever their specific nature, relate to the patterns of electrical activity that form the sensorimotor structure of the brain? Although the exact nature of this physical relationship is not relevant to this theory because internal reconstruction is functional and independent of any specific instantiation, I personally believe—and there is some experimental evidence—that the mind is an emergent property of the brain. Although the term *emergent* is found in many fields and with many meanings, I will use its general description from solid state physics, which consists of three defining features: an emergent system 1) cannot be predicted by a microscopic description of the substrate from which it emerges. In other words, it cannot be described reductively, 2) is substrate-independent, and 3) is characterized by long-range correlations.

Identification of an emergent mind can only be experimental. There has been some direct evidence that suggests this, but other, less direct, data also are consistent with this conclusion. For example, a long-range correlation is

found in low-frequency global brain electrical activity that appears to be associated with transfer of information from short-term to long-term memory. Also, even within an individual, the substrate of the mind, i.e., the physical structure of the brain, varies due to atomic and molecular replacement over time while mental processes are unaffected. Although it is uncertain if, or to what degree, brain cells are replaced, their components definitely are, and, in this limited sense, the mind must be substrate-independent.

If the mind is emergent according to these criteria, at least two important and interesting consequences arise. First, since the mind's substrate is the brain, it cannot be explained—by definition—through any microscopic description of brain structure. This is a reason to be skeptical about explaining consciousness in terms of the physics of the brain, as mentioned below. Second, if the mind is substrate-independent, mental processes are not caused by the specific human brain structures from which they emerge. Therefore, there is no restriction on the potentially diverse forms of astrobiological intelligences, that are otherwise precluded if mental functions are seen merely as a reductive consequence of activity of the human brain (or an identical structure).

Viewing the mind as an emergent property of the brain may be an attractive and intuitive proposition, but it is important to note that this characterization is not essential to any part of internal reconstruction or mental evolution. The critical relation between mind and brain is a functional one, and the physical correlation of cerebral and mental processes is inconsequential. Though an emergent relation seems the most rational explanation for the dual logical structures described above, the specific manner in which their complementarity arises is not relevant. The theory itself is unaffected by the physical or structural relationship of mind and brain, whatever its exact nature.

Beyond Philosophy

The study of the mind traditionally has been within the purview of philosophy, and possibly the major innovation of this theory is that its approach is purely scientific—i.e., it is axiomatically based, it is unambiguous and clearly defined, and it makes verifiable (or falsifiable) predictions. This means that the theory places the study of the mind in a context that can be rationally discussed and argued, and from which progress can be made. Since the nature of mental capacities has traditionally been a subject of conjecture, not a topic of scientific study, a clear and unambiguous definition of the current final level of mental evolution, i.e., consciousness,

has not been possible. And without such an unambiguous definition, it is not possible to suggest axioms and generate a hierarchy of well-defined capacities from which consciousness emerges—in short, to derive a continuity of evolving mental abilities. Conferences and symposia on consciousness and the mind rarely offer such definitions or even recognize their usefulness. Any proposed explanations are *a posteriori*, consisting of a collection of characteristics of consciousness or the mind, and are inadequate to form rigorous definitions amenable to scientific or evolutionary analysis.

One potential method of resolving some of the ambiguity in the purely philosophical approach has been an attempt—especially in recent years—at integrating philosophy of mind and physical science. Unfortunately, this attempt at grafting rigor onto ambiguity cannot clarify the underlying lack of understanding of either mind or consciousness. A rigorous description of how consciousness occurs in the human brain is theoretically possible, but only if the nature of consciousness is first clearly understood. An example that has generated an entire academic industry is the attempt to explain consciousness as a quantum process. The rationale for such investigations seems to be based on a questionable syllogism: consciousness is mysterious and seems *sui generis*; quantum mechanics is mysterious and seems *sui generis*; therefore, quantum mechanics must have something to do with consciousness. Of course, the relevance of a quantum description of consciousness based on shared incomprehensibility is neither heuristic nor convincing. A more serious objection, however, is that, if the mind is an emergent property of the brain, which seems likely, it is by definition impossible for it to be explained by a microscopic or reductive description of the substrate from which it emerges—including any quantum brain processes.

Mental Evolution

The following sections describe the terms and axioms underlying the theory of mental evolution, construct its ordinal stages, explain its predictions, and show how the predictions can be verified. Finally, I preview three significant consequences of the theory emphasized in the collected papers—the nature and temporal origin of consciousness and language, and the limits of knowledge that are implied by these novel explanations.

As a science, this theory of mental evolution consists of a circumscribed body of knowledge—the mind—three explicit axioms, a method of ordinal generalization, and a reproducibly testable and predictive theory of internal reconstruction. The theory begins with clearly stated axioms and progresses

so that each element of each stage is an equivalence class of elements in the previously defined level. This ordinality provides unambiguous definitions of mental capacities and a clear description of how these capacities are related and how they arise. The theory, which is based on a single observation—any organism can know the external world solely as an internally constructed model—culminates in a rigorous and unambiguous definition of consciousness.

Definition of Terms

The following terms have specific meanings in this theory of mental evolution: *concept, concept-space, recognition,* and *definition*.

Concept—a collection of attributes that determines membership in a category, is independent of its representation, and so, can exist in many forms. That is, a concept is an explicit equivalence relation and can be thought of as a category itself, independent of its members.

Concept-space—the place where concepts exist. This idea may initially seem unnecessary since the concepts of modern humans exist in an abstract mental space, and their form and content seem inseparable. But the abstract mental space itself is a product of evolution, and before it existed, concepts were embodied in a modified physical object—a physical concept-space. For example, a carved stone displaying the physical characteristics of an animal is the concept-space of an Upper-Paleolithic-Age concept.

Recognition—a process that involves comparison, as differentiated from definition.

Definition—a process that involves specification. For example, recognition of an object as a horse requires a mental comparison between it and previously encountered horses. Recognition does not require awareness of the characteristics that define a horse, or even that such characteristics exist. In contrast, definition of a horse requires specification of those characteristics that delineate "horseness."

Axioms

This theory is based on three axioms—the first is global, the second applies to the function underlying mental evolution, and the third is required to validate (or falsify) the predictions of the theory.

Axiom 1—The external world of any physical organism can be known only as an internally constructed model. It seems inconceivable that this is not true, but it is assumed and not proven.

Axiom 2—The function of the mind is generalization, or the mind is an inductive machine. A strong rationale for this idea is based on the science of neurology, which is possible only if the brain is seen as a purely sensorimotor machine. Because higher brain functions are increasingly complex, integrated, and interconnected, the afferent function of the brain is the construction of an increasingly complex sensory map of the external world. I posit that the function of the mind is to interpret this construction, and that these internally constructed sensory maps provide the substrate for mental evolution.

Since increasing mental capacity is most generally characterized by increasing inductive facility—the ability to identify equivalence—this axiom states that the specific function that interprets the brain's sensory maps is generalization. The elements of internal sensory maps are generalized and re-generalized in the mind, producing an increasingly complex, ordinally structured hierarchy of mental functions.

Prediction of a sequence of archaeologic artifacts from the hierarchy of evolutionary mental functions requires three corollary assumptions that together form the third axiom: mental capacities 1) can be correlated with a phylogenetic sequence; 2) are reflected in archaeologic data; and 3) are expressed as soon as they arise.

Axiom 3—The minimum mental capacity necessary to produce an artifact is the capacity indicated by the archaeologic appearance of that artifact.

The Process of Mental Evolution

The process of mental evolution proposed in this theory results from the instantiation of the abstract sequence of internal reconstruction in human phylogeny. The first two stages of internal reconstruction—sensory transduction at an organism's boundary surface, or the conversion of diverse external input to a common internal mode, and the identification of separate objects and the origin of time—are the most unfamiliar and may be the most difficult to understand. In addition, although they are arguably the most significant from a purely scientific perspective, they offer no observable predictive consequences. For this reason, the following discussion assumes that separate objects already have been identified within the sensory maps. I will discuss these first two steps of internal reconstruction later in this introduction and in the articles presented in the appendices, trusting that the explanation of the rest of mental evolution will help clarify these more abstract precursors.

The sequence of mental evolution consists of two parts. The first is an ordinal hierarchy that begins with objects and categories of objects and ends by defining a concept that explicitly describes what it means to be a member of the category—the category's equivalence relation. With objects identified as equivalence classes among separate global patterns of activation, the first stage of a predictive process of mental evolution begins. Each stage results from an equivalence identified in the previous level and, therefore, is defined exclusively in terms that already have been defined. This ordinality and precision provide clarity and rigor that previously have not been applied to mental processes.

The second part of the mental evolutionary sequence follows the introduction of the innovative idea of a concept-space, the place where concepts exist. As mentioned earlier, since we know of concepts only as ideas in our minds, form and content seem synonymous. However, this definition of a concept does not include any reference to how it is represented, and so it can exist in many forms. The increasing complexity of an ordinal sequence of concept representation—culminating in the abstract concept-space of consciousness—then forms the final phase of mental evolution.

Although the ordinality of the theory determines its scientific nature and is probably its most significant contribution, it also can complicate understanding. The lack of ambiguity that results from each stage of evolution being defined in terms of the previous one also means that once a higher level is achieved, its predecessors are subsumed and become functionally obsolete. In essence, the stages are erased as they are superseded. Consequently, from the perspective of a fully evolved (conscious) human, it is extremely difficult to perceive the subtle steps that are required to achieve our highest level of mental evolution. This is why consciousness seems unprecedented and why philosophy cannot identify what it is and from where it came.

Because the stages of mental evolution involve ordinal distinctions that are no longer relevant to our fully evolved thought—in both the characterization of categories and the representation of concepts—understanding the nature and origin of our modern mind requires both an altered perspective and an uncommon attention to detail.

Despite these hurdles—and as possible incentive to surmount them—the unambiguous nature of the evolutionary stages not only rigorously defines how mental functions arise and how they are related but ties their expression to the archaeologic record of human phylogeny and, thereby, generates testable predictions. It is also through this phylogenetic perspective

that a clear view of the sequence of subtle capacities that form the stages of mental evolution becomes apparent.

Assuming that the minimum level of mental capacity necessary to produce an artifact is the capacity indicated by the appearance of that artifact, the increasingly complex levels of mental function predict the appearance of specific artifacts—and only those artifacts—in a specific order in the archaeologic record. This theory of mental evolution not only predicts the precise—and exclusive—observed individual artifact classes but also the sequence in which they appear. In contrast, current interpretations of the meaning of early human artifacts—such as Neandertal burials being the birth of religion, or cave wall paintings being the birth of art—depend on nebulous and undefined notions of religion and art and cannot explain their sequential emergence.

Stages of Mental Evolution and Predicted Archaeologic Artifacts

Mental evolution begins with the partition of internal sensory maps into separate objects. It proceeds through repeated generalization in five ordinal stages: a) categorization, b) protoconceptualization, and c) conceptualization in 1) a contingent physical concept-space, 2) an independent physical concept-space, and 3) an abstract concept-space. Categorization involves recognition of the similarity of objects; protoconceptualization, recognition of the commonality that is the basis of categorization; and conceptualization, explicit specification of this commonality. Initially, physical objects embody the explicit characteristics that determine membership in a category and form a contingent physical concept-space that is inseparable from the concept itself.

The mind next recognizes the concept-space as an entity separate from any concept it carries, constituting an independent physical concept-space. Finally, the concept-space is defined, creating an abstract concept-space. ***The formation of an abstract concept-space marks the achievement of consciousness; and the mode of expression of abstract concepts is language***.

Once the process of mental evolution begins, each stage is formed by identifying an equivalence in the previous one, providing an adaptive advantage that contributes to further mental development. This adaptive advantage is the increasing capacity to retrieve elements of the perceived world for consideration, planning, and communication when they are not physically present.

Categorization

Categorization involves recognition of the similarity of perceived objects, and a category is defined as a collection of these equivalent objects. For example, the mind recognizes a category of horses that serves as a standard of comparison to determine if a newly encountered object is a horse. Each object that is identified as a horse inherits the characteristics of previously encountered horses. All animals have the mental capacity for categorization, since they at least recognize the category of food. Increasing complexity of categorization involves recognition of more categories, subcategories, and higher-order categories, such as categories of categories. Once a category is recognized, details of the category's elements, and similarities among them, also can be identified. These similar details form new categories that are subcategories of the original category. For example, once the category of horse is recognized, it is possible to recognize subcategories such as horse's head, tail, and limbs.

Predicted archaeologic artifacts. Since a category serves only as a standard of comparison and has no independent existence, this stage should not be associated with any observable form of representation or include any nonutilitarian artifacts.

Protoconceptualization

Protoconceptualization emerges with the recognition that the elements of a category have been grouped together because they share an equivalence—a set of defining characteristics. This stage is characterized by the recognition that a category exists independent of its members. That is, the defining characteristics of a category form a generic or ideal element—the equivalence relation—that is not an actual member of the category, but rather can be seen as the category itself**.** Once protoconceptualization, or the recognition of the existence of such a generic element, has occurred, a category can be retrieved for consideration independent of an observation. However, although the equivalence is recognized, it cannot be specified, so only a member of the category that manifests the defining characteristics can represent it. The protoconcept "horse" thus must be a particular horse.

Predicted archaeologic artifacts. The minimal mental capacity required to recognize the equivalence underlying a category, or that a category exists independent of its members— protoconceptualization—is marked by the appearance of unaltered natural objects that are intentionally isolated or displayed, each representing the category of which it is a member.

Conceptualization in a contingent concept-space

Conceptualization in a contingent concept-space occurs when the equivalence forming a category is defined. This can take place once subcategorization has produced protoconcepts that recognize sufficient fine detail within a category. For example, the protoconcept of the category of horse's tails is the recognized generic horse-like tail. A collection of these recognized generic characteristics of horses forms a generic horse—the concept "horse." Thus, a concept is defined as a collection of protoconcepts, or attributes. At this stage of evolution, a collection can only be assembled in a physical object, since no other type of space exists.

Therefore, the attributes defining the equivalence relation of the category are assembled in a unique, physical concept-space—a modified object such as a sculpture, a cave-wall painting, or an engraving of a horse. The concept and its concept-space—the image and the material in which it exists—are inseparable. In other words, the physical medium has significance only as the locus of assembly for the defining characteristics of a category, and the defining characteristics can only be brought together in such a physical object—neither can exist without the other. In this sense, the physical concept-space is contingent. It also provides an adaptive advantage, since concepts can be displayed in any convenient physical medium and be of any size, and so can be assembled and manipulated much more easily than protoconcepts, which must be members of their categories.

Predicted archaeologic artifacts. The minimal mental capacity required to explicitly specify the defining characteristics of a category—or concept—in a contingent concept-space is associated with the appearance of reproductions or representations of individual, recognizable objects in a physical medium.

Conceptualization in an independent physical concept-space

Conceptualization in an independent physical concept-space is defined as the recognized equivalence of contingent concept-spaces and results in a concept-space separate from any particular concept. So, for example, a carving of a horse and a carving of a bull are seen to have something in common—the stone that can accommodate either animal. A recognized equivalence must include any property shared by all of a category's elements, and since the contingent concept-spaces are all physical objects, the independent concept-space must also be a physical space. The emergence of the independent physical concept-space enables expression of increasingly complex conceptual relationships.

This capacity is manifested by compositional reproductions in which multiple concepts interact in a common space, e.g., a composition containing both a horse and a bull. In fact, the existence of an independent concept-space is necessary and sufficient for the innovation of composition. It is necessary because a composition cannot exist without a common space in which the individual elements of the composition are unified; and it is sufficient because all elements that appear in a common space are understood to be part of a unified composition. The appearance of an independent physical concept-space contributes to a significant adaptive advantage. For example, the existence of a common space in which multiple concepts can interact would enable, for the first time, expression of a shared story or myth. It also would enable more detailed transmission of information important for communal concerns such as tribal cohesion or hunting, and so represent a significant improvement over previous stages of evolution.

Predicted archaeologic artifacts. The minimal mental capacity required to define an independent physical concept-space is marked by the appearance of compositional reproductions in the archaeologic record.

Conceptualization in an abstract concept-space

Conceptualization in an abstract concept-space results from the specification of the equivalence of the independent physical concept-spaces, constituting the next ordinal stage—a generic independent concept-space. Each independent physical concept-space can accommodate only a limited number of concepts, but the generic independent concept-space must have an arbitrarily large capacity. Since no physical space is infinite, however, this must be an abstract space. In physical concept-spaces the concepts are expressed as visible collections of a category's defining features, which are shared by those who view them. The abstract space is invisible, however, so a means must be created to retrieve and share the concept. A name, e.g., *horse*, specifies the location of the concept in the abstract space, making language the shared expression of abstract concepts. Language demonstrates the ability to express the collection of attributes defining the category of horses that exists in an abstract concept-space—the abstract concept "horse."

Predicted archaeologic artifacts. The minimal mental capacity required to express the collection of attributes defining a category that exists in an abstract concept-space—an abstract concept—is the use of a word describing that abstract concept. This stage is marked by the appearance of

language and, as a result, should be associated with the disappearance of all physical expressions of concepts.

The existence of this abstract concept-space vastly facilitates conceptual operations, since concepts can be juxtaposed and hypotheses tested mentally, rather than by manipulation of physical representations. This increased mental efficiency also enables very high-order categorization identifying fundamental equivalences in the perceived world and defining objects whose existence is completely abstract.

To reiterate, the formation of an abstract concept-space constitutes consciousness, and conscious beings live in a model of the world that is created in this space. This model reproduces conceptual counterparts of the objects and relationships of the external world in the abstract space, and its existence explains the subjective sense of a separate observer that is a prominent characteristic of consciousness. Conscious beings do not simply act, but also are able to observe themselves acting because an action is performed both in the physical world and in the model of that world that resides in the abstract concept-space.

The ordinal nature of the stages of mental evolution provides an explanation for the universal development of language—and other mental capacities—even in possibly isolated human populations. The modern human brain includes the sensorimotor structure necessary to set mental development in motion. The process then unfolds independently, each stage arising as an equivalence identified in the one that precedes it. Although extended over thousands of years in phylogeny, each stage follows a clear path from its formation as a simple equivalence within its predecessor to an intrinsic equivalence that will be identified and then form its successor. As noted previously, the main difficulty in retrospectively perceiving the equivalences that form the levels of evolution is that each is so easily subsumed and made functionally obsolete by the greater mental capacity at the next highest stage. Contact and direct transmittal can alter the evolutionary rate, but not its ultimate occurrence and outcome.

Correlations with Paleolithic-Age Artifacts

The predictions of the theory of mental evolution explain the nonutilitarian Paleolithic-Age artifacts and their sequence, and the exact correlation with the archaeological record simultaneously provides verification of the theory. The artifact classes discussed below encompass the complete archaeological record but do not include all of the details that are discussed in the collected articles.

The only human artifacts found until the end of the Lower--Paleolithic Age and the ascendance of Neandertal Man are stone tools. Although human tools are more modified, more varied, and more permanent than those produced by other animals, the differences are of degree, not kind. Thus, the tool-producing hominids defining the genus Homo (i.e., *Homo habilis* and *Homo erectus*) were at the same level of mental evolution as other animals. This lack of nonutilitarian human artifacts in the archaeologic record correlates with the theoretical prediction of their absence at the evolutionary stage of categorization.

The Middle-Paleolithic Age

The Middle-Paleolithic Age begins with the ascendancy of *Homo sapiens neanderthalensis* approximately 100,000 years ago, and three associated artifactual innovations have been reported: human burial; manganese oxide and ocher used for body painting or dyeing of animal skins; and possible "ritual sites" devoted to cave bears and human skulls. A unifying characteristic of these artifacts is the isolation or display of unaltered natural objects—for burial and body painting, the dead and living human body, respectively; for ritual sites, cave bear skeletons or human skulls. Thus, each of these artifact types exactly demonstrates the predicted products of the evolutionary stage of protoconceptualization.

Although many fanciful meanings have been ascribed to early human burial, it most simply indicates recognition of some property of an individual that exists independent of the body. Since the category of an individual consists of a single object—the body of the individual—recognition of a defining characteristic independent of the body indicates the recognition of a category's existence independent of its element, the definition of protoconceptualization. The absence of burial prior to this era must mean that the body had no meaning beyond its existence as an object, which, in turn, implies that the mental capacity of protoconceptualization had not yet been achieved.

Since a protoconcept must be a member of the category it represents, only the physical body can act as the protoconcept for the human self. In this unique case, both the subject—the proto-I—and the object—"me"—reside in the body of the individual. Distinguishing the two necessitates isolating or displaying the object that is "me," as with any other protoconcept. The presence of pigments in Middle-Paleolithic-Age deposits is thought to indicate that Neandertal man practiced body painting and dyeing of clothes. Dichromatic body painting and the coloring of crude fur

clothing are the most elementary methods of differentiating the proto-I from its object, thereby recognizing the existence of the (proto-) subject.

The beginning of the Upper-Paleolithic Age

The beginning of the Upper-Paleolithic Age is heralded by the dominance of *Homo sapiens sapiens*. This is the time when physically modern humans began developing modern mental capacities. Two classes of nonutilitarian artifacts characterize this era—statuettes, paintings, and engravings; and personal ornamentation. Statuettes, paintings, and engravings are discrete, easily recognizable objects reproduced in a physical medium. These characteristics are precisely the predicted consequences of conceptualization in a contingent concept-space and are *prima facie* evidence of the mental capacity of this stage of conceptualization.

Specific characteristics of images produced on cave walls and engraved in stone blocks provide additional support for the conceptual nature of early Upper-Paleolithic-Age reproductions. Discrete cave paintings and block engravings nearly universally demonstrate superposition of images, often to the point of complete obfuscation. In addition, many block engravings were broken and tossed aside, apparently used for paving or building, like any other stone. It seems likely that these creations would have been more carefully preserved if humans capable of abstract thought had produced them with an artistic purpose. A more probable explanation is that the paintings and engravings functioned as concepts inhabiting discrete concept-spaces. These artifacts can be compared with notes that are written at a particular time for a particular purpose. They can then be balled up and thrown away or overwritten, with only the last message having any meaning.

The appearance of personal ornamentation at this stage provides further artifactual evidence of the definition of a category and existence of a concept. The concept here is the subject "I," and the natural choice for its physical concept-space is the body. In all other cases, a true concept is easily separable from a protoconcept of the same category, since the latter is a member of a category, and the former is a reproduction in a different medium. In this case, however, as with the protoconcept "I" and the category "me" at the stage of protoconceptualization, subject and object coexist in the body; and a way must be found to distinguish them.

As body painting provided this differentiation at the prior evolutionary stage, personal ornamentation differentiates the concept "I" from the object "me" at this higher level of mental functioning. Realistically, the only way to distinguish the protoconcept and concept of an individual is the

sophistication or complexity of the body's decoration. The multicoloured tattoos and vast array of personal ornaments that characterize the archaeologic record of the Upper-Paleolithic Age contrast strongly with the dichromatic body painting and dyeing of skins associated with the Middle Paleolithic. This advance provides evidence of the mental evolution from recognition of the category to its definition.

The final phase of the Upper-Paleolithic Age

The final phase of the Upper-Paleolithic Age is characterized by the appearance of carved and painted compositional reproductions. This achievement is the predicted consequence of the emergence of an independent physical concept-space and, as previously indicated, it is necessary and sufficient to explain this appearance.

The Mesolithic Age

The Mesolithic Age is a rather opaque era that is characterized by two remarkable and apparently contradictory features: 1) the disappearance of Upper-Paleolithic-Age naturalistic reproductions; and 2) rapidly advancing mental capacities as evidenced by the early stages of urbanization, domestication of plants and animals, and the appearance of organized warfare. The disappearance of Paleolithic-Age "art" has been difficult to reconcile with the cultural innovations of the Mesolithic Age, since art is often considered to be the epitome of cultural expression. However, this theory of mental evolution offers an explanation. Paleolithic-Age "art" disappeared because it did not function as art at all. Rather, it served as the increasingly complex expression of concepts in an evolving physical concept-space that became instantly obsolete when the ability to express abstract concepts through language appeared, as it did during the Mesolithic Age.

Innovative Implications of the Theory

The articles emphasize several innovative implications of this theory, among them, a clear and unambiguous nature and temporal origin of language and consciousness, and the limits of knowledge that can be established by applying these novel results. A distinguishing feature of this analysis—when compared with other descriptions of consciousness and language origins—is that its theoretical construct is independent of any philosophical or linguistic assumptions. It confirms the commonly assumed

intimate relationship between consciousness and language, but without reference to philosophy or linguistics. A primary goal of establishing a theory of mental evolution was to explain—as a consequence of a rigorous, ordinal, and independent analysis—what language and consciousness are and when they appeared. The result suggests a late origin of both in human phylogeny supported by the validated predictions of the theory found in the archaeologic record.

A statistical analysis of language also supports its late origin. The argument suggests that the sound-making capacity of the modern human larynx creates a vast redundancy in potential word creation relative to all languages that have existed. Since nature abhors such profligacy, the most reasonable conclusion is that the sound-making capacity of modern humans evolved long before language and for a far less efficient type of communication. When language evolved, only a minuscule fraction of the possible words were needed.

With the preceding articles having established an unambiguous definition of language in the context of an internal model of the world created by each individual, the final article explores an analysis of communication between individuals, societies, and civilizations and identifies semantic limitations on expression that extend from natural language to the most abstract languages of science and mathematics. It demonstrates that all words and symbols have an essential ambiguity that limits the precision of any form of communication and thus limits what can be expressed. For example, the sensory data received by an observer of an epileptic seizure is presumably the same for a Babylonian or a modern neurologist, but the meanings of words used to describe such a natural event are more a result of perception than reception. So the meaning of the Babylonian word for epilepsy, *bennu*, is formed by an observation filtered through the entire social and cultural context that influences its perception. Every aspect of Babylonian life, from notions of the nature of health and disease, to religious and social structures, including hierarchy and kinship, affects it. In other words, the meaning of *bennu* is determined by filtering an observation of an extrinsic recurrent natural event, epilepsy, through the Babylonian model of the world.

This process is, of course, the same for any society, giving rise to two important consequences. First, our observation of a recurrent natural event does not give us enough information to completely determine the meaning of the word describing the same event in another society. And second, since all observations are distorted, no word or symbol used to describe a natural event describes it completely and precisely. A word or symbol denoting any recurrent natural event specifies more information than the simple sensory

data—in short, it includes the multifarious contextual associations and interpretations characteristic of the society in which the observation occurs. Thus, any word used by a society to describe a natural event overspecifies the observation.

Societies are composed of individuals, and each individual has his or her own model of the world. Therefore, no two people have the same collection of experiential or contextual data for the meaning of any given word. For example, although the English word *epilepsy* labels a particular disease, each individual's complete definition of epilepsy depends on a unique set of experiences and associations. If this is so, the word *epilepsy* must have a consensual definition. That is, its meaning is formed by the intersection of all individuals' unique associations that comprise their personal definitions of the disease.

Since a society shares a great deal of context, two members of a society will share a more detailed idea of epilepsy then would either person with a member of a different society. And, of course, the society of neurologists will share an even more detailed and specialized definition. But since the meaning of the word *epilepsy* consists of the common features shared by a group of individuals, and each individual has experiences and associations with epilepsy that are not part of the group definition, the exact meaning does not exist for any single member of the group. In this sense, the meaning of a word used by a society underspecifies any individual's definition.

In summary, a word or symbol in any language and society has three interrelated characteristics. First, its meaning cannot be known completely. Second, it overspecifies the observation it describes, containing contextual elements unique to that society, and third, it underspecifies the definition of any member of that society, since it consists of only the aspects common to all members.

Therefore, the meaning of a word or symbol, as it appears in a document or any other form of interpersonal communication, has a strange existence suspended between the individual and the observed world, partially reflecting extrinsic and intrinsic realities, but not accurately reflecting either. It cannot completely reflect external reality because the model of the world through which it is filtered distorts this reality, and it cannot precisely reflect the internal reality of any individual because it is restricted to only those elements held communally. It is this existence apart from the actual realities of the physical and personal worlds that will be shown to impose limits on scientific and historical knowledge.

Modern Consequences of Mental Evolution

This presentation of mental evolution in terms of neurology and neuroscience has some interesting implications for modern research programs in those disciplines. In particular, it suggests new interpretations of the appropriate context for cognitive imaging in general, and its use in law and society in particular. It also has consequences for the perennially popular studies of the "neurology of art." The articles briefly explore these implications.

The designation "cognitive imaging" itself is oxymoronic, since it's only the activity in the neurological substrate, and not a mental process, that can be imaged with fMRI. There is no causal relationship between the imaged cerebral activity and the mental process, only a correlation, and even a complete understanding of the activity can provide no information about the nature of the mental function involved.

So, for example, the brain activity recorded during an fMRI experiment with mental arithmetic clearly has nothing to do with the nature of arithmetic, but only identifies the cerebral machinery used for performing arithmetic mentally. The same information about arithmetic could be determined by studying an abacus. The manner in which arithmetic is performed by the cerebral substrate is certainly a topic of interest, but this cannot be confused with the understanding of any aspect of the nature of arithmetic itself. In contrast, the functional imaging of sensorimotor processes is causal, and recorded activity does identify the sensorimotor function. So, for example, an fMRI study of recovery from stroke identifies exactly how the cerebral substance is re-wired to compensate for damaged regions. Cognitive-imaging practitioners, of course, are aware of this critical distinction, but sometimes don't honor it in interpreting their studies. Consequently, great care is required in the proposed uses of cognitive imaging in legal proceedings, specifically, or in social contexts in general. The idea of using functional imaging as a lie detector clearly demonstrates this difficulty.

Investigations of creativity by functional imaging, as in the "neurology of art," are similarly suspect. The images obtained indicate only the cerebral machinery used for a particular creative mental activity, or deficits created by damaged machinery, and provide no information about the creative process. In fact, the phrase "neurology of art" is *prima facie* oxymoronic, as there can be no deductive description of an inductive process, or a reductive cerebral explanation of an emergent mental function.