Reference: *Biol. Bull.* **215:** 204–215. (December 2008)

# Science of the Conscious Mind

GIORGIO A. ASCOLI* AND ALEXEI V. SAMSONOVICH

*Center for Neural Informatics, Structure, and Plasticity, and Molecular Neuroscience Department, Krasnow Institute for Advanced Study, George Mason University, Fairfax, Virginia 22030*

**Abstract.** Human beings have direct access to their own mental states, but can only indirectly observe cosmic radiation and enzyme kinetics. Why then can we measure the temperature of far away galaxies and the activation constant of kinases to the third digit, yet we only gauge our happiness on a scale from 1 to 7? Here we propose a radical research paradigm shift to embrace the subjective conscious mind into the realm of objective empirical science. Key steps are the axiomatic acceptance of first-person experiences as scientific observables; the definition of a quantitative, reliable metric system based on natural language; and the careful distinction of subjective mental states (*e.g.,* interpretation and intent) from physically measurable sensory and motor behaviors (input and output). Using this approach, we propose a series of reproducible experiments that may help define a still largely unexplored branch of science. We speculate that the development of this new discipline will be initially parallel to, and eventually converging with, neurobiology and physics.

## Introduction

The stated goal of the editorial invitation to contribute this position paper was "to place the debate on consciousness and how it may emerge from brains firmly in the historical scientific center of biology" (J. L. Olds, Krasnow Institute for Advanced Study, George Mason University, pers. comm.). The debate on consciousness in fact includes the question of whether the relationship between consciousness and the brain pertains exclusively to biology or also involves other scientific disciplines, such as psychology, physics, or informatics (Ascoli and Grafman, 2005). At the risk of failing (but in the spirit of) the original mandate, this article argues that the scientific characterization of the conscious mind, and thus its relationship with the nervous system or other material devices, requires a radical paradigm shift in research. As in previous analogous cases in the history of science, such a shift could result in a novel discipline altogether. At least, by suggesting the first steps in this direction, we take the position that consciousness research does belong to the realm of hard science, which is in and by itself a matter of contention (Chalmers, 1996; Ascoli, 1999a; Kim, 1999; Tononi and Edelman, 2000).

We begin from biology, and give a brief overview of the rapid acceleration in neuroscience progress. Next we notice that the subjective self is excluded from this "picture perfect" scenario, we tackle the issue of what the conscious mind is, and we explain its importance. We proceed with a proposal to approach this missing element as a scientific observable, including the definition of a quantitative, objective, and reliable metric system. Then we illustrate examples of experimental paradigms suitable to investigate the content of conscious mental states empirically. We conclude with a discussion of the future prospect for a grand unified theory of brain (or matter) and mind, and its possible implications.

## The Strides of Neuroscience: Toward a Complete Computer Model of the Mammalian Brain

Neuroinformatics and computational neuroscience are mature and exciting areas of research. The progress made in neuroscience during the last decades benefited greatly from neural modeling and numerical data analysis (Dayan and Abbott, 2005). Empirically driven, mathematically consistent, bottom-up and top-down models have been extensively tested and found to be robustly predictive. Quantitative descriptions cover a broad range of scales, from molecules to the entire brain, and seemingly all aspects, from anatomy and development to biophysics and behavior. Success stories include neuronal electrophysiology and synaptic plasticity (Gerstner and Kistler, 2002; Carnevale and Hines,

Received 15 July 2008; accepted 25 September 2008.

* To whom correspondence should be addressed. E-mail: ascoli@gmu.edu

2006); axo-dendritic morphology, outgrowth, and connectivity (Ascoli, 2002; Samsonovich and Ascoli, 2007a); and complex dynamics, from subthreshold interactions in single neurons, through spikes in cell assemblies, to continuous attractors in large-scale systems and long-range correlations among functional areas (Trappenberg, 2002).

These computational approaches harness Bayesian, information-theoretic, statistical-physics, and machine learning methods, and are not limited to a selected neuron type, brain area, species, or experimental paradigm. They have been applied to analyze, interpret, simulate, and predict experimental data, and to interface the brain in real time, providing monitoring, automated control, replacements, and extensions for parts of the nervous system (neural prostheses). Even the least understood areas of the brain, such as some subcortical nuclei and parts of the prefrontal cortex, appear solvable in the not-so-distant future (*e.g.,* O'Reilly and Frank, 2006).

A similar, complementary scenario applies to cognitive neuropsychology. The reductionist and functionalist approach confidently explains the physical nature of sensory, perceptual, and behavioral activity, all the way to top-level rule-based systems (Anderson *et al.,* 2004). The rapid progress in noninvasive imaging techniques (from electroencephalography and functional magnetic resonance to near-infrared spectroscopic and chemical shift imaging) further enables mapping these functions onto the human brain. Engineers and neuroimagers are simultaneously innovating these technologies and maximizing their benefits by coregistration of multiple methods. As a result, modern cognitive neuroscience is developing what can be called a moderate norm of the "typical" adult brain functions.

The continuous accumulation of data in molecular, cellular, systems, and cognitive neuroscience stimulated the development of electronic archives and corresponding data mining tools. Densely populated databases are already available for neuromorphological reconstructions (Ascoli *et al.,* 2007), gene expression maps (Lee *et al.,* 2008), functional brain imaging (Van Horn *et al.,* 2004), computational models (Davison *et al.,* 2004), and many other types of neuroscience data (Gardner and Shepherd, 2004). Although many technical challenges remain, there is a widespread agreement on the eventual feasibility of decoding the "connectome," the complete architectural blueprint of cellular connectivity in the mammalian brain (Sporns *et al.,* 2005; Swanson, 2007; Lichtman *et al.,* 2008). Overall, neuroscience, engineering, and bioinformatics appear unstoppable and without limits (Samsonovich and Ascoli, 2002).

## Make Up Your Mind on What It Is Like to Be You

Despite the triumphs of neuroscience, something is missing. That element is the conscious mind, meant as first-person experience, or subjective feelings, and their associated meaning (Flanagan, 2007). The functionalist view of reductionism is that "the mind is what the brain does," yet most of what the brain does is unconscious, while some aspects of that activity are definitely responsible for our feelings. This consideration suggests the following variation: *"the conscious mind is what {those things the brain does [which feel like something]} feel like."* In this definition, we interpret "brain" liberally to include electronic artifacts such as (software or hardware) brain-like models, as well as other future instantiating systems. Square brackets and braces are unconventionally used to clarify the logical parsing of the sentence rather than to indicate attributes that could be omitted without altering the definition. The rest of this paper will rely on the notion that consciousness of what it is like to be someone is arguably the most (perhaps the only) important, defining character of human existence (Frankl, 1946; Samsonovich and Ascoli, 2005a).

The question of consciousness will soon acquire substantially higher practical and ethical significance, as human-level computing power becomes available. Machines that think and grow cognitively like humans can help solve social and economic problems of our society (Albus and Meystel, 2001). New initiatives in artificial intelligence intend to tap neuroscience knowledge and to replicate the principles of neural information processing in artifacts (*e.g.,* Haikonen, 2003). On the one hand, the terms "consciousness," "self," "emotion," "qualia," and "episodic memory" are increasingly common in commercial and academic computer science. On the other hand, computational neuroscience lacks higher-level concepts related to agency and subjective experiences (Samsonovich and Ascoli, 2002). These issues are also consequential for biomedical science, as psychiatry and neurology unavoidably move toward an integrated view (Albus *et al.,* 2007). There is an increasingly widespread consensus that it is urgent to fold the conscious mind into the domain of empirical science (Spitzer, 2008).

It may seem to be a natural explanation that the very complexity of the brain should give rise to the emergence of first-person experiences (Tononi and Edelman, 2000; see also Ascoli, 2000; and Tononi, 2008). There is, however, no logical necessity for this to happen, and the notion of this emergence is not yet rigorously defined (Fig. 1A). Objectively, what does it mean that a physical system has subjective feelings? The more we study the brain, the less we understand the reasons for its connection to the conscious mind (Ascoli, 2005). Similarly, while many if not most aspects of neural dynamics, structure, and development already are (or can be) modeled computationally, there are no credible attempts to create an equivalent of a conscious mind *in silico.* First-person experiences seem incompatible with the present scientific paradigm.

Starting from Aristotle's proposal in *de Anima*, philosophers developed an alternative framework to describe real-

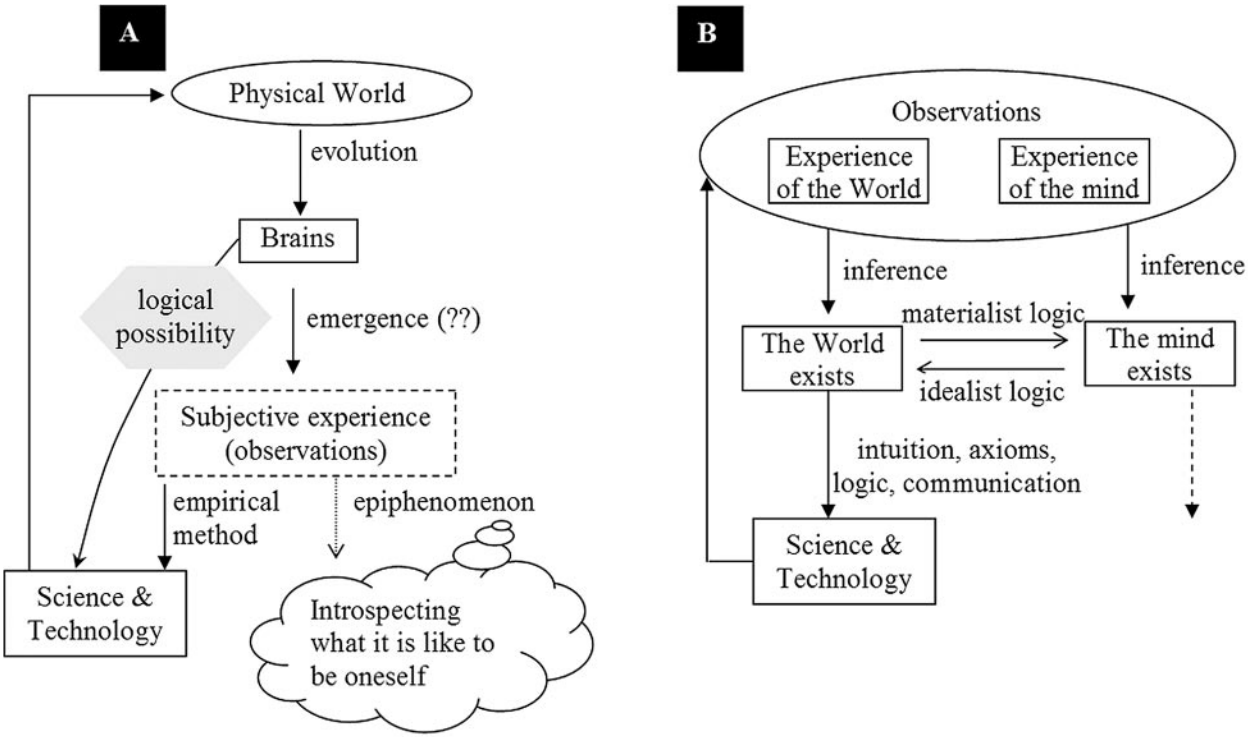

**Figure 1.** Materialist (A) *vs.* idealist (B) views of scientific reality. In both representations, the oval is the starting point, and rectangles are considered "facts." Note that important human endeavors, such as art, are typically considered complementary to science and technology. Following the scheme of panel B, we propose to "complete" the process symmetrically, by applying intuition, axioms, logic, and communication to the inferred knowledge of the existence of the mind, much as is done to the inferred knowledge of the existence of the World. We expect that a new branch of science and technology will result from this extension.

ity (*e.g.,* Kim, 1999). Since all that we know of the world we ultimately learned through subjective experience, the first-person perspective becomes the starting point (Fig. 1B). Both matter and mind are then inferred (if not constructed) from these observations. Nevertheless, the empirical scientific method has not yet been applied to this idealist viewpoint. The subjective, private, and personal nature of the conscious mind may at first seem impossible to reconcile with the need for the objective communication and reproducibility of science. Yet what is objectively communicated and reproduced in scientific research is not the entity itself (a distant galaxy or an enzyme), but rather its features (such as temperature and kinetic constants). Human beings routinely communicate the features of their subjective experiences (*e.g.,* "this ice cube feels cold") and can agree on some objective facts about them at least under normal, controlled experimental settings (*e.g.,* ice feels colder than boiling water). This point of view suggests that a science of subjective experience could be developed in exact analogy with the traditional scientific paradigm.

## Scientific Accessibility of the Conscious Mind

Consistent with the view illustrated in Figure 1B, a plausible cycle of scientific discovery and development can be entertained (Fig. 2). The mind builds mental constructs based on experience and uses them to satisfy its practical needs. The expected future outcome is the emergence of artificial or hybrid minds that might then contribute to the same loop. The first, but typically implicit, step of scientific investigation is to recognize that subjective experiences reflect objective reality. In other words, scientists accept (at least in practical terms) that there exists, independent of them, an objective reality that they can observe through experience. This is a nontrivial statement, and it is not a logical consequence of any known fact. It is rather stipulated in empirical science as one of the axioms of the scientific method. Other fundamental elements of this paradigm include logic, the principle of parsimony, observations, hypotheses, predictions and their experimental tests, and the beliefs that there are persistent universal laws of the world that can be learned and stated as falsifiable theories.

The process produces a mathematical apparatus, metrics, and tools to probe and engineer the object of study, which is limited to the physical world. Experiences in the empirical scientific paradigm are a means of observation, not objects of study. Within this traditional scientific investigation of the physical world, subjective mental states are mere epiphenomena (Fig. 1A), abstractions, or at best are identi-

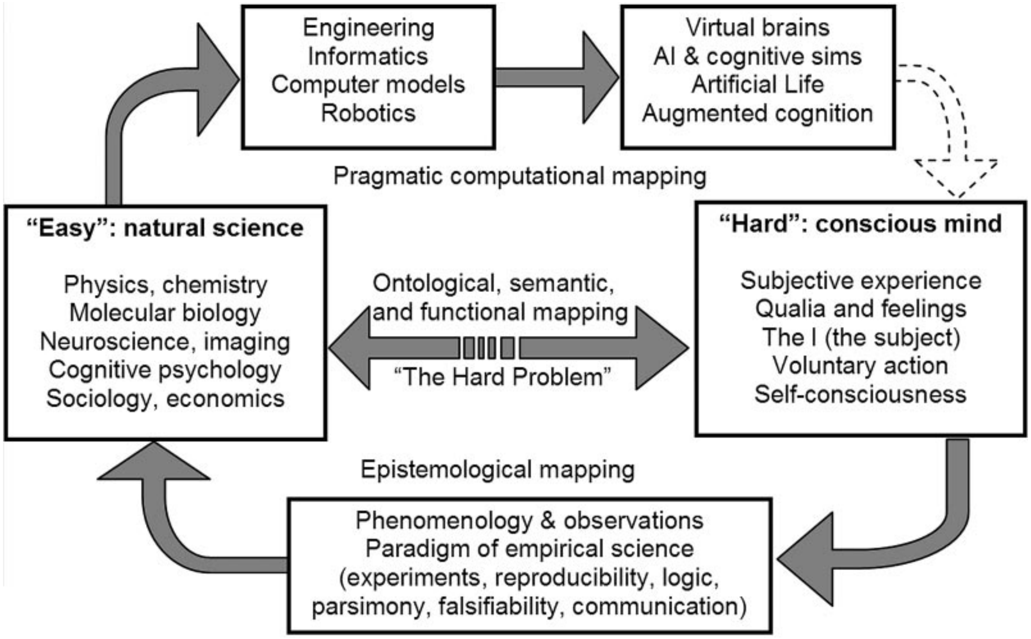

**Figure 2.** Scientific evolution of the mind and the "Hard Problem of consciousness" (adapted from Samsonovich *et al.,* 2008). AI, artificial intelligence; sims, simulations. Qualia is plural for "quale," the qualitative aspect of sensations from the first-person perspective (Lewis, 1929).

fied with patterns of neuronal activity. Yet because they are directly observable *via* experiences of experiences (Fig. 2), first-person experiences, or qualia (Lewis, 1929), should be considered at least as "real" as the postulated physical world that they provide access to. For human researchers, the occurrence of subjective feelings is a fact, not an illusion. But the same researchers cannot scientifically infer the same distinction about any other person. This conundrum has been stated as the "Hard Problem of consciousness" (Chalmers, 1996). Our attribution of subjective experiences to people on the basis of observations and physical measurements is arbitrary. In principle, one can speculate that actual experiences of people could be different, or not present, or present in only 50% of all cases, with no consequences for any physical measurement. Therefore, the questions of how a brain looks to a researcher and how "it feels itself" are not reducible to each other and must have separate answers.

The proposed solution is to extend the scope of empirical research and the scientific paradigm to encompass not only the physical world of the traditional hard sciences, but all of observable reality, including the conscious mind. In particular, the same loop representing the investigation of the physical world (Fig. 2) also applies to the investigation of mental states (Fig. 3). In exact analogy to the previous description, we notice as the first step that some of our experiences convey information not about the physical world, but about the experiences themselves. In particular, we can observe our subjective experiences directly through higher-order experiences (*i.e.,* experiences of experiences; see also "higher-order thoughts," or HOTs: Rosenthal, 1993). This is again a nontrivial observation, which is not a consequence of any previously known scientific fact. Much as we do for the physical world, we can then postulate as a new axiom of science that there are subjective experiences, and that they are observable. Moreover, our previous failure to identify these subjective experiences as features of the physical world suggests the possibility of their being elements of reality irreducible (or at least so far not yet reduced) to a measurable material existence described by traditional science.

There may at first appear to be a fundamental distinction between an observation about a conscious state and that about the physical world. In particular, for hard sciences, "observable" means by many people, from different reference frames, and it implies inter-observer reproducibility. Since only one person (oneself) can observe a conscious state, how can the observation be reproduced? Again, it is crucial to remember that what is reproducible in scientific research is not the entity itself, but rather its features. Observations about subjective conscious states are thus reproducible if the features attributed to these mental states by independent observers are consistent.

Upon accepting the axioms on the existence and observability of conscious mental states, we can proceed by adopting or adapting the same scientific principles developed for the other observable elements of reality (the physical world). These include logic, measurements, hypotheses,

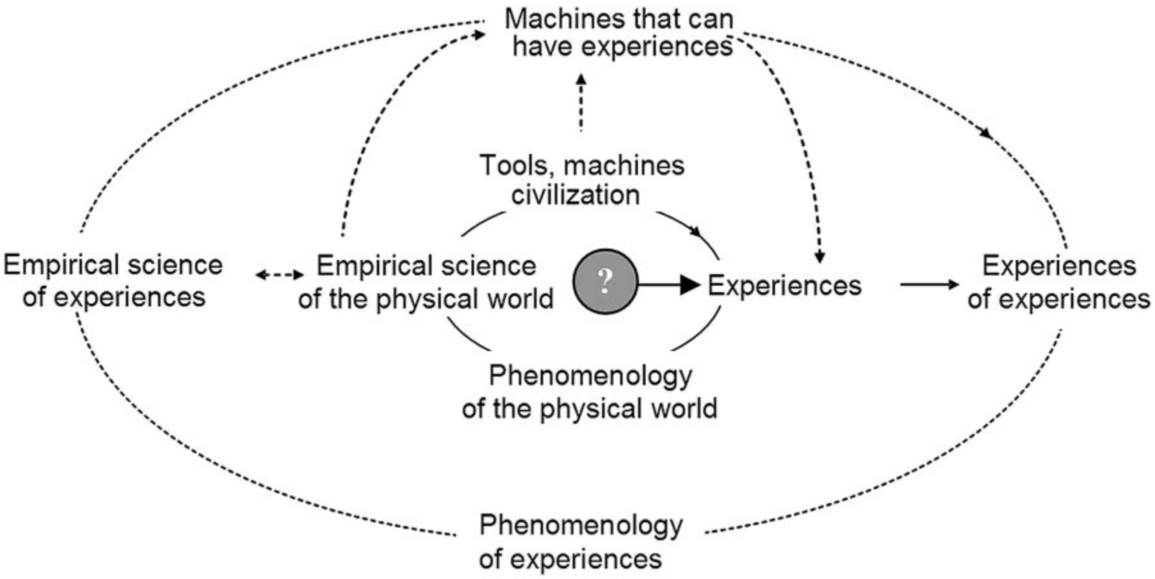
Machines that can have experiences
Tools, machines civilization
Empirical science of experiences
Empirical science of the physical world
?
Experiences
Experiences of experiences
Phenomenology of the physical world
Phenomenology of experiences

**Figure 3.** Representation of the logical flow underlying the science of mental experience as a mirror image of that of physical reality (adapted from Samsonovich *et al.,* 2008). Experiences constitute the axiomatic starting point and the primordial set of data. Upon these, we build a phenomenological representation of the world, which is predictive and thus evolutionarily advantageous. Quantitative and systematic measurement systems, along with the experimental method, lead to the development of empirical science. We argue for the necessity of the corresponding transition in the phenomenology of experiences.

predictions and their experimental tests, and the beliefs that subjective experiences obey persistent universal laws (parsimony), which can be learned and stated as falsifiable theories. Using the scientific paradigm, we can then systematically document the phenomenology of subjective experiences and develop an empirical science of conscious mental states. As a result, we will acquire new knowledge about the mind. The expected outcome might include the means to cure or prevent psychiatric disorders, optimize education and training, engineer ever more powerful computing and communication devices inspired by or optimized for the human mind, and reproduce consciousness in machines.

Even if conscious mental states and corresponding brain states are logically irreducible to each other, the development of a new science of the mind is expected to yield consistent empirical evidence of a tight correspondence between these distinct elements of reality. In other words, there will be a quantitative if unexplained connection between the present existing science of the physical world and the new envisioned science of the conscious mind. For example, the (self-described) feeling of being happy might consistently co-occur with a well-defined pattern of brain activity detected, for example, by noninvasive functional imaging. The mapping between these two scientific realms would encompass both subjective reports referring to internal states, as in the happiness example, and to external features, such as the feeling that an object is cold (and the relationship to its temperature).

In practical terms, it is likely that the experimental frameworks designed to characterize the mind will have fairly complementary (or at least not entirely overlapping) technological limitations relative to those encountered in physics and biology, resulting in parallel research enterprises. However, the two theoretical constructs can be formally unified by postulating a new "supervenience" axiom, known as the principle of organizational invariance (Chalmers, 1996): systems with similar functional organization must have similar experiences. This means that identical brain states give rise to identical mental states. Note that the reverse is not necessarily true (the mapping can be degenerate): small changes in neural organization, such as the activation of an individual voltage-gated ionic channel, may go undetected by the conscious mind.

Several contemporary researchers have also accepted the scientific validity of first-person experiences, including philosophers, computer scientists, and neurobiologists. Among the approaches that resonate more closely with our position, some propose several axioms or constraints based on introspection, to be addressed by scientific theories of consciousness (Metzinger, 2003), or more specifically, to be turned into third-person models by developing the mechanisms that their formulation implies (Aleksander, 2005). The idea that the result of introspection should be amenable to examination as a functional virtual machine has also been surmised in artificial intelligence (*e.g.,* Sloman and Chrisley, 2003). Moreover, considerable work is ongoing to characterize the neural correlates of consciousness (for review, see Tononi and Koch, 2008), which are important for linking introspective reports to brain measurements.

## Semantic Maps as A Metric System for Subjective Meaning

A substantial phenomenology of subjective experiences is already documented, and attempts are underway to establish their neural correlates. Indeed, establishing introspection as the foundation of reality (in the phenomenological tradition of Husserl) has recently been proposed in artificial intelligence as well (a trend referred to as *synthetic phenomenology*). However, including the conscious mind as a target of investigation of empirical research implies extending the theoretical framework of science with new concepts to describe mental states rigorously and precisely. In particular, we need to design mathematically sound metrics reflecting definite aspects and elements of our subjective experiences, and a corresponding system of quantitative measures. Important phenomenological experience may be tied to individuals (consciousness of beauty, responsibility *etc.*), rather than to concrete objects whose features could be explained by the pattern-recognition properties of neural networks. The need of a metric addresses the necessity to compare features of feelings quantitatively both within an individual and across individuals. An example of the first case would be when something feels hot, warm, lukewarm, cool, or cold to me, or when a friendship feels deep and highly valued or casual and lightly valued. The second refers to a situation in which something feels warmer to one individual than it does to another, or my friendship with one person feels deeper to me than your friendship with a different person feels to you. In either case, both a reference point (the neutral position) and the unit of measure must be determined.

The measurement itself cannot be acquired externally by any physical device but must instead be subjective, because so is the target observable. At the same time, traditional introspective techniques, based on discrete ranking or semi-qualitative Likert scales (for example, scores of 1 to 7; Likert, 1932), may fall short of satisfying quantitative scientific standards. Physics started by developing the means for numerically measuring length and time, then mass, velocity, force, *etc.,* and for interrelating the corresponding concepts. When discovering previously unrecognized phenomena, physicists invent new measuring devices. Similarly, the empirical science of the conscious mind should begin with a precise system of metrics, adequate measuring tools, and an underlying system of mathematical concepts (Fig. 4A).

To define a metric system for subjective experiences, we need to identify their principal dimensions. This knowledge can only be derived from first-person human data. Use of qualitative subjective reports could require an impractically large number of entries and human subjects to attain an acceptable level of quantitative accuracy and a representative sample of the diversity and natural variability of conscious mental states. The ideal study should include a large population over many generations. These conditions may be approximated by mining natural language directly. Every word (or in general, concept represented by a word) has qualitative and quantitative semantic aspects. For example, "large" and "warm" relate different qualities, but orders can be conceived relative to, for example, "huge" or "hot." These relations can be quantified with measures of size and temperature, but what about less physical attributes such as "friendship," or the content of a complex discourse in a specific context?

The idea of semantic space, defined as the set of all possible meanings that words can express, may be formalized with the notion of cognitive mapping. Cognitive maps index representations by their context, such as spatial location, and are employed by mammals for path-finding and navigation (Samsonovich and Ascoli, 2005b; McNaughton *et al.,* 2006). The idea of describing the content of meaning geometrically (Gärdenfors, 2004) is to embed concepts in an abstract metric space, capturing semantic relations with distances and angles (Fig. 4B). Orthogonal directions in this space would correspond to qualitatively distinct aspects of meaning, while relative positions along an axis would reflect quantitative semantic relations (*e.g.,* Ploux and Ji, 2003).

We have recently demonstrated the existence of a semantic map of human language, in the form of a multidimensional space in which the relative position of each word quantitatively reflects the content of its meaning (Samsonovich and Ascoli, 2007b). The proof by construction consists of a simple, robust, and reproducible computational procedure to derive such map from a dictionary of synonyms and antonyms with a thermodynamic approach. To start, all words are randomly allocated in a high-dimensional space. Their positions are then optimized to minimize an energy functional defined such that synonyms and antonyms tend to align parallel or anti-parallel, respectively. The emerging principal components of this map correspond to easily recognizable semantics (Fig. 4C). The primary and secondary axes are aligned along the meanings of good /bad ("value") and of calm/excited ("arousal"), respectively. Additional orthogonal components include the senses of open/closed and copious/essential.

This semantic map is sufficiently robust to allow the automated extraction of synonyms and antonyms not originally in the dictionary, and to predict the particular "twist" of their meanings on the basis of their spatial coordinates. Interesting geometric characteristics include the low (~4) dimensionality (Fig. 4D), a bimodal distribution of the first component, increasing kurtosis of all subsequent (unimodal) components, and a "U-shaped"

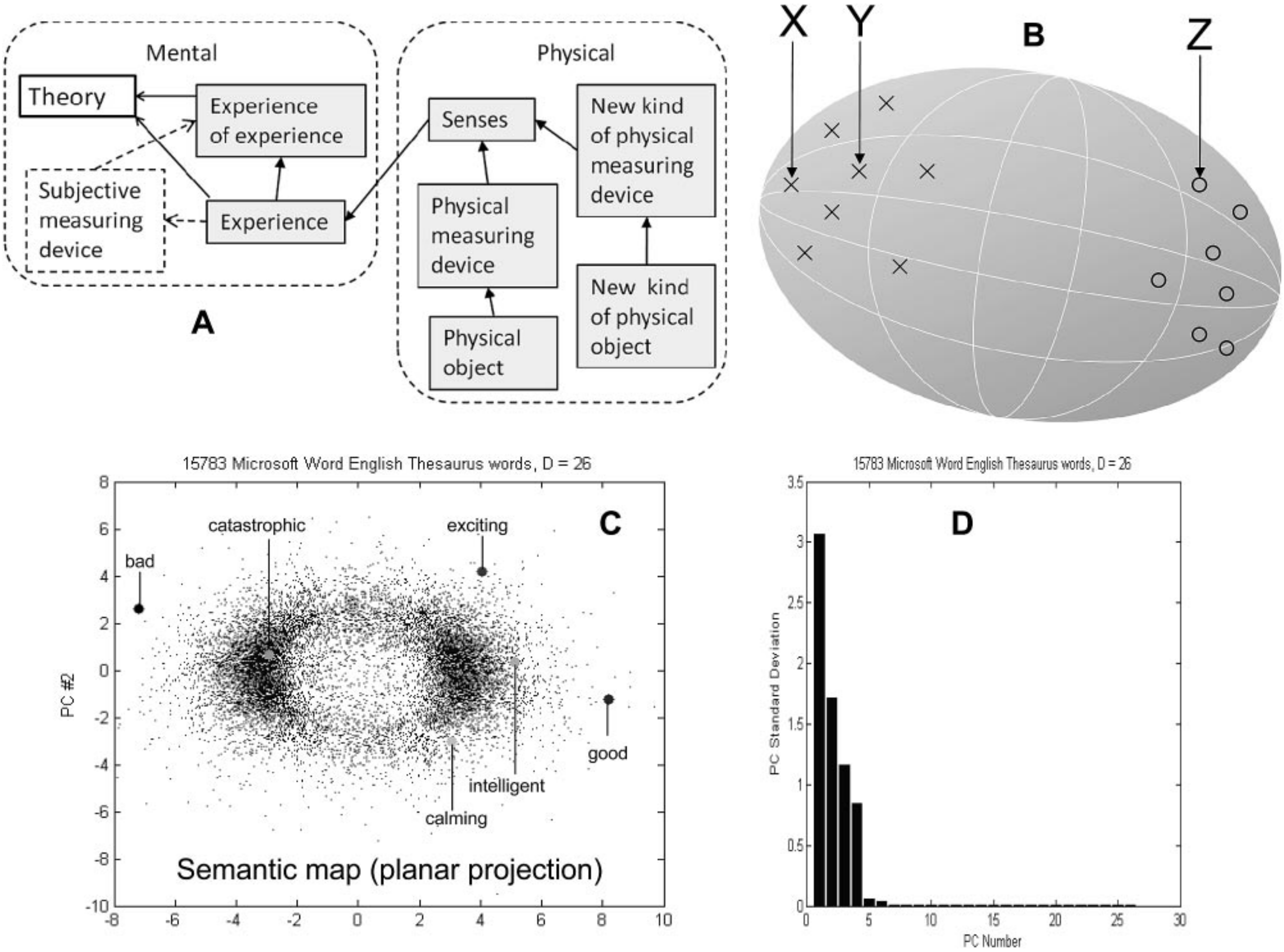

**Figure 4.** (A) Logic of measurement and its extension from the traditional to the extended scientific paradigm. (B) The concept of semantic cognitive map viewed as a manifold. Geometric distances between symbolic representations (words) allocated on the map reflect their semantic relationships. Words X and Y are synonyms; words X and Z, Y and Z are antonyms. (C) Projection of the semantic map obtained from a dictionary of synonyms and antonyms on the first two principal components. The position of each word quantitatively reflects its value on the good/bad (horizontal) and calming/exciting (vertical) axes. (D) Dimensions of the map correspond to the principal semantic values of the representations, in this case corresponding to subjective experiences (A and B adapted from Samsonovich *et al.,* 2008).

maximum-spread planar projection. Both the map's semantic content and main geometric features are consistent between dictionaries (Princeton's WordNet and Microsoft Word), among languages (English, French, German, Russian, Spanish), and with known psychometric measures and established linguistic theories (Leary, 1957; Osgood *et al.,* 1975; Chomsky, 2006). The fact that these principal semantic components correspond to abstract conceptual and emotional dimensions (value, arousal, *etc.*) rather than concrete things also suggests that they may constitute fundamental universals of phenomenological experience.

Elements of this semantic map consist of general concepts each of which can be expressed in a word. This framework can be expanded to describe segments of text or the specific content of entire topics expressed in extensive documents, by combining the frequency and semantic content of their individual words with syntactic processing and latent semantic analysis (Latent Semantic Analysis@Cu Boulder, no date). In addition, elements of more general nature, such as pictures, sounds, abstract symbols, can be similarly included in the map. In principle, semantic maps can operate as subjective measuring devices. For example, a researcher can memorize a set of "landmark" concepts that uniformly cover a semantic map, and learn to allocate any subjective experience with respect to those landmarks. By facilitating the input-output with a powerful human-computer interface, it would be possible to communicate any ongoing subjective experience in real time and digitally. This method would thus record dynamics of the conscious mind "on-line" in a quantitative and objective format.

The present instantiation of our semantic map, as presented here, is based on individual words, which can be interpreted as representing a subset of all possible concepts. Other concepts, and mental states in general, may be difficult or even impossible to express in words. Thus, in the strict sense, this semantic map should be taken as just an example of the kind of metric system needed to quantify the content of mental states. To work as a general measurement device, the map would need to be expanded to represent nonverbal cognition. At the same time, language is often considered one of the most useful windows on the mind that is available to researchers (*e.g.,* Pulvermüller, 1999; Arbib, 2005; Hampe and Grady, 2005; Pinker, 2007). Thus, semantic maps based on language can in and by themselves constitute a powerful tool toward the establishment of a science of the conscious mind.

## The "Hard Problem" Is Hard, not Impossible

Semantic maps initially derived from natural language dictionaries and corpora can be gradually and systematically enriched with subjective measurement techniques as outlined above. The resulting complete semantic map of human experiences provides a possible theoretical framework for the new science of subjective mental states. A mathematical model of the conscious mind can be formulated as a distribution of activity on this map, along with corresponding symbolic content of representations and the laws of dynamics, including input and output. The experimental reproducibility of the scientific method mandates pooling together results of subjective measurements by many researchers. Therefore, either a single universal semantic map should be adopted by all subjects, or an isomorphism among different personal semantic maps must be established (Palmer, 1999).

Modern imaging and psychometric tools can then be leveraged to co-register the complete models of the human mind and of the human brain. There is abundant evidence for cognitive maps in the brain (Thivierge and Marcus, 2007), but new studies are required to map the most abstract representations and the implementation of value or meaning in the nervous system. Nevertheless, recent breakthroughs indicate both the scientific and technical feasibility of the advances envisioned in this position paper. In particular, a key challenge in noninvasive functional imaging consists of inferring higher mental states from the corresponding activity patterns in naïve contexts. In a relevant development, Kay *et al.* (2008) demonstrated that it may soon be possible to reconstruct a picture of a subject's visual experience from measurements of brain activity alone. Their decoding method relies on quantitative models relating visual stimuli to brain activity in early visual areas, and enables an observer to identify the specific image from a large set of completely novel natural images. These receptive-field models tune individual voxels (besides their spatial retinotopic organization) for spatial frequency and space directly based on responses evoked by natural images. (A voxel is a three-dimensional (volume) data point, by analogy with pixel, a two-dimensional point.) These results suggest that a similar technique may be applicable to higher cognitive states.

Indeed, another newsworthy study reports similar "mind-reading" abilities with respect not just to a sensory modality, but to abstract semantics instantiated by common words (Mitchell *et al.,* 2008). Subjects were scanned while contemplating the meaning of 60 nouns, one by one. The resulting brain activation patterns were decomposed along 25 dimensions, each corresponding to a specific verb. The decomposition matrix (*i.e.,* the weights) were based on the co-occurrence of a given verb-noun pair, measured from Google's trillion-word corpus of web pages. The same matrix and frequency statistics were then used to generate predictions based on the resultant virtual imaging signatures associated with the verbs. The model predicts activation patterns for thousands of concrete nouns in the text corpus for which functional imaging data have not yet been acquired, with highly significant statistical accuracy over the words for which activity data are available.

The latter study constitutes a crucial breakthrough, and misses only one critical element: a universal metric system for subjective experiences. As the authors themselves note, co-occurrence counts constitute a popular but crude approximation of the semantic content of a word (Mitchell *et al.,* 2008). Nonetheless, this same method (and even this same data set) can be deployed to yield virtual imaging signatures corresponding to the principal semantic components of language obtained from the dictionaries of synonyms and antonyms. In this case, the weights of the decomposition matrix would be based not on co-occurrence, but directly on the coordinates of each word on the semantic map. The resulting virtual signatures would constitute quantitative predictors of the corresponding subjective dimensions of a person's mental state. These predictions would be directly suitable to experimental tests and could include an estimate of expected inter-subject variability. The choice of the 25 manually selected verbs was rationally justified in the original study (Mitchell *et al.,* 2008), but ultimately arbitrary. In contrast, these alternative coordinates constitute the emerging principal semantic components of language, and are thus expected to capture the most significant and consistent aspects of concepts that can be expressed in words.

## Times of Radical Departures

Why do we need a paradigm shift to develop a science of the conscious mind? After all, "psychology" means "mind science," so mental states could be assumed to be precisely what psychologists investigate. Instead, modern psychology is heavily involved in the study of objective behavior, rather than subjective experiences. There seems to be a widespread conviction that behavior and neural activity are in fact the only observable aspect of the mind. As we discussed above, this conviction is true only if we further qualify it as "observable by a person other than the subject." To illustrate the magnitude of this distinction, let us consider the following thought experiment.

Two biopsychologists, Ann and Jane, are on a flight to the annual neuroscience meeting. They are from the same university and department, where they had similar, distinguished careers. Now they are both silently reading the same thesis draft from a joint student, but soon their minds drift off. Ann thinks of her home, and her thoughts wander to her son in college. Contemplating for a few minutes her feelings of pride, love, and satisfaction, she remembers that he promised to leave her a message with the results of his first midterm exam. Ann's eyes are still fixating on the middle of the page, and she is enjoying these moments of quiet and relaxation.

Jane also starts thinking of her home, but is immediately assailed with the fear that she might have left the stove on when heading for the airport. Hurriedly reviewing her memories of closing the carry-on and seeing the taxi from the window, she convinces herself that indeed she forgot to switch off the burner under the tea kettle. Jane's heart is now pounding as she pictures the all-wooden cabinetry surrounding the stove. Her eyes are still fixating on the middle of the page, and she is paralyzed with fear and embarrassment for her carelessness. Just then the plane lands. With the announcement allowing use of cell phones, Jane and Ann simultaneously reach for their pockets, and they both speed-dial #1. Ann is checking her voice mail; Jane is calling the fire and rescue emergency line.

From the third-person perspective (*e.g.,* one of the many other biopsychologists on the same flight to the neuroscience convention), Ann's and Jane's observed behaviors are nearly identical, yet their subjective mental states are radically divergent. Although ontologically, as discussed above, the two subjective experiences are not directly measurable with physical devices, they have observable consequences in the physical world, as could be confirmed by, for example, listening to their cell phone conversations. Logically speaking, these subsequent phone calls, and even the heart rate, breathing, muscular tension, *etc.* (which would likely all be altered in opposite directions in these two cases), could also be considered as aspects of a broadly defined extended behavior. Skin conductance and neural recordings, which are also obviously "observable," could similarly distinguish between Ann's and Jane's mental states, although in the specific example they were not "observed." Internal brain dynamics recorded with physical instruments could in principle even predict the difference in the contents of Ann's and Jane's phone calls. In practice, however, the most telling behavioral aspects of the conscious mind are seldom objectively quantified in cognitive science. In particular, most quantitative research in psychology does not concentrate directly on what the aspects of cognition that feel like something actually feel like. Studies that do tap into this crucial characterization of the mind arguably do not rely on a system of quantitative, reproducible metrics (Dawes, 2008).

Although it may not be generally possible to decode conscious mental states on the basis of observable behavior alone, simultaneous recording of neural activity while also keeping track of inputs and outputs can lead to significant progress, as discussed in previous sections of this paper. Yet the specific content of subjective experience is typically neither associated with overt behavior nor quantified, even when the corresponding emotional state, attention, memory recall, mental rehearsal, *etc.,* are studied using neuroscientific approaches. In particular, several aspects of human subjective experiences have been examined within a rigorous statistical framework, especially in the field of autobiographic memories. For example, the temporal distribution of autobiographic memories in normal adults has consistently been shown to obey a power decay from recent to more remote episodes (Rubin and Wenzel, 1996). This means that when humans recall their past, they are more likely to think of the previous week than the previous year, and this probability can be estimated with high precision, confidence, and reliability. However, these are relatively peripheral aspects compared to the very content of those autobiographic memories. An attempt to probe a more central aspect is constituted by the Cue-Recalled Autobiographical Memory (CRAM) test, which measures the number of specific details (objects, people, locations, *etc.*) retrieved in a memory (CRAM, date unknown). However, the quantitative characterization of the conscious mind independent of (or complementary to) sensory and motor behavioral observables is still largely missing in other crucial areas of cognitive science, such as decision-making and situation awareness.

## Empirical Dissociation of Mental States From Behavior

It would be useful to design an experimental paradigm to induce distinct mental states (including independent perceptual interpretation and intentional planning) while controlling for all possible behavioral variables, and in particular maintaining identical input (sensory stimulus) and output (motor response). In such a situation, subjects could be scanned and their patterns of brain activity would constitute direct correlates of their subjective experiences (which they could later report for confirmation). Here we propose a general framework for such an experimental paradigm, starting from a specific example.

Individual participants are instructed to monitor street traffic in the role of security agent. Each participant is randomly assigned (blind to the investigator) to one of four groups, receiving the following instructions before entering the scanner.

- Group 1: If you see a blue sedan with darkened windows, that's the Ambassador's car on a classified mission. You should immediately follow it by pushing the joystick forward. If you see a black jeep with no license plates, that's a terrorist cell. You should stay calm and wait for backup by pulling the joystick backward.
- Group 2: If you see a black jeep with no license plates, that's the Ambassador's car on a classified mission. You should immediately follow it by pulling the joystick backward. If you see a blue sedan with darkened windows, that's a terrorist cell. You should stay calm and wait for backup by pushing the joystick forward.
- Group 3: If you see a blue sedan with darkened windows, that's the Ambassador's car on a classified mission. You should stay calm and wait for backup by pushing the joystick forward. If you see a black jeep with no license plates, that's a terrorist cell. You should immediately follow it by pulling the joystick backward.
- Group 4: If you see a black jeep with no license plates, that's the Ambassador's car on a classified mission. You should stay calm and wait for backup by pulling the joystick backward. If you see a blue sedan with darkened windows, that's a terrorist cell. You should immediately follow it by pushing the joystick forward.

In the scanner, all subjects see the same scene (blue sedan with darkened windows) and presumably respond in the same way (pushing the joystick forward). Thus, the sensory and motor controls are identical. However, some subjects see the Ambassador while others see a terrorist, and some subjects follow their targets while others hold their positions. Although all external behaviors were indistinguishable, mental states and intents were different (as later subjective reports can confirm).

In a variation of this experimental design, participants are instructed to monitor a meter representing a hypothetical patient's blood pressure. As the hospital chief surgeon, the participant must decide on surgical intervention on the basis of vital sign readings according to the following instructions for the four groups.

- Group 1: If the pressure drops below 100, the patient is finally stable. Begin surgery immediately by pushing the joystick forward. If the pressure rises above 150, the patient is in critical condition. Stay calm and continue monitoring the pressure by pulling the joystick backward.
- Group 2: If the pressure drops below 100, the patient is finally stable. Stay calm and continue monitoring the pressure by pushing the joystick forward. If the pressure rises above 150, the patient is in critical condition. Begin surgery immediately by pulling the joystick backward.
- Group 3: If the pressure drops below 100, the patient is in critical condition. Begin surgery immediately by pushing the joystick forward. If the pressure rises above 150, the patient is finally stable. Stay calm and continue monitoring the pressure by pulling the joystick backward.
- Group 4: If the pressure drops below 100, the patient is in critical condition. Stay calm and continue monitoring the pressure by pushing the joystick forward. If the pressure rises above 150, the patient is finally stable. Begin surgery immediately by pulling the joystick backward.

In the scanner, all subjects see a pressure drop and will push the joystick forward, yet half of them are intervening while half are not, and half of their patients are critical while the others are stable.

These two examples can be clearly extended into a generalized framework, in which an identical sensory stimulus induces different percepts, giving rise to different intents, and finally resulting in an identical motor response (Fig. 5). Thus, four different mental states can be distinguished on the basis of the combination of binary interpretation and intention possibilities, in the face of a single input-output pairing. This framework can be easily expanded to multiple choice scenarios and can be further controlled by repetition with other combinations of sensory stimulus (B instead of A in Fig. 5) or motor responses (K instead of J).

## Conclusions

The conscious mind is a natural phenomenon known to all of us, and as such it should be considered a topic of biology, psychology, and physics, as well as of the new scientific discipline outlined here and based on the recognition that science can and should describe quantitatively what the first-person perspective feels like. This requires, as the first essential steps, accepting the empirical observability of subjective experiences, developing a consistent metric system possibly based on semantic maps, and carefully distinguishing subjectively observable mental states from physically measurable input and output behaviors.

A few theoretical and experimental attempts have been made to relate human abstract cognition to well-understood biological mechanisms or reproducible brain correlates (*e.g.,* Ascoli, 1999b; Mitchell *et al.,* 2008). A systematic and robust account of these connections will eventually be necessary to interface computational neuroscience with cognitive science by characterizing the semantics of the neural code at the highest functional level. However, the majority

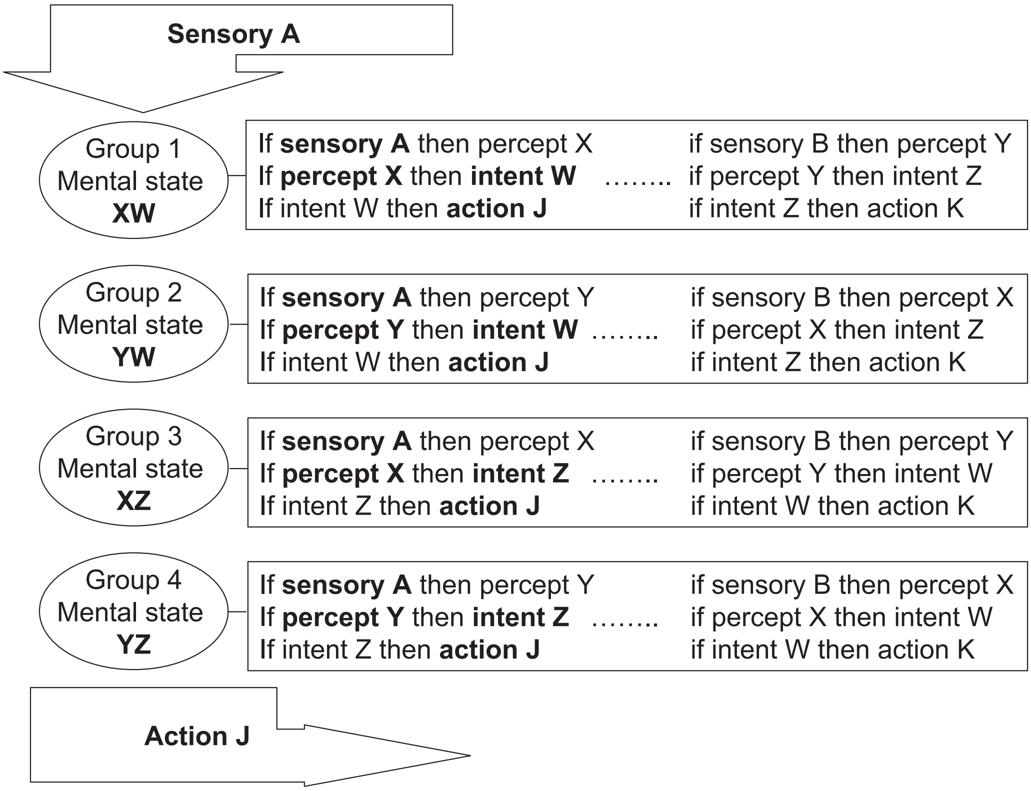
Sensory A
Group 1
Mental state
XW
If sensory A then percept X
If percept X then intent W ……..
If intent W then action J
if sensory B then percept Y
if percept Y then intent Z
if intent Z then action K
Group 2
Mental state
YW
If sensory A then percept Y
If percept Y then intent W ……..
If intent W then action J
if sensory B then percept X
if percept X then intent Z
if intent Z then action K
Group 3
Mental state
XZ
If sensory A then percept X
If percept X then intent Z ……..
If intent Z then action J
if sensory B then percept Y
if percept Y then intent W
if intent W then action K
Group 4
Mental state
YZ
If sensory A then percept Y
If percept Y then intent Z ……..
If intent Z then action J
if sensory B then percept X
if percept X then intent W
if intent W then action K
Action J

**Figure 5.** A general experimental framework to probe inner mental states. Subjects are divided into four groups, and each is given a context for interpreting two possible sensory inputs A and B (leading to distinct percepts X and Y, and intents W and J), as well as instructions for corresponding actions J and K (left and right column within each rectangle). All subjects, however, are then exposed to one and the same input A, and will accordingly output an identical action J. Nonetheless, each of the four groups instantiate unique combinations of perceptual and intentional mental states.

of cognitive neuropsychology reports are to date semiqualitative or do not directly address the characterization of the content of mental states. Studies that describe the content of mental states (as proposed in the last section) with a quantitative metric system (such as semantic maps) are still the exception in the current scientific trend. To create a new science of the mind, they need to become the rule. The recent strides of neuroscience and the breathtaking advances in computing power provide a fertile historical junction that can trigger a new scientific revolution if paralleled by a new conceptual framework and research paradigm for a science of the mind.

## Acknowledgments

The authors are indebted to Drs. Aurelio Ascoli, Kenneth A. De Jong, Gerald A. Goldin, Rebecca F. Goldin, M. Layne Kalbfleisch, Harold J. Morowitz, and James L. Olds for many inspiring discussions.

## Literature Cited

**Albus, J. S., and A. M. Meystel. 2001.** *Engineering of Mind: an Introduction to the Science of Intelligent Systems*. Wiley, New York.

**Albus, J. S., G. A. Bekey, J. H. Holland, N. G. Kanwisher, J. L. Krichmar, M. Mishkin, D. S. Modha, M. E. Raichle, G. M. Shepherd, and G. Tononi. 2007.** A proposal for a decade of the mind initiative. *Science* **317:** 1321.

**Aleksander, I. 2005.** *The World in My Mind, My Mind in the World: Key Mechanisms of Consciousness in Humans, Animals and Machines.* Imprint Academic, Exeter, United Kingdom.

**Anderson, J. R., D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. 2004.** An integrated theory of the mind. *Psychol. Rev*. **111:** 1036–1060.

**Arbib, M. A. 2005.** From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* **28:** 105–124.

**Ascoli, G. A. 1999a.** Is it already time to give up on a science of consciousness? A commentary on mysterianism. *Complexity* **5:** 25–34.

**Ascoli, G. A. 1999b.** Association, abstraction, and the emergence of the Self. *Noetic J.* **2:** 9–20.

**Ascoli, G. A. 2000.** The complex link between neuroanatomy and consciousness. *Complexity* **6:** 20–26.

**Ascoli, G. A. 2002.** *Computational Neuroanatomy: Principles and Methods.* Humana Press, Totowa, NJ.

**Ascoli, G. A. 2005.** Brain and mind at the crossroad of time. *Cortex* **4:** 619–620.

**Ascoli, G. A., and J. Grafman. 2005.** *Consciousness, Mind and Brain.* Massom Publisher, Milan, Italy.

**Ascoli, G. A., D. E. Donohue, and M. Halavi. 2007.** NeuroMorpho.Org:

A central resource for neuronal morphologies. *J. Neurosci.* **27:** 9247–9251.

**Carnevale, N. T., and M. L. Hines. 2006.** *The Neuron Book*. Cambridge University Press, Cambridge.

**Chalmers, D. J. 1996.** *Conscious Mind: in Search of a Fundamental Theory.* Oxford University Press, New York.

**Chomsky, N. 2006.** *Language and Mind,* 3rd ed, Cambridge University Press, Cambridge.

**CRAM (Cue-Recalled Autobiographical Memory Test). Date unknown.** [Online] WebCRAM version 3.1.0. Available http://cramtest.info [2008, September 30].

**Davison, A. P., T. M. Morse, M. Migliore, G. M. Shepherd, and M. L. Hines. 2004.** Semi-automated population of an online database of neuronal models (ModelDB) with citation information, using PubMed for validation. *Neuroinformatics* **2:** 327–332.

**Dawes, J. 2008.** Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int. J. Market Res*. **50:** 61–77.

**Dayan, P., and L. F. Abbott. 2005.** *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press, Cambridge, MA.

**Flanagan, O. J. 2007.** *The Really Hard Problem: Meaning in a Material World.* MIT Press, Cambridge, MA.

**Frankl, V. E. 1946.** *Man's Search for Meaning.* 1997 English ed., Washington Square Press, New York.

**Gärdenfors, P. 2004.** *Conceptual Spaces.* MIT Press, Cambridge, MA.

**Gardner, D., and G. M. Shepherd. 2004.** A gateway to the future of neuroinformatics. *Neuroinformatics* **2:** 271–274.

**Gerstner, W., and W. M. Kistler. 2002.** *Spiking Neuron Models: Single Neurons, Populations, Plasticity.* Cambridge University Press, Cambridge.

**Haikonen, P. O. 2003.** *The Cognitive Approach to Conscious Machines.* Imprint Academic, Exeter, United Kingdom.

**Hampe, B., and J. E. Grady. 2005.** *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. Cognitive Linguistics Research, 29. Mouton de Gruyter, Berlin.

**Kay, K. N., T. Naselaris, R. J. Prenger, and J. L. Gallant. 2008.** Identifying natural images from human brain activity. *Nature* **452:** 352–355.

**Kim, J. 1999.** *Philosophy of Mind*. Westview Press, Boulder, CO.

**Latent Semantic Analysis@CU Boulder. Date unknown.** [Online]. Available: http://lsa.colorado.edu [2008, September 30].

**Leary, T. 1957.** *Interpersonal Diagnosis of Personality.* Ronald Press, New York.

**Lee, C. K., S. M. Sunkin, C. Kuan, C. L. Thompson, S. Pathak, L. Ng, C. Lau, S. Fischer, M. Mortrud, C. Slaughterbeck, *et al.* 2008.** Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. *Genome Biol.* **9:** R23.

**Lewis, C. I. 1929.** *Mind and the World Order*. C. Scribner's Sons, New York.

**Lichtman J. W., J. Livet, and J. R. Sanes. 2008.** A technicolour approach to the connectome. *Nat. Rev. Neurosci.* **9:** 417–422.

**Likert, R. 1932.** A technique for the measurement of attitudes. *Arch. Psychol*. **140:** 1–55.

**McNaughton, B. L., F. B. Battaglia, O. Jensen, E. I. Moser, and M. Moser. 2006.** Path integration and the neural basis of the 'cognitive map.' *Nat. Rev. Neurosci.* **7:** 663–678.

**Metzinger, T. 2003.** *Being No One*. MIT Press, Cambridge, MA.

**Mitchell, T. M., S. V. Shinkareva, A. Carlson, K. Chang, V. L. Malave, R. A. Mason, and M. A. Just. 2008.** Predicting human brain activity associated with the meanings of nouns. *Science* **320:** 1191–1195.

**O'Reilly, R. C., and M. J. Frank. 2006.** Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput*. **18:** 283–328.

**Osgood, C. E, W. H. May, and M. S. Miron. 1975.** *Cross-Cultural Universals of Affective Meaning.* University of Illinois Press, Urbanna, IL.

**Palmer, S. E. 1999.** Color, consciousness, and the isomorphism constraint. *Behav. Brain Sci.* **22:** 923.

**Pinker, S. 2007.** *The Stuff of Thought: Language as a Window into Human Nature*. Viking, New York.

**Ploux, S., and H. Ji. 2003.** A Model for matching semantic maps between languages (French/English, English/French). *Comput. Linguist*. **29:** 155–178.

**Pulvermüller, F. 1999.** Words in the brain's language. *Behav. Brain Sci.* **22:** 253–279.

**Rosenthal, D. R. 1993.** Multiple drafts and higher-order thoughts. *Philos. Phenomen. Res.* **53:** 911–918.

**Rubin, D. C., and A. Wenzel. 1996.** One hundred years of forgetting: a quantitative description of retention. *Psychol. Rev.* **103:** 734–760.

**Samsonovich, A. V., and G. A. Ascoli. 2002.** Towards virtual brains. Pp. 423–434 in *Computational Neuroanatomy, Principles and Methods*, G. A. Ascoli, ed. Humana Press, Totowa, NJ.

**Samsonovich, A. V., and G. A. Ascoli. 2005a.** The conscious self: ontology, epistemology and the mirror quest. *Cortex* **41:** 621–636.

**Samsonovich, A. V., and G. A. Ascoli. 2005b.** A simple neural network model of the hippocampus suggesting its pathfinding role in episodic memory retrieval. *Learn. Mem.* **12:** 193–208.

**Samsonovich, A. V., and G. A. Ascoli. 2007a.** Computational models of dendritic morphology: from parsimonious description to biological insight. Pp. 91–113 in *Modeling Biology: Structure, Behaviors, Evolution,* M. D. Laubichler and G. B. Muller, eds. The Vienna Series in Theoretical Biology, MIT Press, Cambridge, MA.

**Samsonovich, A. V., and G. A. Ascoli. 2007b.** Cognitive map dimensions of the human value system extracted from natural language. Pp. 111–124 in *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms,* B. Goertzel and P. Wang, eds. Proceedings of the AGI Workshop 2006, Frontiers in Artificial Intelligence and Applications, vol. 157. IOS Press, Amsterdam, The Netherlands.

**Samsonovich, A. V., G. A. Ascoli, H. J. Morowitz, and M. L. Kalbfleisch. 2008.** A scientific perspective on the Hard Problem of consciousness. Pp. 493–505 in *Artificial General Intelligence 2008,* P. Wang, B. Goertzel, and S. Franklin, eds. Proceedings of the First AGI Conference, Frontiers in Artificial Intelligence and Applications, vol. 171. IOS Press, Amsterdam, The Netherlands.

**Sloman, A., and R. Chrisley. 2003.** Virtual machines and consciousness. *J. Conscious. Stud.* **10:** 133–172.

**Spitzer, M. 2008.** Decade of the mind. *Philos. Ethics Humanit. Med.* **3:** 7.

**Sporns, O., G. Tononi, and R. Kötter R. 2005.** The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **1:** e42.

**Swanson, L. W. 2007.** Quest for the basic plan of nervous system circuitry. *Brain Res. Rev.* **55:** 356–372.

**Thivierge J., and G. F. Marcus. 2007.** The topographic brain: from neural connectivity to cognition. *Trends Neurosci*. **30:** 251–259.

**Tononi, G. 2008.** Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* **215:** 216–242.

**Tononi G., and G. M. Edelman. 2000.** *A Universe of Consciousness: How Matter Becomes Imagination.* Basic Books, New York.

**Tononi, G., and C. Koch. 2008.** The neural correlates of consciousness: an update. *Ann. NY Acad. Sci.* **1124:** 239–261.

**Trappenberg, T. P. 2002.** *Fundamentals of Computational Neuroscience.* Oxford University Press, Oxford.

**Van Horn, J. D., S. T. Grafton, D. Rockmore, and M. S. Gazzaniga. 2004.** Sharing neuroimaging studies of human cognition. *Nat. Neurosci.* **7:** 473–481.