The Cognitive Binding Problem: From Kant to Quantum Neurodynamics

George A. Mashour

Abstract
The cognitive binding problem is a central question in the study of consciousness: how does the brain synthesize its modal and submodal processing systems to generate a unity of conscious experience? This essay considers several solutions to the binding problem, as well as their shortcomings. In particular, the current theory of neural synchronization as the basis for binding and consciousness is explored in its relationship to the relativity of simultaneity. This discussion of cognitive binding and simultaneity in the brain incorporates the philosophy of Kant, notably the principles of the transcendental unity of apperception and the transcendental aesthetic found in his *Critique of Pure Reason*. This leads to a more general consideration of consciousness and time, and explores the possibility of non-temporal theories of consciousness. The emerging field of quantum neurodynamics is discussed in this context, and its remarkable relationship to Kant is elucidated. Finally, the relevance of Kant's philosophy to cognitive binding is used as a basis for the discussion of a neurophilosophical method in the investigation of consciousness.

Key Words: cognitive binding problem, neural synchrony, neurophilosophy, Kant, quantum neurodynamics, consciousness

NeuroQuantology 2004; 1: 29-38

All theories that hold the neuron as the functional basis of consciousness must bridge a gap between a property of conscious experience and a fundamental tenet of neuronal processing. While evidence suggests that the brain subdivides perceptual processing into modality (e.g. the visual, the tactile) and submodality (e.g. color, temperature), our perceptions themselves are a unified experience. If the anatomic substrate of perceptual modality is functionally and spatially discrete neuronal subpopulations, how is information ultimately synthesized to create the manifest oneness of experience? This quandary is referred to as the cognitive binding problem, a term attributed to Christof von der Mahlsburg (1981). Binding is thought to occur at virtually all levels of perceptual (and motor) processing, and is thought to be a crucial event for consciousness itself (Crick and Koch, 1994). Although a relatively recent question in neuroscience, the binding problem may have made its first appearance in Immanuel Kant's *Critique of Pure Reason*. Kant's principle of a "transcendental unity of apperception" describes the synthesis of the "knowledge of the manifold" (Kant, 1965):

"There can be in us no modes of knowledge, no connection or unity of consciousness of one mode of consciousness with another, without that unity of consciousness which precedes all data of intuitions, and by relation to which representation of objects is alone possible... For the mind could never think its identity in the manifoldness of its representations, and indeed think this identity *a priori*, if it did not have before its eyes the identity of its act, whereby it subordinates all synthesis of apprehension (which is empirical) to a transcendental unity, thereby rendering possible their interconnection according to a priori rules."

The relevance of the unity of apperception to the binding problem, as well as the relationship of Kant's philosophy to cognitive science in general, has been previously described (Brook, 1994). There are, however, critical differences between the *a priori* principle of unity and some *a posteriori* event of neural binding. For Kant, one cannot posit consciousness, object, or self, without presupposing a unity that "precedes all data of intuitions." Nevertheless, Kant does address a "synthesis of apprehension" that is empirical, and thus *a posteriori*. There are therefore remarkable similarities between cognitive binding and the unity of apperception, and it is clear that understanding the source of unity is critical for both philosophic and neuroscientific theories of consciousness.

There have been various solutions proposed for the binding problem, which can be summarized as binding by convergence, binding by assembly, and binding by synchrony (for review see Singer, 1994; Singer, 1996; Neuron, September, 1999).

1. *Information is bound by higher order neurons that collect various responses and fire as a "binding unit" when the full set of inputs converge.*

This theory may be limited as it requires a higher order neuron or binding unit for every submodal or intermodal set of inputs that need to be bound. Second, even if this were the case, it would limit flexibility of higher order neurons, as they would have to be specific for the particular object bound. Furthermore, how would novel objects then be bound? This framework would require a population of uncommitted neurons maintaining latent connections. Finally, identifying neurons responsive to single objects (the so-called "grandmother cell") has been difficult, and met with limited success. It has also been demonstrated that even such higher-order neurons are still responsive to lower-order features. In short, "binding by convergence" does not seem to be an appropriate solution, although it has been suggested that it is a possible binding strategy for highly specialized representations.

2. *Binding occurs as a result of self-organizing Hebbian cell assemblies.*

In this proposed solution the single neuron or binding unit is replaced by a cell assembly whose interconnections define a set of neurons that need to be bound for a particular feature, object, etc. This allows a more dynamic flexibility than the schema above, because the same neuron is capable of participating in more than one cell assembly, and thus could participate in multiple binding patterns. This flexibility also results in a potential ambiguity: how is a particular neuron to be specifically associated with a single binding assembly, especially in the situation of superposed binding demands? In other words, if neuron "A" is a member of two cell assemblies that bind the objects "X" and "Y," what happens when both

"X" and "Y" need to be represented? This problem of superposition is a limiting feature of "binding by assembly."

3. *Binding results from the synchronized firing of neuronal subpopulations that are spatially discrete.*
Correlation of neurons in the temporal dimension has been proposed as a mechanism of unambiguously defining them as part of a binding assembly. Evidence for large-scale synchronization during a perceptual event has been obtained using electroencephalography (EEG) and magnetoencephalography (MEG) (Joliot et al, 1994; Rodriguez et al, 1999). Furthermore, spatially discrete neuronal subpopulations may be synchronized by resonance with a common structure that oscillates. An example of this would be cortical neurons resonating with the 40 Hz oscillation of thalamic neurons. While "binding by synchrony" successfully accounts for many features of the binding problem, it introduces the interesting issue of time in neuronal function. Although the concept of spatial mapping of neural activity has been considered since the 19th century, it has been noted that temporal mapping is "far more difficult to conceptualize, since its study requires an understanding of simultaneity in brain function not usually considered in neuroscience" (Llinas and Ribary, 1994).

BINDING AND SIMULTANEITY, CONSCIOUSNESS AND TIME

Although the study of simultaneity is a relatively recent development in neuroscience, it has been considered extensively in physics. In particular, the theory of relativity has radically changed our notion of time and simultaneity. It is of significance to note that the first theoretical step of Einstein's *On the Electrodynamics of Moving Bodies* is a definition of simultaneity (Einstein, 1952). He states unambiguously that all judgments of time are judgments of simultaneous events, and that such judgments presuppose an observer since the absolute time of Newtonian physics is incompatible with special relativity (Einstein, 1961). Determinations of time are determinations of simultaneity, and are thus dependent in principle on observation itself, i.e., they are dependent on consciousness.

Establishing simultaneity as the basis for consciousness leads to an interesting conclusion. Simultaneity is assumed to be a precondition of binding and consciousness (cognitive theory), while consciousness is itself a precondition of simultaneity (relativity theory). Thus, it appears that simultaneity and binding are somehow mutually conditioning. The relationship of cognitive binding and simultaneity may reflect the deeper relationships of consciousness and time. The notion of time (as well as space) being a feature or function *of* consciousness—rather than an objective physical entity perceived *by* consciousness—is once again attributable to Kant. He states in the Transcendental Aesthetic of the *Critique of Pure Reason*: "Time is not an empirical concept that has been derived from any experience. For neither coexistence nor succession would ever come within our perception, if the representation of time were not presupposed as underlying them *a priori*. Only on the presupposition of time can we represent to ourselves a number of things as existing at one and the same time (simultaneously) or at different times (successively)." The logician and mathematician Kurt Gödel argued that the Transcendental Aesthetic is supported by the theory of relativity (Gödel, 1990). Given events "A" and "B," it can be said with equal validity, depending on the frame of reference of the observer, that "A" happened at the same time as "B," before "B," or after "B." Gödel concluded that such phenomena support the absence of objective time (posited by Kant), as well as the absence of change posited by (Parmenides).

A Kantian perspective would assert that cognitive binding is not the result of simultaneous events, but rather the mechanism by which the categories of "simultaneous" and "successive" are delineated. There is neuroscientific evidence supporting this assertion. It has been previously shown that multiple auditory clicks occurring within a 12 ms timeframe are perceived by subjects as a single click (Joliot et al, 1994). The 12 ms threshold roughly corresponds to the phase shift of 40-Hz oscillations between the rostral and caudal poles of the brain, suggesting this transcortical electrical activity as the neural substrate for the observed perceptual binding. Thus, auditory events occurring within this 12 ms period were bound together as one. Clearly, binding must be a mechanism by which the brain represents simultaneous events (one click) versus successive events (two clicks). The 40-Hz resonance, however, is only of significance because it is a proposed mechanism by which the activity of spatially discrete neuronal subpopulations are synchronized. Thus, it is suggested *both* that cognitive binding can be explained in temporal terms ("synchronized" and "oscillation"), and that temporal terms (such as "simultaneous" and "successive") are themselves generated by cognitive binding. This circular conclusion seems to be a concrete example of the difficulty of interpretation that we may come upon when investigating time and the brain.

In order to have a theory of consciousness that does not lead to self-referential contradictions, perhaps we must, to some extent, "detemporalize" consciousness. One may argue that this is impossible given the fact that we perceive everything in time. It is the very fact that we must perceive in time, i.e., that time is a necessary *a priori* condition of our perception, that we are ultimately led to paradoxes of temporal theories of consciousness. Although at first glance such non-temporal theories may seem absurd, they have already been constructed by individuals in philosophy, physics, and neuroscience. Once again, Kant anticipates these arguments with his framework of the noumena and phenomena. Consciousness interacts with some intelligible objective reality (the noumena) to generate our sensible subjective reality (the phenomena). Because space and time are modes of consciousness rather than objective entities, the noumena are clearly non-spatial and non-temporal. The noumena/phenomena division bears similarity to the implicate/explicate framework put forth by the quantum physicist David Bohm. The implicate represents an enfolded, distributed reality that is brought into the explicate by consciousness (Bohm, 1980). This terminology was further adopted by the neuroscientist Karl Pribram, who describes a spectral domain (similar to the noumena or the implicate) and a space-time domain (similar to the phenomena or the explicate) (Pribram, 1997). Pribram emphasizes the possibility of bi-directional transformations between the spectral and space-time domains, and suggests that computational events of Fourier or Gabor functions serve to convert the potential/distributed/enfolded nature of the spectral domain into the actual/local/unfolded properties of the space-time domain. Thus, although we perceive in the space-time domain, the "deep structure" of consciousness is in the spectral domain.

Theories such as Pribram's and Bohm's may help us to integrate neuronal accounts of consciousness into meaningful frameworks that are compatible with physics and philosophy.

QUANTUM NEURODYNAMICS—A RETURN TO KANT?

The discussion of distributed domains of reality and consciousness leads us into a recent view of the brain that does not regard the neuron as a unit of cognitive perception, but rather regards the brain as one indivisible entity. This view is based on the emerging field of *quantum neurodynamics*. The brain, instead of being a Newtonian object obeying classical laws, is posited to be a macroscopic quantum object obeying the same laws found at the Planck scale (Penrose, 1994). The large-scale quantum coherence would render the brain a Bose-Einstein condensate, where properties of quantum wavefunctions hold at a macroscopic level. These properties would include non-locality and quantum superposition. Such large- scale coherence can be found in superconductivity and superfluidity, where the environment is disentangled from the quantum events. Superconductivity and superfluidity occur at temperatures just above absolute zero, where it is difficult for the environment to interfere with the quantum coherence. Herbert Fröhlich, however, predicted that quantum coherence could also occur in the temperature of a biologic environment (e.g. the brain) with high metabolic energy and extreme dielectric properties of the material involved (Fröhlich, 1984). It has been posited by Stuart Hameroff that the cytoskeletal elements of microtubules could be just such a material (Hameroff, 1987). Microtubules are composed of tubulin dimers, which can exist in multiple conformations based on dielectric properties, leading to dipole oscillations that could be transmitted through the length of the tubule. These superconducting "Fröhlich waves" could be transmitted throughout the protein network and gap junctions: from cytoskeleton to membrane protein to extracellular matrix to adjacent membrane protein to cytoskeleton, and so on. It is of interest that the supporting glial cells, the cellular majority in the brain, would contribute to this network, recalling Camillo Golgi's conception of the syncytium. It is currently a point of theoretical contention whether quantum coherence in the brain would decohere from a waveform collapse of superposed states (termed obective reduction) (Hameroff and Penrose, 1996), or from the "warm, wet, and noisy" environment of the brain (Tegmark, 2000). It has been suggested that the hollow core of microtubules, ordered water, and actin gelation may buffer the quantum events from the environment (Hagan et al, 2001). Magnetic resonance imaging of the brain by quantum coherence (albeit an induced artifact of the procedure) has also been suggested as proof of principle (Hagan et al, 2001).

Hameroff has suggested that quantum computation within the microtubules could be involved in consciousness, and proposes the activity of general anesthetics as support for this claim (Hameroff and Watt, 1983; Hameroff, 1998). He has argued that the similar activity of diverse chemical structures within the class of anesthetics can be explained by a common action of binding to hydrophobic domains and modulating dipole moments in the tubulin components of microtubules. This is supported by the fact that anesthetics inhibit prokaryotic motility, a function mediated by microtubules. It is of interest that recent studies using quantitative EEG suggest that anesthetics of various pharmacologic properties may act by interrupting cognitive binding (for review of these findings see John, 2002).

It is critical to note, however, that anesthetics and other substances that modulate consciousness are also known to act through specific neuronal receptors mediated by specific intracellular second messenger systems. This raises the important question of how cognitive activity mediated by quantum events in microtubules could interact with cognitive activity mediated by the essentially classical events in neurons. Metabolic energy generated through the local production of ATP may influence quantum-electrical events such as Fröhlich waves. Conversely, the electrical events in cytoskeletal proteins can affect membrane proteins, especially voltage-sensitive ion channels. It is not as yet clear how the classical aspect of neuronal function can modulate the quantum aspect without disrupting its coherence. Nonetheless, it appears that there is a potentially bi-directional communication between the quantum and classical components of brain function. It is of interest to speculate whether these communicating components are related to the spectral and space-time domains of consciousness proposed by Pribram, and whether Gabor functions are descriptive of their information transfer.

The quantum theory of the brain is as yet hypothetical and theoretical, with no empirical confirmation. It is of interest, however, to examine how quantum neurodynamics explains cognitive binding and unity of consciousness. First, because it is a theory that views the brain as a quantum unity in itself, it eliminates the problem *a priori*. In short, if there are no spatially discrete information processors, then there is no binding problem. Binding ceases to be a difficulty for the brain to solve *a posteriori*, but simply follows from the quantum structure, i.e. the quantum unity, of the brain. This unity, however, transcends the mere interconnectedness of microtubules and other brain proteins. Quantum unity also implies quantum non-locality, in which entangled particles can influence one another instaneously.

Such entanglement and non-locality has long been confirmed (Aspect and Grangier, 1986), suggesting a unity of two particles that is independent of spatial separation and temporal constraints on information transfer. If the brain has such quantum entanglements, then the consequences are remarkable. The brain would have a unity that is independent of spatial and temporal relationships, i.e., it would have a unity that is transcendental in the Kantian sense. While entangled quantum particles are clearly extant in time and space, their *unity* is not, because it defies common intuitions of spatiotemporal causality. This form of quantum unity might ground theories of cognitive binding, because the presupposition of time becomes unnecessary. Indeed, it has been proposed that the ontologic interpretation of quantum physics can be described independently of spatial and temporal terms (Hiley, 1998).

Quantum neurodynamics does not merely pose an *a posteriori* solution to the binding problem, but rather creates a framework within which binding is an *a priori* principle rather than a problem at all. The result is that cognitive binding based on quantum unity would be a return to Kant's concept of a transcendental unity of perception, and may form a link between old philosophical questions and current neuroscientific theories. Quantum brain theory is also proposed to be complementary to conventional neuronal theories of cognitive processing, and therefore need not be inconsistent with other proposed mechanisms of binding. It has, indeed, been suggested by Hameroff that the brain as Bose-Einstein condensate is consistent with the theories of binding by assembly and binding by synchrony. Thus, the view of quantum neurodynamics satisfies Kant's conception of both the empirical or *a posteriori* synthesis (neuronal/classical binding) and the *a priori* unity to which it is subordinate (quantum binding).

IMPLICATIONS FOR A NEUROPHILOSOPHICAL METHOD

The foregoing analysis is remarkable in that its conclusion in quantum physics was a return to its origin in Kant. It is further remarkable because concepts developed introspectively by Kant were relevant, indeed prescient, with respect to physical and neuroscientific data and theory. This suggests a potentially deeper relationship between philosophy, neuroscience, and physics within the investigation of consciousness. Although the need for a multidisciplinary discussion of consciousness has finally been recognized, we are not yet at the stage of a multidisciplinary method. The implicit relationship of Kant to the physical and neuroscientific aspects of the cognitive binding problem supports the general possibility of an explicit relationship between philosophy and neuroscience.

The current hypothetical grounding for experiments in neuroscience is invariably the conclusions of other experimental projects. It is not being suggested that this successful normative approach be replaced with a neurophilosophical method. It is being suggested, however, that neuroscientific experimentation based on philosophical "data" may engender access to information in a way that bypasses the normative course. Consider the unity of consciousness and binding problem discussion as an example. Kant's discussion of a unity of apperception stands in direct relationship to his discussion of the various faculties of the mind. A unity of consciousness was a natural and necessary development from the concept of faculties, given the nature of our perception. Consider now the neuroscientific pathway to the binding problem. It was perhaps a century between the empirical recognition of spatial mapping in the brain and the conceptual and experimental genesis of the binding problem. In short, while Kant proposed "faculty" and "unity" contemporaneously in the *Critique of Pure Reason*, experimental neuroscience only arrived at questions of "unity" after an extensive normative investigation of "faculty." It is thus suggested that the binding problem could have made its appearance neuroscientifically had the philosophical relationship to Kant been considered and disclosed at an earlier time. It would undermine the genius of Kant to suggest that any philosophical work possesses such a rich relevance to neuroscience, but have we yet made a thorough consideration of the prospective experimental value of philosophy?

It is not within the scope of this essay to make a full consideration of a potential neurophilosophic method, and the author is aware that more questions than answers are created by its incomplete presentation here. Its introduction, however, is appropriate in light of the neurophilosophic consideration of the binding problem. In particular, Kant's concept of transcendental unity may be a philosophical formulation of quantum unity, and thus a quantum hypothesis for the unity of consciousness could potentially be considered through one philosophical-experimental method.

CONCLUSION

Cognitive binding appears to be essential for cognitive activities ranging from lower order processes to consciousness itself. Furthermore, the concept of a unity of consciousness is also considered to be an essential philosophical principle according to Kant. Various solutions to the binding problem have been proposed, including mechanisms of convergence, assembly, and synchrony. The most developed of these theories—binding by synchrony—introduces complexities of time and consciousness that appear paradoxical. These paradoxes suggest the need for complementary, albeit counterintuitive, theories that incorporate a non-temporal and even a non-spatial component to binding and consciousness. Bohm and Pribram have developed such models, reflecting Kant's delineation of noumena and phenomena. Similarly, quantum neurodynamics has provided an account of binding that eliminates the contradictions found when dealing with time and consciousness, and provides a neuroscientific expression for Kant's concept of the transcendental unity of apperception. Finally, the relevance of Kant's thought to multiple aspects of neuroscientific and physical theories of consciousness heralds the possibility of a neurophilosophical method.

REFERENCES

Aspect A, Grangier P. Experiments on Einstein-Podolsky-Rosen-type Correlations with Pairs of Visible Photons, in *Quantum Concepts in Space and Time* (eds. Penrose R, Isham CJ) (Oxford University Press, Oxford, United Kingdom, 1986).

Bohm D. *Wholeness and the Implicate Order* (T.J. Press (Padstow) Ltd., Padstow, Cornwall, Great Britain, 1980).

Brook A. *Kant and the Mind* (Cambridge University Press, New York, New York, 1994).

Crick F, Koch C. Some Further Ideas Regarding the Neuronal Basis of Awareness, in *Large-Scale Neuronal Theories of the Brain* (eds. Koch C, Davis J) (MIT Press, Cambridge, Massachusetts, 1994).

Einstein A. On the Electrodynamics of Moving Bodies, in *The Principle of Relativity* (Dover, New York, New York, 1952).

Einstein A. Relativity: the Special and the General Theory (Crown Publishing, New York, New York, 1961).

Fröhlich H. General Theory of Coherent Excitations in Biological Systems, in *Nonlinear Dynamics in Biological Systems* (eds. Adey, W.R., and Lawrence, A.F.) (Plenum Press, New York, New York, 1984).

Gödel K. A Remark About the Relationship Between Relativity Theory and Idealistic Philosophy, in *The Collected Works of Kurt Gödel Vol II* (eds. Feferman et al.) (Oxford University Press, Oxford, United Kingdom, 1990).

Hagan S, Hameroff S, Tuszynski J. Quantum computation in brain microtubules: decoherance and biological feasibility. Phys Rev E Stat Nonlin Soft Matter Phys 2002; 65(6 Pt 1): 061901

Hameroff SR, Watt RC. Do anesthetics act by altering electron mobility? Anesth Analg 1983; 62:936-940.

Hameroff SR. *Ultimate Computing: Biomolecular Consciousness and Nano-technology* (North-Holland, Amsterdam, 1987).

Hameroff SR, Penrose R. Conscious events as orchestrated spacetime solutions. Journal of Consciousness Studies 1996; 3(1):36-53.

Hameroff S. Anesthesia, consciousness and hydrophobic pockets. Toxicology Letters 1998; 100/101:31-39.

Hiley B. Quantum theory, the implicate order and the mind/matter relationship. Toward a Science of Consciousness, Tucson III 1998.

John ER. The neurophysics of consciousness. Brain Res Rev 2002; 39:1-28.

Joliot M, Ribary U, Llinas R. Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. Proc Natl Acad Sci 1994; 91:11748-11751.

Kant I. *Critique of Pure Reason* (trans. Norman Kemp Smith) Discussions of Transcendental Unity of Apperception pp. 135-161, Transcendental Aesthetic pp. 74-81, Phenomena and Noumena pp. 257-275, (St. Martin's Press, New York, New York, 1965).

Llinas R, Ribary U. Perception as an Oneiric-like State Modulated by the Senses, in *Large-Scale Neuronal Theories of the Brain* (eds. Koch C, Davis J) (MIT Press, Cambridge, Massachusetts, 1994).

Penrose R. *Shadows of the Mind* (Oxford University Press, Oxford, United Kingdom, 1994).

Pribram KH. The Deep and Surface Structure of Memory and Conscious Learning: Toward a 21st Century Model, in *Mind and Brain Sciences in the 21st Century* (ed. Solso RL) (MIT Press, Cambridge, Massachusetts, 1997).

Rodriguez E, George N, Lachaux JP, Martinerie J, Renault B, Varela FJ. Perception's shadow: long-distance gamma band synchronization of human brain activity. Nature 1999; 397:430-433.

Singer W. Putative Functions of Temporal Correlations in Neocortical Processing, in *Large-Scale Neuronal Theories of the Brain* (eds. Koch C, Davis J) (MIT Press, Cambridge, Massachusetts, 1994).

Singer W. Neuronal Synchronization: A Solution to the Binding Problem? in *The Mind-Brain Continuum* (eds. Llinas R, Churchland PS) (MIT Press, Cambridge, Massachusetts, 1996).

Tegmark M. The importance of quantum decoherence in brain processes. Phys Rev E 2000; 61:4194-4206.

von der Mahlsburg C. The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81-2 1981.

von der Mahlsburg C. The Binding Problem of Neural Networks, in *The Mind-Brain Continuum* (eds. Llinas R, Churchland PS) (MIT Press, Cambridge, Massachusetts, 1996).