Disambiguation, Binding, and the Unity of Visual Consciousness

Eric LaRock
OAKLAND UNIVERSITY

Abstract. Recent findings in neuroscience strongly suggest that an object’s features (e.g., its color, texture, shape, etc.) are represented in separate areas of the visual cortex. Although represented in separate neuronal areas, somehow the feature representations are brought together as a single, unified object of visual consciousness. This raises a question of binding: how do neural activities in separate areas of the visual cortex function to produce a feature-unified object of visual consciousness? Several prominent neuroscientists have adopted neural synchrony and attention-based approaches to explain object feature binding. I argue that although neural synchrony and/or attentional mechanisms might function to disambiguate an object’s features, it is difficult to see how either of these mechanisms could fully explain the unity of an object’s features at the level of visual consciousness. After presenting a detailed critique of neural synchrony and attention-based approaches to object feature binding, I propose interactive hierarchical structuralism (IHS). This view suggests that a unified percept (i.e., a feature-unified object of visual consciousness) is not reducible to the activity of any cognitive capacity or to any localized neural area, but emerges out of the interaction of visual information organized by spatial structuring capacities correlated with lower, higher, and intermediate levels of the visual hierarchy. After clarifying different notions of emergence and elaborating evidence for IHS, I discuss how IHS can be tested through transcranial magnetic stimulation and masking. In the final section I present some further implications/advantages of IHS.

Key Words: ambiguity, attention, binding, consciousness, disambiguation, emergentism, object perception, spatial structure, unified experience

A fundamental issue in philosophy and related neurosciences is the object feature binding problem. Empirical evidence indicates that an object’s features (e.g., its color, texture, shape, etc.) are represented in separate areas of the visual cortex (Felleman & van Essen, 1991). In fact, the neural connections that exist between certain areas of the visual cortex, such as those that
represent color and motion, are sparse at best (Bartels & Zeki, 2006; Zeki, 2003). It is also unlikely that a single area in the brain could underlie the unity of an object’s features at the level of visual consciousness because of the combinatorial capacity problem: the possible combinations of features that are represented throughout our lives would seem to far exceed the neural machinery with which we are equipped (von der Malsburg, 1996, 1999). The question of how neural activities in separate areas of the visual cortex function to produce a feature-unified object of consciousness is often referred to as the object feature binding problem (see Bayne & Chalmers, 2003; de Kamps & van der Velde, 2001; Milner 1974; von der Malsburg, 1981, 1996, 1999). Some prominent neuroscientists argue that neural synchrony and/or attention could play the role of binding an object’s features into a unified object (Crick & Koch, 1990; Engel, 2003; Luck & Beach, 1998; Milner, 1974; Raffone & van Leeuwen, 2003; Roelfsema, 1998; Singer, 1996, 1999; von der Malsburg, 1996, 1999). As von der Malsburg (1996) observes, ‘If, during a time interval, the signals on a set of neurons are found to be significantly correlated, the set is interpreted as being bound during that interval’ (p. 137). De Kamps and van der Verlde (2001) point out that a related problem arises if multiple objects are present in the subject’s visual field: for example, if a red circle and a blue square were present in the subject’s visual field, the brain might easily mark features as belonging to the wrong object. Why does the brain not represent the circle as blue and the square as red? This is the feature ambiguity problem (see also Crick, 1994; Engel, 2003; von der Malsburg, 1996, 1999). A solution to the feature ambiguity problem would require explaining how the brain is able to disambiguate an object’s features (i.e., mark, or selectively tag, the correct set of features as members that belong to its respective object) when two or more objects are present in the subject’s visual field. One proposal is that synchrony and/or attention plays a critical role in selectively tagging the distinct set of neurons that correlate with an object’s respective stimulus features (see Crick & Koch, 1990, 2003; Luck & Beach, 1998; Raffone & van Leeuwen, 2003; Reynolds & Desimone, 1999; Tanaka, 1993; von der Malsburg, 1996, 1999). For example, Engel (2003) says that ‘the synchrony would selectively tag the responses of neurons that code for the same object and demarcate their responses from those of neurons activated by other objects’ (p. 134). I argue that although synchrony and/or attention might function to disambiguate an object’s features, it is difficult to see how either of these mechanisms could fully explain the unity of an object’s features at the level of visual consciousness. In addition to presenting a critique of neural synchrony and attention-based approaches to object feature binding, I develop an alternative view called interactive hierarchical structuralism (hereafter, IHS). This view suggests that a unified percept (i.e., a feature-unified object of visual consciousness) emerges out of the interaction of visual information organized by spatial structuring capacities correlated with lower, higher, and intermediate levels of the visual hierarchy.
The paper begins with a discussion of the visual system and vision-related binding problems. The work of von der Malsburg (1996, 1999) is then utilized to elaborate the neural synchrony approach to object feature binding. In the context of articulating the limitations of the neural synchrony approach, I also discuss why an attention-based approach to binding will not suffice as an explanation of object feature unity in visual consciousness. After the limitations of the neural synchrony and attention-based approaches are presented, I elaborate the IHS view.

Distributed Visual Areas

Recent findings in neurobiology strongly suggest that visual consciousness depends on the specialized activity of neuronal areas distributed throughout the visual hierarchy (Bartels & Zeki, 2006; Felleman & van Essen, 1991). Neurons in areas V1 and V2 respond to primitive visual features, such as contours and edges, in a variety of orientations (Grosof, Shapely, & Hawken, 1993). Neurons in V1 exhibit inhibitory or excitatory responses when their receptive field properties are sufficiently stimulated by objects that have certain orientation preferences. The receptive field of a neuron refers to a specified area in which a stimulus will excite or inhibit a neuron’s firing rate (Kalat, 1992). Visual information about an object or event is fed forward to specialized areas in the ventral and dorsal pathways, also known as the ‘what’ and ‘where’ systems (Mishkin, Ungerleider, & Macko, 1983). The ventral system’s neuronal pathway projects from the occipital lobe down to the inferior temporal lobe (IT), and plays the functional role of processing information about an object’s shape, color, and texture. The dorsal system’s neuronal pathway projects from the occipital lobe up to the parietal lobe (7a), and performs the functional role of processing information about an object’s location and size. The middle temporal and middle superior temporal lobes (MT and MST) process movements associated with individual objects or collections of objects. Thus, visually perceiving a rattlesnake slithering in the sand would involve the activities of at least V1–4, IT, 7a, MT, and MST (see Engel, 2003; Kosslyn & Koenig, 1995; Stein, 1992).

Neuropsychological evidence confirms this distributed view of visual feature representations. For example, damage in V4, an area of the ventral system, produces achromatopsia, that is, color blindness (Meadows, 1974); damage in IT produces associative agnosia, that is, the inability to identify shapes (Kosslyn & Koenig, 1995); and damage in MT produces akinetopsia, that is, motion blindness (Zihl, von Cramon, & Mai, 1983). Moreover, specialized neuronal areas have been identified within the dorsal system, which correlate with spatial attention, spatial representation, and the ability to differentiate within and between objects (Robertson, 1999). Thus, what we currently know about the visual system is that it relies upon several specialized subsystems...
distributed throughout the visual hierarchy. Even though this functional specialization has enabled the brain to process information efficiently, it also implies several vision-related binding problems.

**Binding Problems for Visual Consciousness**

Because of the distributed nature of feature representations, it is difficult to see how a feature-unified object arises in visual consciousness. In other words, how do we go from a distributed account of feature representations at the neurobiological level to an explanation of a feature-unified object at the conscious level? If there were direct correlations between feature representations in distributed neural areas and visual experience, it would seem that visual experience would consist of an unconnected set of features minus object unity. Normal subjects, in any case, do not visually experience objects as disunities, and so merely telling a story about the neural correlates of the distinctive feature representations of an object cannot be the complete story. Some have suggested that **combination-coding cells** could solve the object feature binding problem. Defenders of this view have hypothesized that a single unit/neuron could play the role of object feature binding. Von der Malsburg (1996) points out, however, that the possible combinations of visual features we experience throughout our lives would far exceed such binding units: ‘Our nervous system cannot afford to contain combination-coding neurons that represent all possible bindings, combinatorics quickly leading to astronomical numbers’ (p. 135; see also van der Velde & de Kamps, 2002, 2006). Presumably, the brain evolved specialized subsystems in order to process information efficiently in an attempt to solve the combinatorial explosion problem. Paradoxically, the distributed character of feature representations within and across the visual cortex has justifiably led neuroscientists and philosophers to conclude that the object feature binding problem is a serious vision-related binding problem in its own right. The question of how neural activities in separate areas of the visual cortex operate to produce a feature-unified object of consciousness is usually referred to as the object feature binding problem. Bayne and Chalmers (2003) articulate this problem as follows:

> The notion of objectual unity is closely tied to a central issue in cognitive psychology and neurophysiology. When I look at a red square, the color and the shape may be represented in different parts of my visual system. But somehow these separate pieces of information are brought together so that I experience a single red square. ... This phenomenon is often referred to as binding, and the question of how it is achieved is often referred to as the binding problem. (p. 25)

Because evaluating neural synchrony and attention-based approaches to the object feature binding problem are central targets of this paper, I will mention only briefly one further vision-related binding problem.
Scenes are comprised of objects located in space. Explaining how each object is bound to its present location is also an important vision-related binding problem. This form of binding is known as ‘location binding’ (Treisman, 1996, p. 171). In other words, since recognizing objects is a function of the ventral system and locating them is a function of the dorsal system, it is critical to show how the visual system connects ‘what’ properties to ‘where’ properties. I might also mention that location binding is a species of inter-spatial binding, since its principal function involves binding objects to their present locations relative to other objects. For example, there is an external spatial relation that holds between a vertically oriented cup and a table’s horizontal surface when the cup is on top of the table (see Hummel & Biederman, 1992). Given combinatorial possibilities, the cup could have been to the left of or below the table, etc. Bayne and Chalmers (2003) have aptly dubbed the experience of this external aspect of spatial relatedness spatial unity: ‘We can say that two conscious states are spatially unified when they represent objects as being part of the same space’ (p. 25). The location of an object can change both egocentrically (i.e., relative to the head or body) or allocentrically (i.e., relative to the locations of other objects in the environment). These novel object locations must somehow be represented in visual consciousness.

I should mention that one can study this specific binding issue from the perspective of a location binding question. To solve the problem of location binding, then, a theory of visual cognition must be able to answer questions such as ‘Where is this object?’ It is important to mention that some compelling empirical work has already been done to answer this location binding question (see, e.g., de Kamps & van der Velde, 2001). This does not suggest that the mechanisms involved in location binding are not also involved in object feature binding (van der Velde & de Kamps, 2006). In fact, IHS takes this insight into account. (See Treisman, 1996, 2003, for a discussion of other vision-related binding problems.)

Von der Malsburg’s Approach to Object Feature Binding

Let us now consider the work of von der Malsburg (1981, 1996, 1999). Von der Malsburg argues that the neural synchrony hypothesis can shore up the deficiencies implicit in the classical neural network approach to object representation. Although the classical framework has helped to explain how short- and long-term memory operates in our cognitive architecture, it nonetheless has difficulty modeling typical visual scenes comprised of multiple objects (von der Malsburg, 1996, pp. 131–132; 1999, pp. 95–96). Von der Malsburg asks us to consider how a neural network model could represent two visual objects (a triangle and a square) at the same time. The neural network model has an internal structure that enables it to extract four distinct propositions; the propositions themselves are represented by specialized output neurons.
Two of those output neurons represent the shapes of objects: one neuron represents triangle and the other neuron represents square. The other two output neurons represent locations of the objects (top and bottom) relative to their position on the retina (von der Malsburg, 1996, p. 134).

When the connectionist model is presented with only one object, it has no problem representing the correct shape with the correct location (triangle, top) or (square, bottom). Nevertheless, the classical model has difficulty representing the correct shape with the correct location when two or more objects are present in the subject’s visual field (von der Malsburg, 1996, p. 134; also 1999, p. 96). As von der Malsburg (1999) observes:

A problem arises, however, when two objects are presented simultaneously. If the output reads (triangle, square, top, bottom) it is not clear whether the triangle or the square is in the upper position. This is the binding problem: the neural data structure does not provide for a means of binding the proposition top to the proposition triangle, or bottom to square. ... The difficulty is that simply coactivating the elementary symbols leads to binding ambiguity when more than one composite symbol is to be expressed. (p. 96)

The notion of combination coding is a standard reply to the feature binding problem. On the combination-coding hypothesis, a single neuron represents both shape and locational features of an object. Although theoretically sound, this solution entails an untenable outcome. The virtually infinite combinations of attributes that are coded throughout one’s visual life would far exceed the purported combination-coding units (von der Malsburg, 1996, pp. 134–135). Furthermore, it seems plausible to think that evolution would have equipped the brain with a flexible and efficient capacity of representing ‘all combinations that can ever play a role in our life’ (von der Malsburg, 1996, p. 135). Von der Malsburg suggests that what is needed in the classical neural network architecture is a dynamic mechanism that would obviate the requirement of combination-coding cells.

Von der Malsburg (1981, 1996) proposes the dynamic link architecture. The dynamic link architecture accepts the classical connectionist idea that neurons are to be understood as ‘elementary symbols’ (1996, pp. 137–138). It also emphasizes two important temporal characteristics of neurons: (1) the speed of neural firing indicates the ‘intensity with which the elementary symbol is presently active’ (1996, p. 137); (2) the coordinated activity among a group of cells is indicative of its ‘fine temporal signal structure’ (1996, p. 137). These temporal dimensions, von der Malsburg (1996) believes, have important implications for a plausible view of binding: ‘If, during a time interval, the signals on a set of neurons are found to be significantly correlated, the set is interpreted as being bound during that interval’ (p. 137). To substantiate this claim, von der Malsburg appeals to experimental data, which shows that neurons exhibit temporal synchrony in response to moving figures (see Engel, Konig, Kreiter, Schillen, & Singer, 1992; see also Singer, 1996, 1999). I might add that neural synchrony has been observed between neurons whose locale is in
entirely different hemispheres of the brain. Thus, the neural synchrony view is consistent with an account of large-scale temporal correlation, not simply local temporal correlation (see Varela & Thompson, 2003).

In a spirit of fairness, however, von der Malsburg (1996) anticipates and articulates a rejoinder on behalf of the combination-coding advocate: the combination-coding proponent will point out that the empirical support for combination-coding neurons shows that binding does not require a temporal notion of coding (p. 139). But this response, von der Malsburg argues, fails to understand the nature of the problem: ‘The binding problem arises only when the combination-coding neurons in the brain cannot disambiguate the situation’ (p. 139, italics added). By implication, von der Malsburg believes that if one could show how the brain disambiguates the visual situation, then one would have a solution to the binding problem.

So, how does the dynamic link architecture disambiguate the visual situation? Though the classical model is blind to fine temporal signals, the dynamic link architecture emphasizes that neurons are sensitive to incoming temporal signals and can be correlated in time (von der Malsburg, 1996, pp. 141–142). This notion of neural dynamics is compatible with Abeles’ (1991) idea of synfire chains, which refer to chains of distinct groups (‘pools’) of neurons whose activation patterns occur in succession: ‘A synfire chain can be traversed by an activity process in which all the cells in a pool fire simultaneously and thus succeed in firing the cells in the next pool simultaneously’ (von der Malsburg, 1996, p. 142). There is temporal synchrony within each neuronal group without temporal overlap between different groups. These temporal distinctions have an important consequence: temporal differentiation between neural groups entails disambiguation between the features of distinct objects. Additionally, synfire chains express the flexible character of neurons: each neuron has the capacity to perform its specialized function in several synfire chains, since its participation is temporally differentiated. For example, the neurons that represent the color of object A could also function to represent the color of object B insofar as the distinct feature representations of A and B were generated at separate times. The concept of synfire chains underscores the idea that neurons operate in a flexible and adaptive fashion, which is precisely what we would expect as evolved beings. Along with Bienenstock (1994), von der Malsburg (1996) maintains that synfire chains are the fundamental building blocks of our dynamic cognitive architecture (pp. 141–143).

Disambiguation is ultimately accomplished by the temporal correlation of neural signals within a segment and the de-correlation of neural signals in different segments of the visual scene. As von der Malsburg (1996) explains,

The dynamic link architecture affords the infrastructure required for scene representation. Complex (sensory) patterns can be segmented in a meaningful way into subpatterns that correspond to separate objects or functional components of the scene. This is done by temporally correlating signals within a segment and decorrelating signals between elements in different segments. (p. 143)
The dynamic link architecture seems to provide a framework on how the brain might represent a scene’s component objects by means of significant temporal correlations among pools of neurons that underlie the distinctive representational features of objects at successive times via synfire chains. It is useful to recall that von der Malsburg maintains that the binding problem arises only when the brain cannot disambiguate the visual situation (see 1996, p. 139). By implication, he thinks the dynamic link architecture solves the binding problem because the brain is able to disambiguate a scene’s component objects by correlating the signals that respond to an object’s attributes within a segment, and by de-correlating the signals that respond to the attributes of objects in different segments.

A final reflection should be made in this context. Von der Malsburg’s analysis motivates a similar distinction between sustained simultaneous activation and activation with phase synchrony. The difference between these forms of activation is found in the treatment of displays with multiple objects, for example a red circle and a blue square. In this sense, the brain could disambiguate the features of the two objects in the display on grounds of activation with phase synchrony. Mechanisms of this sort provide a way of answering the question ‘Which color belongs to which shape?’ The ability to answer a question like this is a minimum requirement for any theory of object perception, in the sense that it would show how the correct features of an object are disambiguated (i.e., marked as members of a specific object) when two or more objects are present in the subject’s visual field. In this sense ‘synchronous activation of neurons could be used as a measure of the fact that neurons are responding to the same object’ (van der Velde & de Kamps, 2002, p. 292).

Limitations of Neural Synchrony and Attention-Based Approaches to Binding

First, neural synchrony probably cannot account for the enduring character of a feature-unified object representation because of its fleeting nature. For example, O’Reilly, Busby, and Soto (2003) point out that once an object is removed from a subject’s visual field and the neuronal assembly that subserves that object’s features ceases firing, there is no indication of that assembly’s neurons being bound and ‘yet we can have enduring representations of bound features, so somehow this problem needs to be addressed’ (p. 171). O’Reilly and colleagues are referring to our capacity to represent a feature-unified object over time (i.e., diachronic object unity). From a phenomenological point of view, the experience of diachronic object unity bears the distinct phenomenal feature of continuity, the enduring stream of consciousness that James (1890) and Dainton (2000) have so deftly characterized in their phenomenological analyses. Now, because the representation of diachronic object unity persists beyond the subpopulation of cells that fire in synchrony,
the representation of diachronic object unity is not reducible to the synchronous activation of neurons.

One might reply with the following short-term working memory (hereafter, SWM) objection. SWM probably functions to temporarily store visual and other modes of information for several seconds at a time. Thus, SWM might underlie the continuity (or diachronic character) of a feature-unified object representation in visual consciousness. To bolster this claim, one might appeal to the research of Patricia Goldman-Rakic (1987), who discovered that prefrontal neurons function to temporarily store locations of objects within one’s visual field. Without an awareness of temporarily stored spatial/locational information, individuals would have difficulty fulfilling behavioral tasks; presumably the spatial information temporarily stored by SWM could be what underlies the individual’s ability to guide the relevant motor outputs over time: for example, reaching for the red ball to my left instead of the blue ball to my right. The neural correlate of this spatial/locational mode of SWM is related to the association areas of the dorsal prefrontal cortex, which can be decomposed into three regions in reference to the principal sulcus: the cortical area surrounding the sulcus, as well as the cortical areas ventral and dorsal to the sulcus. Researchers have utilized delayed response tasks to corroborate these claims. One such task involved showing a morsel of food to a monkey; then, after a brief delay, the monkey was allowed to reach for the food. Researchers found that prefrontal neurons fire selectively in response to an object’s location and then continue to fire during the delay period of such tasks. Other experiments on monkey cortex show that a small lesion to the principal sulcus causes a deficit in SWM (Kandel, Schwartz, & Jessell, 2000, pp. 356–361). Hence, behavioral continuity depends on an awareness of the relevant spatial information temporarily stored by SWM. Apart from temporarily storing locational information, it would seem that SWM would have to play a role in transiently storing information about an object’s shape and color for the purposes of object recognition. The cortical region that is ventral to the principal sulcus is the likely region where SWM functions to transiently store information about an object’s ‘what’ properties, for example its color, shape, texture. These data suggest that SWM is modular, in the sense that separate neuronal areas selectively respond to and temporarily store information about an object’s ‘what’ and ‘where’ properties (Kandel et al., 2000).

Having looked at some of the neural correlates of SWM, we have yet to discuss the cellular mechanisms of working memory. Kandel’s research on the sea slug, Aplysia, is illuminating in this respect because it suggests that ‘the duration of short-term memory storage depends on the length of time a synapse is weakened or strengthened’ (Kandel, 2006, pp. 204–205). Kandel recognizes, however, that SWM is not located ‘at a single specialized site,’ but involves the synapses of thousands of neurons distributed throughout the relevant specialized cortical areas (p. 204; see also Kandel et al., 2000). Kandel is not alone in making the observation that SWM is distributed across specialized areas of the brain. For example, Crick and Koch (1990) also maintain:
Both iconic and working memory are likely to be distributed throughout the appropriate cortical areas, with auditory events transiently stored in auditory cortices, visual events in the visual cortices, and so on (p. 270; see also Bear, Connors, & Paradiso, 2001).

This raises an important binding question: if the cellular mechanisms that underlie SWM are distributed throughout each cortex, then why would binding problems not arise at the level of short-term working memory? For example, how does SWM bind an object’s ‘what’ and ‘where’ properties if the cellular mechanisms that correlate with SWM are distributed in separate neuronal regions? How does SWM bind an object’s color, shape, texture, and motion properties into a single unified object if the cellular mechanisms that temporarily store information about those properties are distributed in specialized areas of the ventral pathway and MT? Will not the spatial spread of the neural events that correlate with SWM thwart the unification process? Somehow the separate pieces of information regarding an object’s properties are bound together so that I am consciously aware of a single, property-unified object over time, despite the limitations of SWM. Based upon these data, one could infer that explaining how an object’s separate feature representations are temporarily stored (or transiently sustained) by SWM is one thing; while explaining how the distributed operations of SWM could bind such features into a unified object over time (i.e., diachronic object unity) is another matter. Which is to say: temporal storage does not entail diachronic object unity. Even if SWM is necessary for diachronic object unity, it is not also sufficient for it. Now, we might be tempted to suppose that a combinatorial site exists in the frontal lobes, since some neurons in that region respond to an object’s shape and location (Kandel et al., 2000). Though these neurons might be necessary to guide behavior, we have already observed that the supposition of a combinatorial site is improbable in virtue of combinatorial capacity limitations (see, e.g., Singer, 1996; von der Malsburg, 1996, 1999).

Perhaps there is another way to formulate the objection from SWM. Lisman and Idiart (1995) have proposed that multiple memories can be temporarily stored in a neural network by subcycles of non-overlapping oscillations, much like those that have been observed in the brain. Their model suggests that the slow increase in the post-depolarization state could be an important mechanism that stores different memories in different subcycles. The network’s capacity to parse different memories into different subcycles is also a function of an inhibitory mechanism: ‘The neurons receive continuous oscillatory input and pooled feedback inhibition, the function of which is to partition a cycle into subcycles’ (Lisman & Idiart, 1995, p. 1514). What do Lisman and Idiart say their model demonstrates? It shows how SWM could function to temporarily store and thereby ‘distinguish’ (i.e., disambiguate) the respective features of approximately seven objects ‘by phase (oscillatory subcycles)’ (p. 1514). While I do not disagree with this inference, I will later argue that object feature disambiguation \textit{per se} does not guarantee an account of object
feature binding. So, let us now proceed to the next limitation for the binding by neural synchrony hypothesis.

Second, some experimental research indicates that neural synchrony results from ‘moving stimuli and is notoriously difficult to measure with stationary stimuli’ (Hummel & Biederman, 1992, p. 509; see also Gray, 1999). In fact, the experimentation of Tovee and Rolls (1992) has shown that neuronal activity in IT of alert monkeys revealed no evidence of synchrony when the cells in IT were activated by stimuli that were stationary (see also Gray, 1999). This has dramatic consequences for the neural synchrony hypothesis. It is difficult to see how neural synchrony could be an adequate binding mechanism since some of the objects we perceive in both natural and artificial environments (e.g., mountains, boulders, houses, etc.) are stationary. Moreover, recent evidence derived from studies on primate visual cortex show that motion detection is not correlated with neural synchrony. For example, Thiele and Stoner (2003) constructed an experiment that employed perceptually coherent plaid patterns and two types of non-coherent plaid patterns to test the binding by neural synchrony (BBS) hypothesis. By exposing these distinctive patterns to alert monkeys, they could test the predictive value of BBS. If BBS is plausible, the coherent plaid should elicit synchronization in neuronal area MT of awake, fixating primates. However, the experimental results indicate that, contrary to the BBS prediction, the ‘synchrony elicited by our coherent plaids was statistically indistinguishable from that of non-coherent plaids’ (Thiele & Stoner, 2003, p. 367; see also Mashour, 2004). Of course, these specific objections do not entail the denial that other tests have been implemented in order to invoke synchronized oscillations with respect to other cognitive/behavioral tasks, such as motor and detection tasks (see Tallon-Baudry & Bertrand, 1999). These particular objections merely imply that neuronal synchrony falls short of a complete account of feature-unified objects of visual consciousness on grounds that it can explain neither stationary nor moving objects.

Third, the burgeoning body of experimental work used to corroborate the temporal correlation hypothesis has been obtained from both anaesthetized and awake animals, namely, cats and monkeys (see Engel, Konig, Gray, & Singer, 1990; Gray & Singer, 1989; Shadlen & Movshon, 1999; Tallon-Baudry & Bertrand, 1999). Indeed, Crick (1994) claims that considerable experimental work is performed on ‘anesthetized animals who are not conscious’ (p. 15, italics added). This seems to suggest that the activity of oscillatory synchronization in the gamma frequency range (approximately 35 to 75 Hz) occurs in response to visually presented objects in the brains of both conscious and unconscious animals. We, therefore, cannot maintain that oscillatory synchronization in the 35 to 75 Hz range is the distinctive neural correlate of a feature-unified object of visual consciousness. Furthermore, this type of oscillatory activity correlates strongly with pre-attentive awareness, that is, unconscious activity in V1 (see Gray, 1999; Luck & Beach, 1998). This finding motivates an important question: how could oscillatory synchronization in
the 35 to 75 Hz range be the direct neural correlate of a feature-unified object of visual consciousness if such ‘specialized’ neuronal activity correlates with pre-attentive awareness, that is, unconscious activity in V1? Now, this does not mean that neuronal synchronization could not play an important cognitive role in V1. Recent evidence, for example, has shown that neurons in different columns of V1 that respond to the same orientation will respond in a temporally correlated fashion to an object that exemplifies that orientation. This is consistent with the idea that the temporally correlated activities of lower-level neurons could play a role in binding the points that comprise oriented lines of objects (see Hardcastle, 1998; Terzis, 2001). It is likely that the receptive field properties of these neurons have become specialized (or ‘tuned’) over long periods of evolutionary development in order to carry out these basic operations. This might explain why the receptive fields of neurons in V1 are comparatively smaller than those at higher levels of the visual cortex. Because this lower-level form of binding is limited to the construction of oriented lines, and occurs at the pre-attentive level of visual information processing, V1 is probably not the locale (or neural correlate) of visual consciousness (see Gray, 1999). Empirical evidence suggests further that visual hallucination can take place for a period of time after V1 has been severely damaged. And persons with blindsight—a visual disorder caused by damage to cells in V1—can still be conscious of certain visual features, such as motion (Prinz, 2000). This is because the neuronal area that correlates with the awareness of motion receives a direct visual input that bypasses V1 (Zeki, 2003). Consequently, the evidence gathered thus far indicates that V1 might be necessary as a source of inputs to higher regions of the visual cortex, but could not be the neural correlate of a feature-unified object of visual consciousness (cf. Crick & Koch, 1995; see also Prinz, 2000).

Fourth, the research of Luck and Beach (1998) adds further reason to doubt that neural synchrony could suffice as the mechanism of object feature binding at intermediate/higher levels of the visual system. This is because the receptive fields of cells in these areas are too large, hence increasing the probability of accidental synchronizations of output neurons that correlate with the attributes of different objects:

> Temporal tagging alone is probably insufficient to solve the binding problem under realistic conditions of complex, multiple-object stimulus arrays. As the number of objects increases, the number of active units increases, and there is a dramatic increase in the probability of accidental simultaneity in the outputs of neurons that represent objects. (Luck & Beach, 1998, p. 459)

The principal reason why the problem of accidental synchronizations does not arise at the lower level of the visual system (i.e., V1) is that the receptive fields in this area are comparatively smaller than those at higher reaches of the visual cortex. What this suggests, then, is that one might consider supplementing this lower-level form of temporal binding with a higher-level form of binding that utilizes ‘some other mechanism’ (Luck & Beach, 1998, p. 461).
Luck and Beach (1998) and Crick and Koch (2003), as well as several other neuroscientists (see also Reynolds & Desimone, 1999), maintain that attention could be the missing explanatory piece of the feature binding puzzle associated with higher levels of the visual system. For example, although Crick and Koch (2003) no longer think that neural synchrony is the neural correlate of consciousness, they still maintain—in their revised neural coalitions view—that attention plays an essential role in binding an object’s features when competition occurs in the cortical neural network:

Several objects/events can be handled simultaneously—more than one object/event can be attended to at the same time—if there is no significant overlap in the cortical neural network. ... If there is such an overlap, then (top-down) attention is needed to select one of them by biasing the competition among them. This [attention-based] approach largely solves the classical binding problem, which was mainly concerned with how two different objects/events could be ‘bound’ simultaneously. On this view, the ‘binding’ of the features of a single object/event is simply the membership in a particular coalition. (p. 123)

Much like their previous theorizing, Crick and Koch propose that visual attention selects a coalition of neurons when competition occurs, and that it is in virtue of this attentionally selected coalition that the perceived unity of an object’s features is produced: ‘In general, at any moment the winning coalition is somewhat sustained, and embodies what we are conscious of’ (p. 121).

This inference, however, is not warranted. Even if an explanation of feature disambiguation could show how an object’s representational features are selected (‘marked’ or ‘labeled’) when competition occurs, this would not automatically guarantee an account of how such features, once selected, are bound together to form a unified object of consciousness. Which is to say: feature disambiguation does not entail feature binding. To provide further strength to this criticism, evidence from neuropsychology strongly indicates that attention is not necessary for binding, though it seems to play a central role in feature disambiguation at intermediate and higher levels of visual information processing. For example, a variety of experiments have been performed on normal individuals indicating that selecting the correct set of features when two or more objects are presented depends upon the focus of attention (Treisman, 1996, 2003). In one experimental set-up, individuals were briefly shown two colored letters at the same time: a green T and a red O. The experimenters found that when the focus of attention was prevented by means of a brief presentation of the letters, individuals would experience illusory conjunctions: the individuals reported seeing ‘a red T when a green T and a red O’ were presented at the same time (Treisman, 2003, p. 99). Consequently, red, rather than green, was bound to T and the result was an experience of illusory conjunctions. What is interesting is that these normal individuals have still performed the function of binding, albeit of an illusory conjunction sort. Therefore, focal attention is not necessary for binding.
Now, if binding can take place independent of the disambiguating role of attention (and/or some other neural mechanism), then an account of disambiguation does not automatically guarantee an account of binding. This empirically based insight directly challenges the widespread assumption that if one could show how the brain disambiguates the visual situation, then the binding problem has been solved. Von der Malsburg (1996) is not the only binding theorist who takes this assumption. For example, Tanaka (1993) infers that the binding problem ‘disappears’ on grounds that attention could disambiguate the features of competing objects: ‘If the representation of features of an attended object is enhanced and that of other objects is suppressed, the binding problem disappears’ (p. 687). The empirical data on illusory conjunctions presented above show why this inference is invalid. Moreover, it would not matter whether one appealed to attention, or neural synchrony, or some other neural mechanism to explain feature disambiguation; the above criticism would still hold. To illustrate this point, Engel (2003) supposes that the temporal synchrony account implies an ‘elegant solution’ to the problem of feature binding because temporal synchrony could ‘selectively tag’ the responses of neurons that code for the same object when competition occurs and hence ‘demarcate their responses from those of neurons activated by other objects’ (p. 134). How the brain selectively tags (i.e., disambiguates or ‘demarcates’) the correct stimulus features (when two or more objects are present in the subject’s visual field) is conceptually distinct from how the features of an object are brought together as a single, unified object of visual consciousness. It is, therefore, a conceptual mistake to run the questions of disambiguation and binding together. One can imagine, though, that some of these theorists might say, for example, that all they mean by ‘binding’ is feature disambiguation. However, just explaining feature disambiguation does not explain the harder question about the unity of visual consciousness. A further conclusion we might take from this discussion is that the functional role played by attention at intermediate and higher levels probably implies a normative dimension to the disambiguating process: the perceptual subject’s ability to select the correct set of features when competition occurs in the cortical hierarchy is directly tied to the subject’s attentional capacities.

An Interactive Hierarchical Structuralist Theory of Visual Consciousness

Having presented a critique of neural synchrony and attention-based approaches to object feature binding, I would now like to discuss interactive hierarchical structuralism (IHS). This view suggests that object feature binding could result from the interaction of information organized by spatial structuring capacities correlated with lower, higher, and intermediate levels of the visual hierarchy. IHS coheres with some of Lamme’s (2003) views on this issue. Like Lamme,
the IHS account draws a distinction between the initial feedforward sweep of information from lower to higher levels of the processing hierarchy and the subsequent feedback interactions of information between these levels. On the IHS account, a unified percept (i.e., a feature-unified object of visual consciousness) results from the subsequent feedback interactions. A metaphysical implication of IHS is that a unified percept is not reducible to the activity of any cognitive capacity or to any localized neural area, but emerges out of the interaction of visual information organized by spatial structuring capacities correlated with lower, higher, and intermediate levels of the visual hierarchy. This raises a crucial question.

What notion of emergence is compatible with IHS? It is important to point out that there are weak and strong notions of emergence in the philosophical and scientific literature. This has largely to do with the diverse kinds of criteria invoked to determine whether a property is emergent. Some philosophers who defend a weaker conception of emergence appeal to the criterion of non-deducibility to establish the emergence of consciousness, but wind up denying that consciousness, as an emergent property in relation to the brain, could produce top-down influences independent of the causal powers endemic to neural properties (cf. Varela & Thompson, 2003, pp. 274–275). After looking at a species of weak emergence, we will then examine a notion of strong emergence in order to motivate the IHS account.

Weak Emergence

A notion of weak emergence is found in the philosophical work of Searle (1992). Searle proposes an emergentist view of consciousness called biological naturalism. He bolsters his view by drawing distinctions between (1) the elements that compose a system, (2) ‘system features,’ and (3) ‘causally emergent system features’ that can be deduced (or predicted) neither from an analysis of the sheer physical structure of any given element nor from their environmental relations (Searle, 1992, p. 112). Let us suppose, for example, that we have a neural system, NS, composed of neurons, a, b, c, etc. NS has features that are not the features of a, b, c, etc. NS weighs four pounds, but each neuron does not weigh four pounds. The weight of NS is a ‘system feature’ (Searle, 1992, p. 112). Even so, the weight of NS is a system feature that can be deduced from the weight of its individual neurons. However, there are some features of NS that cannot be deduced from the features of its individual neurons (Searle, 1992, pp. 111–112). Searle calls such features ‘causally emergent system features’ (p. 112). One such feature is ‘transparency’ (p. 111). Searle thinks that consciousness is more like transparency than weight. Transparency itself can be deduced neither from the sheer physical structure (e.g., weight and shape) of its underlying hydrogen and oxygen molecules nor from its environmental relations. That is because an explanation of transparency must also take into account the causal interactions among its underlying molecules
Similarly, consciousness itself cannot be deduced ‘from the sheer physical structure of the neurons without some additional account of the causal relations between them’ (p. 112). Consciousness emerges only if there are sufficiently organized neurons present to allow for suitable causal relations. But, on Searle’s model, the causal relations of neurons that underlie the emergence of consciousness can also account for the causal efficacy of consciousness: if the causal relations of neurons (CRN) bring about the emergence of consciousness (EC), then CRN could also explain how EC affects the physical world. Causation is, therefore, asymmetrical in the sense that the causal powers of physical base properties, predicated on the basis of inter-neuronal relations, are sufficient to account for the causal powers of consciousness. Searle calls this weaker conception of causal emergence ‘emergent 1’ (p. 112). This notion of emergence is distinct from a stronger form of causal emergence known as ‘emergent 2,’ which states that consciousness could cause events, including neural ones, that could not be fully ‘explained by the causal behavior of neurons’ (p. 112). Because Searle holds to emergent 1 properties only, it is not surprising that he defends emergentism merely on grounds of the non-deducible character of consciousness, rather than also appealing to ‘downwards causation’ in the sense described and defended by those who hold to a stronger notion of emergentism (e.g., Hasker, 1999; O’Connor, 1994; Varela & Thompson, 2003; see also Chalmers, 2006; Davies, 2006). We will look at the idea of ‘downwards causation’ in a moment, but only after we have provided a brief criticism of Searle’s weaker notion of emergence.

First of all, Searle’s weaker variety of emergentism is predicated upon the assumption that if CRN is the cause of EC, then CRN explains the causal power of EC. As Hasker (1999) observes, this would suggest that ‘there is no further requirement on explanation beyond what is already implied in causation’ (p. 173). Most, if not all, advocates of emergentism would not deny that consciousness has a biological basis, but this does not entail that the causal powers of consciousness are (or could be) wholly explicable in biological or physicalistic terms. Second, Searle’s notion of causation and explanation (as it applies in the case of consciousness) would be true only if science were in a state of completion. But science is not in a state of completion (Davies, 2006). So, Searle cannot mean that we currently possess a sufficient account of consciousness, and thus its respective causal powers, in terms of its underlying microphysical causal bases (see also Hasker, 1999). Finally, Searle’s (1992) ‘biological naturalist’ approach to the emergence of consciousness does not seem all that biological after all: for it eliminates the need to take into account the role that environmental relations could (and probably did) play in the evolutionary emergence of consciousness by supposing that causal interactions among microphysical elements are sufficient to account for the emergence and causal powers of consciousness. Despite his claims about the biological nature of consciousness, Searle does not appear to emphasize
What is consciousness for?’ in his account. This is an odd consequence for a view whose title is ‘biological naturalism.’

**Strong Emergence: Downwards Causation**

In contrast to Searle (1992), I agree with Varela and Thompson (2003) and Hasker (1999) in insisting that we should widen the criteria of emergence to include ‘downwards causation’ (or, more precisely, *downwards control*), but that we should do so within an empirically informed framework. On Varela and Thompson’s construal of emergentism, the relation between consciousness and the brain holds a relation of reciprocal causality, from ‘upwards causation’ to ‘downwards causation.’ Upwards causation leads to the production of novel features that have their own transient careers. Downwards causation refers to holistic emergent processes that control or constrain lower-level interactions (Varela & Thompson, 2003, p. 273). Varela and Thompson are not alone in describing downwards causation in terms of controlling, organizing, ordering, or modifying. Kelso (1995) suggests that consciousness ‘molds the metastable dynamic patterns of the brain’ (p. 288). And Freeman (2000) proposes that consciousness is a higher-level principle of organization ‘that constrains the chaotic activities of the parts [or subsystems of the brain] by quenching local fluctuations’ (p. 135). These reflections on the causal power of consciousness are compatible with the idea that the cognitive subject is not merely an *efficient cause* but also a *formal cause* (i.e., an organizational cause), a distinction that goes all the way back to the genius of Aristotle (1941: *Physics*, II, 3; see also LaRock, 2002).

Having defined the sense in which the term ‘emergence’ is used in this context, one wonders whether there is further empirical evidence in support of the claim that higher-level activities of the cognitive subject are causally efficacious in relation to lower-level neural activity. We might consider, for example, how the cognitive subject’s interpretations of ambiguous figures supply implicit evidence for a stronger variety of emergentism. A well-known ambiguous figure is the Necker cube. While viewing a Necker cube, many subjects consistently experience the perception of the cube flipping back and forth between alternative orientations. To elaborate how this occurs, one could utilize the experimental research of Kelso and colleagues (1995; see also Varela & Thompson, 2003). Kelso and colleagues devised an experiment in which the persons were asked to view a Necker cube in eight different spatial orientations. These spatial orientations were presented randomly to the persons involved in the experiment. They were instructed to push a button each time they detected a change in the cube’s spatial orientation. For the most part, there was no regular pattern exhibited by the person’s response to observed switching times for each spatial orientation. However, as the orientation of the cube approached that of a square (oriented at 80 degrees), the spatial orientation...
was perceived for a longer period of time without switching. In light of these results, one could infer that different interpretations of the Necker cube initiated by the cognitive subject will influence slightly the neuronal bias that underlies the cube reversal (see Varela & Thompson, 2003, pp. 277–278). This suggests that ambiguous figures are experienced differently depending on how they are interpreted by the cognitive subject. Hence, higher-level interpretations carried out by the cognitive subject act as top-down influences on the lower-level neural assemblies responsible for generating differing orientations. The cognitive subject, in turn, experiences the result of these higher-level influences on lower-level processes.

IHS is compatible with this stronger notion of emergence in the sense that a unified percept (i.e., a feature-unified object of visual consciousness) emerges out of the interaction of information organized by spatial structuring capacities that correlate with lower, higher, and intermediate levels of the visual hierarchy. Before looking at the evidence for such spatial structuring capacities, it might also be useful to point out the sense in which the IHS approach differs from other approaches, such as the usual varieties of philosophical functionalism and neural reductionism. These usual sorts of philosophical functionalism say that when we talk about mind, we are actually referring to a set of mental states defined in terms of causal roles between perceptual inputs, internal mental processes, and behavioral outputs. For example, my belief that a hurricane is about to form is caused in me by my perception of weather patterns consistent with hurricanes; and in relation to my desire to preserve my life, the fear of a hurricane will cause me to run for shelter. Parting company with the type-type identity theory in the philosophy of mind (Place, 1956; Smart, 1959), functionalists of this usual sort do not hold that mental states can be identified solely with physical states of the human nervous system, but instead contend that mental states can be realized in any suitably organized system. For example, systems composed of silicon, living cells, or perhaps even ‘spiritual energy,’ if properly organized, could be sufficient to instantiate mental states, so long as the right causal roles took place (see Fodor, 1981). Thus, the functionalist can accept conceptual distinctions between mental and neural terms without also embracing ontological distinctions. If the concept of pain is logically equivalent to the concept of a state that ‘occupies a certain causal role,’ then whatever state occupies that role is, in fact, pain (Lewis, 2000, p. 112). Although the concept of pain is not the concept of a neural state, it is nonetheless applicable to the relevant causal role in some system or other, an endorsement of the mind’s multiple realizable character (Fodor, 1981; Lewis, 2000).

I have argued elsewhere that one of the principal errors implicit in the usual varieties of philosophical functionalism is that they downgrade the significance of the biological underpinnings of cognition and experience and therefore fail to appreciate findings in brain science (see Kafetsios & LaRock, 2005; LaRock, 2001, 2002). At the opposite end of the spectrum, Prinz (2000) points out that some neuroscientists who are sympathetic to the claims of
neural reductionism talk as though it is enough to correlate conscious states with neural states without ever discussing function. This kind of talk implicitly assumes that correlation entails identity. But correlation no more entails identity than the property of having three sides entails the property of having three angles. Not all three-sided objects are objects with three angles. Thus, even though identity entails correlation, correlation does not entail identity (LaRock, 2002; Swinburne, 1997). Moreover, in the preceding section I provided an empirical case against the claim that certain neural mechanisms fully explain a feature-unified object of visual consciousness. Thus, providing a case against neural reductionism (and, for that matter, eliminative materialism) is no longer limited to conceptual analyses of mental and neural terms.

If the usual varieties of philosophical functionalism and neural reductionism are too narrow, due to their respective methodological assumptions, this might suggest that a broader approach should be adopted whereby reciprocal influences between psychology and neuroscience take place for the sake of refining both disciplines (see Kafetsios & LaRock, 2005; LaRock, 2002; see also van Eck, Looren de Jong, & Schouten, 2006). Similarly, IHS implicitly seeks to weaken the battle lines between neuroscience and psychology by accommodating both disciplines within its theoretical framework. Having understood the sense in which IHS is compatible with emergentism, and yet differs from philosophical functionalism and neural reductionism, we are now prepared to consider the evidence for spatial structuring capacities at lower, higher, and intermediate levels of the visual hierarchy.

Lower-Level Spatial Structuring

As we observed earlier in this paper, there is considerable evidence indicating that V1 is probably not the locale (or neural correlate) of visual consciousness (see Crick & Koch, 1995; Gray, 1999; Prinz, 2000). However, this does not mean that the spatial structuring activities correlated with V1 are irrelevant to contributing to the spatial structure of objects in visual consciousness. For one thing, the relation between the spatial properties of retinal images and the spatial properties laid out on the surface area of V1 is a relation of structural coherence (see LaRock, 2002). Some of the empirical evidence that can be marshaled in support of structural coherence is derived from experiments performed on monkeys with visual anatomy similar to humans. In one of the experimental set-ups, scientists would train a monkey to stare at a bull’s eye pattern. While staring at the bull’s eye pattern, the monkey is injected with a radioactive sugar called 2Deoxyglucose. The monkey is then sacrificed. Shortly thereafter, a slice of the monkey’s primary visual cortex is removed and chemically developed much like a photograph. Upon chemical development the effects of the radioactive tracer create an outline of the impression caused by the stimulus object, that is, the bull’s eye. The
radioactive tracer identifies the neural regions most actively involved while perceiving the bull’s eye (Kosslyn & Koenig, 1995, pp. 67–68). Kosslyn and Koenig (1995) summarize the experimental results as follows:

As expected, an image of the pattern was spread out over the back of the animal’s brain, particularly in area V1. ... This visual area, V1, is retinotopically organized: The pattern that falls on the retina is itself sent into the brain, and this pattern is physically laid out on the surface area of V1. The area is organized in the same way as the retina, and hence preserves the spatial properties of images that fall on the retina. If you held up your hand during the experiment, and kept the animal focused on it throughout, presumably a ‘picture’ of your hand would have appeared on its brain. (pp. 67–68)

This evidence is supportive of the view that there is structural coherence between the spatial properties of retinal images and the spatial properties displayed in V1. This would suggest that the visual subject is indirectly related to object structures in the external world by means of the spatial properties embedded in and represented by the retinotopically organized V1.

**Higher-Level Spatial Structuring**

How might spatial structuring function at higher levels? How do higher levels influence lower levels? Recent evidence suggests that structural information correlated with IT cortex contributes to object perception in the sense that what is stored in (or correlated with) IT are 3-D structures whose *top-down interaction* with incoming information from striate cortex assists in establishing shape assignment (Parker, Cumming, Johnston, & Hurlbert, 1995; cf. Marr, 1982). In fact, evidence gathered from neurophysiological studies indicates that stored structural information correlated with IT can be activated approximately 60 ms after the onset of a stimulus, which ever so slightly precedes the time in which the receptive fields of neurons in V2 respond to shaped figures rather than their respective grounds (Peterson, 2001). Thus, the initial retinotopically organized V1 embodies information that operates as salient cues (e.g., shape from shading), causing feedback from 3D structures stored in IT (Braun, 1993; Ramachandran, 1988; Sun & Perona, 1997). The visual information that *interacts* between V1 and IT is organized by these spatial structuring capacities in order to assist in achieving shape assignment. I should also mention that a higher-level cognitive function correlated with the fusiform gyrus, an area of the temporal lobe, is an area often correlated with face recognition. For example, if persons undergo damage in this region, they have severe difficulty recognizing faces. The *inability* to recognize face structure is called *prosopagnosia*, derived from the Greek words πρόσωπον (or face) and αγνωσία (non-knowledge). As it happens, recent empirical data indicate that the fusiform gyrus also correlates with our ability to establish the structure of shapes (McCarthy, Puce, Gore, & Allison, 2004). The representation
of the structural character of objects would, therefore, most likely require higher-level analysis by the cognitive subject, and this would involve access to information about the nature and structure of previously experienced objects and scenes (cf. Shadlen & Movshon, 1999).

Based upon the above data, one might surmise that object memories in the form of 3D structures fully explain object feature binding in that they detect the internal spatial structure of visual objects. Although this might work in the case of previously experienced objects, we can also represent objects that we have never seen before. So, the appeal to object memories per se cannot be a sufficient explanation of object feature binding. It looks as though, in many cases, we have to be able to represent object feature binding without access to previously encoded object memories. Thus, although IT is necessary for the representation of unified objects, it is not also sufficient. In fact, neuropsychological evidence has shown that persons with associative agnosia disorder have undergone damage to cells in IT and, as a result, have severe difficulty recognizing the shapes of objects they consciously perceive (Farah, 1990; Kosslyn & Koenig, 1995). It is possible, therefore, to have visual consciousness of objects even if one loses the ability to recognize such objects. It would be a mistake, then, to conflate visual consciousness and visual recognition. Moreover, the size of the receptive fields of neurons in IT is comparatively large and, as a consequence, these cells are indifferent to the size, spatial position, and orientation of objects in one’s visual field (Luck & Beach, 1998). The implication is that recognition tasks, correlated with activity in IT, abstract away from the vantage point of the perceiver. Nevertheless, the mechanism(s) that underlies visual consciousness of objects must account for the vantage point of the perceiver (presumably because visual consciousness of objects is always tied to a certain vantage point). Neural activity in IT cannot account for the vantage point of the perceiver. It follows that neural activity in IT cannot be the sole mechanism that underlies visual consciousness of objects. The findings discussed so far might suggest that visual awareness is somehow linked (or arises in relation) to intermediate-level processing, which is compatible with an interactivist conception of information processing involving lower and higher levels.

**Intermediate-Level Spatial Structuring**

Thus far, we have seen how spatial structuring capacities organize visual information in an interactive manner between lower and higher levels of the visual cortical hierarchy. A consideration of certain pathologies of consciousness substantiates the idea that a spatial structuring capacity is operative at the intermediate level and is integral to the unity of an object’s features at the level of visual consciousness. First, apperceptive agnosics with damage to the inferior parietal lobes exhibit deficits in representing the intrinsic spatial and structural relations of objects, especially when the perspective is novel or the
lighting conditions are uneven (Farah, 1990). Having difficulty representing structural wholes, these patients ‘describe the world as chaotic or fragmented’ (Kosslyn & Koenig, 1995, p. 112). Second, experiments on illusory conjunctions indicate that the color feature of a large red circle can be selected and then falsely conjoined to coincide with the shape of a small blue triangle, or vice versa (Treisman, 2003; Treisman & Schmidt, 1982). This would suggest that spatial structure provides the necessary binding principles: the person is aware of the triangle as red because the redness is bound to the same spatial areas as the triangle. Third, persons suffering from Balint’s syndrome with bilateral damage to the parietal lobes have severe problems with binding because they have lost the ability to establish the spatial structure in which to individuate and bind features together. For example, after extensive studies on the Balint’s patient RM, Robertson, Treisman, Friedman-Hill, and Grabowecky (1997) concluded that ‘he has no perceptual space in which to separate and individuate the letters and bind the colors to the shapes’ (p. 307). These data provide further support to the idea that our ability to bind color to shape relies, in part, upon a spatial structuring capacity at the intermediate level.

**Testing IHS: Feedforward and Feedback Interactions**

Since IHS implies that visual consciousness depends upon feedforward to feedback interactions of information from lower areas to higher areas, it might be useful to discuss how various tests have been performed to support this type of claim. In this context, I will focus briefly on tests involving transcranial magnetic stimulation (TMS) and backward masking. TMS enables an experimenter to deploy a magnetic pulse in order to disrupt activity in a specialized neural area of a normal, healthy participant for a short period of time (Walsh & Cowey, 1998). TMS studies indicate that motion stimuli encoded in MT can be perceived unless V1 is disrupted at some later moment in the information processing stream. Because neural activity in MT is higher than V1, this would suggest that ‘feedback from MT to V1 is necessary for motion awareness’ (Lamme, 2003, p. 15). Moreover, masking experiments can be deployed in order to distinguish unconscious and conscious activity. This type of test seeks to confirm or disconfirm the notion that unconscious activity is correlated with the initial feedforward sweep of information and that conscious activity arises only after recurrent interaction occurs. This test requires presenting a participant with two visual stimuli approximately 40 ms apart from each other. By presenting a second stimulus within this brief time period, the first stimulus is no longer visible to the participant. Even though the second stimulus masks the first, and is therefore not consciously visible to the participant, it still produces feedforward activity in several visual areas. On the other hand, the neural activity that correlates with information feedback (or recurrent interactions) is significantly suppressed by such masking tests (Lamme, 2003; Lamme & Roelfsema, 2000).
Further Advantages of IHS

In addition to presenting a testable account of unified visual consciousness, IHS has further philosophical and scientific advantages. For example, unlike traditional Cartesian varieties of dualism which suppose that the conscious mind is an unextended thinking substance that must be superadded to the brain ‘ab extra by a special divine act of creation’ (Hasker, 1999, p. 189), IHS holds that consciousness emerges in connection with information implemented upon suitably organized biological systems. Rather than being excluded from biology and neuroscience, the emergentist conception of consciousness defended here is indispensable for a fuller explanation of human evolution and conscious behavior (Eccles, 1994; Ellis, 2006; Hasker, 1999; Sperry, 1991). In other words, IHS coheres with findings in nature: humans and animals bear conscious continuity with each other because of their evolutionary past. If consciousness is a product of emergent evolution, there would be a close link between our biology and psychology; and this would be true even if the conscious mind, qua emergent subject, were ontologically distinct from the brain. Contrary to Descartes, consciousness is not merely a special property essential to human beings, but is better characterized as a continuum in living nature. Just how far consciousness extends down the phylogenetic tree is another issue—whether worms and gnats possess it is debatable (cf. Nagel, 1974). In summary, nature testifies that animals have consciousness. Their conscious experiences are probably less complex than ours because ‘they are generated by less complex nervous systems’ (Hasker, 1999, p. 193). Another chief advantage of IHS is that it attributes some causal powers to the conscious mind itself. These causal powers cannot be explained away by the preemptive powers of physical base properties, as they could be in the case of the supervenience theory (Kim, 2000; see also LaRock, 2001, 2002).

Furthermore, IHS has important explanatory advantages over some of its contemporary dualist competitors. For example, Chalmers’ (1995, 1996) naturalistic dualism (ND) implies that phenomenal consciousness accompanies all information processes. Because information is taken to be a fundamental feature of the physical universe, consciousness is taken to be a fundamental feature, much like mass, space, and time are fundamental features of the universe (Chalmers, 1995; cf. Wheeler, 1990). Chalmers (1995) recognizes, however, that if phenomenal experience accompanies all information processes, then ND seems to be underdetermined. For example, perhaps a thermostat, in virtue of embedding information states about temperature, has the experience that it feels just right at certain times during the day. Unlike ND, the emergentist view defended here implies that suitably organized biological systems could count as a further constraint on a theory of consciousness. Consequently, the problem of underdeterminism is not implied by this emergentist account. In fairness to Chalmers, however, some views of emergentism appear as if they cannot provide a solution to the hard problem of consciousness (i.e., the
problem of why experience should arise from matter in the first place; see also Locke, 1689/1979). ND purports to solve this problem by taking consciousness to be a fundamental feature of the universe. Without going into any details here, I have argued elsewhere that an Aristotelian conception of form—the fundamental organizing principle of matter in Aristotle’s metaphysical model—could bridge the gap between matter and consciousness, in the sense that forms could extend all the way down to the basic particles of the universe without entailing panexperientialism (see LaRock, 2002; see also, Lowe, 1998).

Finally, because IHS takes subjectivity seriously as part of an overall theory of consciousness, it can account for other properties of visual consciousness, such as the diachronic (or persistence) aspect of the subject of visual consciousness. Recent data suggest that the cognitive subject bears a persistence relation to the processing stream of temporally segregated feature representations. The ability to exemplify this kind of persistence relation appears to be a mode of the visual subject, and this could very well shed light on how diachronic object unity is possible. For example, experiments conducted by Zeki (2003) have shown that the perception of color occurs prior to the perception of motion by approximately 80 ms, and that location is perceived prior to color and orientations. An implication thus far is that location awareness appears to be temporally foundational to the perceptual processing stream. Moreover, these data reveal an asynchrony of the neuronal events correlated with the individual feature representations of an object. Thus, not only are the neuronal events that correlate with the distinctive features of an object distributed in space (i.e., correlated with separate areas of the visual cortex), but they are ‘distributed in time as well’ (Zeki, 2003, p. 215). As a result of these findings, Zeki (2003) has drawn a useful distinction between ‘micro-consciousness’ and ‘macro-consciousness’ (pp. 216–217). Micro-consciousness refers to the preliminary awareness of single object features at separate times. Macro-consciousness refers to the awareness of two or more object features over time. Now, it would be hasty to infer that an account of a feature-unified object of visual consciousness can be achieved by means of macro-consciousness, since macro-consciousness only refers to the awareness of two or more separate features over time. By implication, macro-consciousness increases conscious representational scope (i.e., the visual subject’s ability to be aware of more than one separate object feature over time), but this does not entail an account of a feature-unified object. Becoming aware of temporally segregated feature representations of an object over time is not the same as being aware of the bound percept. As Zeki explains, ‘subjects become conscious of the bound percept only after they become conscious of the attributes that are bound, again suggesting a temporal hierarchy in perception’ (p. 216). Evidently, the subject of visual awareness bears a persistence relation to the temporally distributed feature representations correlated with separate areas of the visual cortex. These findings are compatible with IHS: our ability to be aware of an
object’s *spatial location* could play an integral role in the perception of a bound object, in the sense that becoming aware of succeeding feature representations would co-refer to this earlier form of spatial/locational awareness. This idea is developed further in the succeeding section, where I respond to a potential binding problem for IHS.

**A Binding Problem for IHS?**

A residual issue remains, however. IHS might imply a binding problem of its own. Consider, for example, an object consisting of a triangle on top of a square, and an object consisting of a square on top of a triangle. In both objects, we have the shapes square and triangle, and the relations *on top of* and *below*. But the binding between the relations and the shapes is different for these objects. So, IHS can work only if this location type of binding problem can be solved. In response to this problem, one need only mention that just as V1 and IT form an information feedback loop, so too V1 operates in tandem with parietal cortex through a feedback loop to process spatial information, such as the *size* and locational properties of objects (see de Kamps & van der Velde, 2001; Treisman, 1996, pp. 174–175; 2003). Thus, *locational* information about both objects would be initially processed in V1 and this information, in turn, would become input for further processing in parietal cortex, ultimately leading to subsequent interactions between lower (V1) and higher (parietal) levels of the processing hierarchy. In fact, PET scans have shown that occipital, temporal, and parietal areas are active during object perception. For example, in one experiment, participants were shown complex colored geometric shapes and, as a consequence, all three areas exhibited heightened neural activity (Roland & Gulyas, 1995; Treisman, 1996). It is worth emphasizing that these findings are compatible with IHS: our ability to be cognitively aware of an object’s *spatial location* could play an integral role in the perception of a bound object, in the sense that becoming aware of succeeding feature representations would co-refer to this earlier form of spatial/locational awareness. Presumably, information about the location of an object is fed forward from V1 to specialized areas of the parietal cortex for deeper processing: information feedback from parietal cortex would then occur so that the cognitive subject could maintain the locale of an object. The cognitive subject’s ability to maintain the locale of an object *over time* could play a role in individuating the space in which an object’s color and shape are subsequently bound together.

**Conclusion**

I argued that although neural synchrony and/or attentional mechanisms might function to disambiguate an object’s features, it is difficult to see how either
of these mechanisms could fully explain the unity of an object’s features at the level of visual consciousness. A philosophical upshot is that providing a case against neural reductionism is no longer limited to conceptual analyses of mental and neural terms. As knowledge about the brain increases, evidence grows against neural reductionism; and the light of this evidence casts shadows over the promissory notes of eliminative materialism. A new view of visual binding was presented called *interactive hierarchical structuralism*. This view suggests that a unified percept (i.e., a feature-unified object of visual consciousness) is not reducible to the activity of any cognitive capacity or to any localized neural area, but *emerges* out of the interaction of visual information organized by spatial structuring capacities correlated with lower, higher, and intermediate levels of the visual hierarchy. Unlike neural reductionism and the usual varieties of philosophical functionalism, IHS incorporates within its explanatory framework both personal (i.e., the phenomenological/first-person experience of unity) and subpersonal (i.e., neurobiological/third-person) aspects of our ontology (cf. Hurley, 1998). Hence, IHS is able to embrace both mental realism and scientific realism at the same time. In addition to its explanatory power, the theory of IHS has broader methodological implications. It points toward a methodology that makes room for reciprocal influences between psychology and neurosciences for the sake of refining both disciplines (see Kafetsios & LaRock, 2005; LaRock, 2002; see also van Eck et al., 2006; Varela & Thompson, 2003).

**References**


**ACKNOWLEDGEMENTS.** I am especially thankful to three reviewers for *Theory & Psychology* and Associate Editor, Huib Looren de Jong, for helpful comments. Parts of this manuscript were presented at the Southern Society for
Philosophy and Psychology (April 2006); ‘Toward a Science of Consciousness,’ the seventh biennial Tucson conference on consciousness, sponsored by the Center for Consciousness Studies at the University of Arizona (April 2006); and ‘Neurophilosophy: The State of the Art,’ held at the California Institute of Technology (June 2005). I thank John Bickle, David Chalmers, Uriah Kriegel, Anthony Landreth, Jesse Prinz, Robert van Gulick, and other participants at these conferences for their helpful suggestions.

ERIC LAROCK is an Assistant Professor in the Department of Philosophy at Oakland University. His research interests include philosophy of mind, consciousness and neuroscience, ancient Greek philosophy (especially Aristotle’s psychology), and metaphysics. Some of his recent publications include: ‘Cognition and Emotion: Aristotelian Affinities with Contemporary Emotion Research’ (with Konstantinos Kafetsios), Theory & Psychology, 15 (2005); ‘Against the Functionalist Reading of Aristotle’s Philosophy of Perception and Emotion,’ International Philosophical Quarterly, 42 (2002); ‘Dualistic Interaction, Neural Dependence, and Aquinas’s Composite View,’ Philosophia Christi, 3 (2001). ADDRESS: Department of Philosophy, Oakland University, 2200 N. Squirrel Rd, Rochester, MI 48309, USA. [email: larock@oakland.edu]