Understanding Consciousness—Have We Cut the Gordian Knot or Not? (Integration, Unity, and the Self)

Robert Van Gulick
Syracuse University

In his 1992 book *Consciousness Reconsidered,* the philosopher Owen Flanagan noted that there are three great mysteries: Why does the universe exist? What is life? And what is consciousness? What is the status of the three mysteries today in the early twenty first century? Have these mysteries been solved? Or do they remain unexplained? And if any remain unsolved today, what are our prospects for doing so in the near or distant future?

The first mystery concerns the very fact of existence: Why does the universe exist? Why is there something rather nothing? Contemporary physicists can tell us a lot about the origin of the universe. According to our best understanding of the data, the universe began in the so called 'big bang" approximately 13.8 billion years ago. Physicists are able to determine the state of the universe back to the earliest nanoseconds of its existence. They may even be able to explain the occurrence of the big bang itself, perhaps in terms of cosmic inflation models, for which there is strong though not yet conclusive evidence. Those inflation models explain how local quantum fluctuations in the vacuum energy state might lead to exponentially expanding spacetime bubbles, i.e. to spacetimes that rapidly expand at an accelerating rate, before slowing to a nonaccelerating rate giving rise to a complete spacetime universe like our own. The models match and explain many features observed in the large scale structure of the universe, such as the uniformity of the background radiation.

Do such models resolve the mystery of why there is something rather than nothing or do they merely push the question back a step? Following an ancient philosophical doctrine going back to the Greek Parmenides and the Roman Lucretius, Saint Thomas Aquinas famously affirmed the principle, "Ex nihilo, nihil fit"—"From nothing comes nothing"—or as we might put in more contemporary terms, "You can't get something from nothing." But perhaps we can. According to the inflation models expanding spacetime bubbles can arise from the vacuum or zero energy state. From the perspective of quantum mechanics even the zero energy

state has quantum fluctuations that could lead to inflation. Does the fact that such exponentially expanding bubbles could result from the zero energy state show that you can indeed get *something from nothing*? Or does it merely show that from the metaphysical perspective, the zero energy state, though zero, is still a *something rather than a nothing*, though if it were a *something* it would be a very strange something, one that stands outside of space and time.

What is the status of the second mystery: What is life? If we were to go back hundred years ago to 1916 in the early twentieth century, we would find some genuine disagreement within the scientific and philosophical community about whether or not life could be understood in terms of the same factors and laws that operated in the nonbiological world. Some argued that explaining life and the processes within living organisms required appeal to factors over and above those involved in physics and chemistry—such as vital forces, so called "elan vital" or special organizing forces or "entelechies." These were posited to be independent additional causal aspects of reality beyond those present in the nonbiological realm. Appeal to these factors was regarded as essential, and thus any attempt to explain life solely in terms of underlying chemical or molecular structure was doomed to fail. In 1913 the eminent British physiologist, J.B.S. Haldane rejected the mechanistic theory of life in favor of organicism. He wrote, "We perceive the organism as a self-regulating entity,…. every effort to analyze it into components that can be reduced to a mechanical explanation violates this central experience" (Haldane 1913).

However, vitalism and anti-mechanism ceased to be serious hypotheses with the revolution in genetics and molecular biology that began with the discovery in 1954 of the structure of DNA by Francis Crick and John Watson and soon thereafter of its basic operation in processes of coding, transcription and replication. Admittedly there is a great deal that we are still learning about the relevant genetic and molecular processes, such as the recent recognition of the important role played by epigenetic factors, molecules that can attach to genes and modulate their activity, sometime as the result environmental interactions. Epigenetics helps explain how gene activity gets shaped by one's individual history and environment. Our knowledge of the specifics is far from complete. We certainly cannot at present tell the full story that gets you from a fertilized moose zygote to an adult bull moose, but we have a sense that we know in general "how the trick is done." We can fit living organisms into reality as complex, self-regulating and self-replicat-

ing biochemical systems that have evolved through Darwinian selection. The problem of explaining life no longer seems like a mystery, but just another difficult but solvable problem for ongoing scientific research to fill in the details.

The third mystery—the mystery of consciousness, is the most specific of the three. Even if we were to understand why there is something (a universe), and how living things exist, we might still be puzzled about how any of those things – living or not – is conscious? It seems we can imagine a possible world with physics, chemistry and even biology, but nothing that is conscious. Atoms, oceans and even living things might exist in the total absence of consciousness, e.g. a world of rivers, mountains bacteria and even plants but devoid of any conscious mind to observe it. The third mystery is perhaps also the one that appears most elusive. Can we understand what consciousness is, and how it fits into the rest of reality? In particular, can we explain consciousness scientifically in terms of underlying physical, chemical or neural structures and processes? Or to frame the issue more generally: Can we explain the nature and existence of consciousness in terms of components that are not conscious? In current debates, some say "yes", but others say "no."

The problem of consciousness has a long history. In the nineteenth century, the English biologist and ardent defender of Darwinian theory, Thomas Huxley famously wrote, "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just *as unaccountable as the appearance of the Djinn, when Aladdin rubbed his lamp*"(Huxley 1866). In mid twentieth century, Ludwig Wittgenstein expressed the sense of bafflement we supposedly feel when considering consciousness. He imagines someone clasping his forehead and exclaiming, 'This is supposed to be produced in the brain!." He goes on to describe it as "the feeling of an unbridgeable gulf between consciousness and brain process… The idea of a difference in kind is accompanied by a slight giddiness – which occurs when we performing a piece of logical sleight of hand" (Wittgenstein *Philosophical Investigations* I, 412, 1953). Wittgenstein himself was suspicious of such intuitions, but he leaves no doubt that they were widely and powerfully held.

What is the status of our third mystery today in the early twenty first century? To what extent have we succeeded in explaining the nature and existence of consciousness, and how it fits into the physical world? In terms of our title, have we succeeded in "cutting

the Gordian knot"? In ancient history or myth, Alexander the Great, confronted a knot in the ancient kingdom of Gordia that bound a ritual cart as a symbol of kingship to the wall of the royal palace, a knot so strong and intricate that Alexander could find no ends from which he might begin to untie it. He decisively solved the problem by slicing the knot with his sword and freeing the ends to undo it. "Cutting the Gordian knot" has thus become a metaphor for solving an seemingly intractable problem by bold and novel action that opens new avenue to solution, an ancient version of "thinking outside the box." Have philosophers or scientists found a way to cut the Gordian knot of consciousness, a way to gain the new perspective needed to resolve our third mystery? Current opinion differs on the answer.

In his seminal 1974 paper, "What is it like to be a bat?", Thomas Nagel argued that the subjective nature of consciousness places severe, and perhaps insuperable, limits on our ability to understand consciousness from the objective perspective of natural science. Nagel uses the example of bats that perceive and navigate through echo-location to illustrate his point about the special sort of understanding associated with conscious experience. Using Nagel's famous phrase, "there is something that it is like to be a bat", i.e. a qualitative phenomenal aspect of what bat experience is like for the bat itself, how it experiences the world from its particular sensory perspective. Nagel argues that no amount of knowledge of the bat's neurophysiology or of the organization of its information processing will enable us to fully understand "what it is like for the bat" from the inside. Our sensory experience is too unlike the bat's for us to be able to comprehend bat's experience in the sympathetic way in which we can understand human experiences. We cannot adequately take up the bat's experiential perspective since our own is too dissimilar.

Nagel goes on to argue that any reductive program that aims to fully explain conscious experience in terms of objective science will fail to capture the distinctive what-it-is-likeness of subjective experience. Indeed, he claims we have no real understanding of how the two might fit together; they seem so dissimilar that trying to link them leaves us blank. Our current position, he suggests, is akin to that of ancient pre-Socratic philosophers trying to understand the contemporary equation of matter and energy. We lack the basic conceptual framework to begin to see how the two might be linked.

Colin McGinn also argues that we unable to understand the physical basis of consciousness in an intelligible way, though his reasons are different than Nagel's. According to McGinn (1989), the residual mystery results not from any metaphysical facts but from facts about our human cognitive limits, especially limits on our capacities for forming concepts. McGinn argues that we humans are unable to adequately grasp how "the water of brain is turned into the wine of consciousness" because we are unable use to either perception, introspection or inference to acquire the concepts needed to make the psycho-physical nexus intuitively transparent. Though other beings with different cognitive capacities may be able grasp the link, for us humans it must remain a mystery.

In considering how the mystery of consciousness fares relative to the other two, it is worth noting that all three mysteries have a *similar form*. Where *F* is some property, each mystery is of the form: "How can you get something *that is F* from things *that are not F* ?" The first mystery #1 is: "How can you get something that *exists* from what does *not exist*? The second mystery #2 is: "How can you get something that is *alive* from things (parts) that are *not alive*? And mystery #3 is: "How can get something that is *conscious* from things (parts) that are *not conscious*?

Despite their similar form, our ability to answer the question and resolve the mystery may be quite different in the three cases, and even those who believe we can answer the second and perhaps the first, may be more pessimistic about our prospects with respect to the third. One hundred years ago in 1916, life was still a mystery; today it is not. Will consciousness be explained in the 21st century as life was explained in the 20th century? Will we soon be able to explain how consciousness depends on the brain just as we can explain how life depends on DNA and the rest of our molecular biochemistry? Are the two problems basically similar or different in some important way?

Many philosophers and neuroscientists believe it is just a matter of time–and not too much time–until the mystery of consciousness is solved and non-physical views are as 'dead' as vitalism is today. Among philosophers for example, John Searle (1997) argues for a biologically naturalistic view of consciousness, Paricia Churchland (1994) appeals to neural network models, and Jesse Prinz (2012) offers his Attended Intermediate Representation (AIR) theory of consciousness which includes specific claims about how conscious-

ness in neurally realized by gamma oscillations in the relevant mid-level sensory neurons.

By contrast, other philosophers argue that the mystery of consciousness is fundamentally different from the problem of life and the method that worked for explaining life cannot be used for explaining consciousness. That mode of argument is offered most prominently by David Chalmers (1996), but is also supported by Galen Strawson (2006) who advocates panpsychism, and by dualists like Brie Gertler (2007). Chalmers famously distinguished between what he called the "Hard Problem" of Consciousness, and various "Easy Problems"(or at least far *easier* problems) of consciousness. The "easy problems" all deal with the functional aspects of consciousness (e.g., how it processes information, interacts with memory, or controls behavior) and with how those functional processes might be realized by neural mechanisms. These Chalmers says are tough ongoing scientific problems but not mysteries, and they are analogous to the problems that were solved by modern molecular biology in the last half century. Those biological problems also involved accurately modeling functional processes—such as reproduction, growth or metabolism—and explaining how those functional processes might be realized by underlying molecular mechanisms. By contrast, the "Hard Problem" is that of explaining how subjective phenomenal consciousness, what-it-is-likeness, comes about. How could qualitative phenomenal consciousness be the result of purely physical neural processes? It is this so called "Hard Problem" that generates an air of mystery and leaves us deeply perplexed. The resistance to solution derives in part from the supposed fact that being phenomenal conscious is not a functional property, or at least not in essence a *merely* functional property. Phenomenal consciousness may have some functional aspects and being phenomenal consciously may give one various functional capacities, but according to Chalmers the central feature of consciousness, very much like Nagel's 'what-it-is-likeness, is not captured by those functional properties. Thus the two-part strategy that worked for explaining life—specify the essential properties of life in terms of functional processes, and then explain how those processes are realized by underlying physical mechanisms—supposedly cannot be used to solve the mystery of consciousness. If as Chalmers argues, the thing to be explained—phenomenal consciousness itself—is not in essence a functional property or process, then the first part of the strategy cannot be carried out. Describing the func-

tional aspects of consciousness speaks only to the supposed easy problems, and fails to capture the essence of consciousness, leaving the strategy inadequate to address the Hard Problem.

However, one might question Chalmers views about the relation between the two sorts of problems. Chalmers assumes that making progress on the functional problems about the organization and operation of consciousness and their neural basis – the so called "easy problems"—will not be of much help in solving the "Hard Problem" because he believes they are fundamentally different in kind. But is that assumption correct? Perhaps as we build our theories and detailed models of the functional aspects of consciousness we will get the insight we need to solve the Hard Problem—or at least make progress on it.

The special divide that supposedly exists in the case of consciousness is sometimes described as that between the *objective* and s*ubjective* facts about conscious experience. Fact about 'what-it-is-likeness' are said to be subjective in that they can be fully known or understood only *empathetically* by one's being able to *imagine the relevant quality of experience* (or what-it-is-likeness)—i.e., to imagine having that type of experience oneself—which requires being able experience the world from a similar point-of-view. It is claimed that you cannot get a full or adequate understanding of subjective consciousness from any theory of objective facts of the sort you would get from neuroscience or any other empirical science.

Once again we can see this question as taking the same basic logical form as our other mysteries*: How* can you get something that is *subjective from* things (or parts) that are *not subjective* (that are *objective*). Indeed, it is possible that particular how-question has no answer. Thus the even more basic question is: *Can* you get something *subjective from* things (or parts) that are *objective*? Is it even possible to do so? It is worth noting in this regard that *subjectivity requires a subject*. There cannot be any experiential states that have a *subjective* aspect—some way in which they appear—unless there is some *subject to whom* they appear in that way. The notion of a pain that is not the pain of any subject is not really coherent. A state cannot be a pain with its subjective aspect of hurtfulness unless there is some subject who feels or experiences that pain, some subject to whom it appears as hurtful. *Subjective facts require a subject or self*. Thus in considering whether or not we can get subjective facts from objective facts, we need to address the question of whether and how one might construct a subject out of objective parts.

So where do things stand at present with regard to explaining consciousness? And what would explaining consciousness involve? What sorts of questions would we need to be able to answer to count as successfully explaining it? As I have described elsewhere (Van Gulick 2005), the various explanations required can be grouped under three main questions.

1. The What Question (descriptive): *What is consciousness?* What are its *principal features?* By what means can they be best *discovered, described* and *modeled?* Of special importance is the contrast between so called *first-person* and *third-person* methods. First-person methods might be described as those by which we learn about the mind *subjectively "from the inside"* through various forms of *introspection* and *self-awareness,* including both our *everyday self-awareness* and more highly developed forms of self-awareness made possible by phenomenological analysis or meditative practices. Third-person methods, by contrast, are often described as those in which we learn about mind *"from the outside."* Third-person methods for studying consciousness would include psychological lab studies, brain-imaging studies, and deficit studies in which diverse mental deficits are observed to correlate with damage to various specific brain regions. The consensus is that both sorts of methods are relevant and useful for studying consciousness, and multiple methods of both types should be used in complementary and mutually supporting ways. A successful explanation of consciousness should integrate our first-person and third-person data and theories.

2. The How Question (explanatory): *How* does consciousness of the relevant sort *come to exist?* Is it a *primitive* aspect of reality*?* If it is not primitive, then *how* does (or could) consciousness in the relevant respect *arise from* or *be caused by nonconscious* entities or processes*?*

3. The Why Question (functional): *Why does consciousness of the relevant sort exist? Does it have a function, and if so what it is it?* Does it *act causally,* and if so with *what sorts of effects?* Does it *make a difference* to the *operation of systems* in which it is *present,* and if so *why* and *how?*

Attempts to answer these questions empirically often make use of contrast studies, i.e. experimental setups that allow one to compare the differences between conscious and unconscious mental state and processing, e.g. the differences between conscious and

unconscious visual perception or between conscious and unconscious cases of memory and learning. The experiments aim to produce pairs of mental states—e.g. visual states—that are as much like each other as possible except that one is conscious and the other is not. It is possible to produce such conscious/ unconscious pairs by varying key stimulus parameters such as the exposure duration of a visual stimulus, or by manipulating other contextual factors such as the subject's focus of attention. Given sets of such contrasting states, it is possible to assess what differences consciousness makes to the nature of such states, including both functional differences in the kinds of processing involved and also neurological differences in the brain regions and types of neural activity involved respectively in the conscious and unconscious cases. The specific methods used in such contrast studies include: near and below threshold stimulation, backward masking, binocular rivalry (with attentional modulation), the "attentional blink" effect, and various sorts of deficit studies such those involving hemi-neglect or blindsight. For present purposes, we can consider two of those methods—backward masking and binocular rivalry—which will give a good sense of how such contrast studies work.

In a backward masking experiment a visual stimulus is presented to the subject for a very brief period of time, typically between forty and fifty milliseconds (ms), and then followed by another stimulus (the "mask") which is presented for a longer interval. Subjects typically do not report having seen the first stimulus at all. They report seeing only the second. Yet experimental tests can show that the first stimulus was visually processed to a high degree and recognized despite the subject lack of any conscious awareness of it. For example, the unconscious stimulus may bias the subject's subsequent response to conscious but unclear or ambiguous stimuli. If the masked stimulus of which the subjects remain unconscious showed an image of a tree, they are more likely to interpret the ambiguous word "palm" as referring to a palm tree rather than to a part of the hand. Thus the unconscious visual perception of the masked image would seem to have been of a fairly high level that involved a recognition of its meaning.

One can look at the subjects' patterns of brain activity data during a backward masking experiment by using by functional magnetic resonance imaging (fMRI). What one finds are both similarities and major differences between the cases of conscious and unconscious perception. Both the consciously seen and

unconscious masked stimuli evoke a succession of neural activetions across the full visual cortex, beginning in areas V1 and V2 where initial processing of the retinal output occurs, and continuing all the way up to later levels of processing in inferior temporal (IT) cortex that are involved in higher level processes such as object recognition and conceptual categorization. However, in the conscious case, activation also spreads to the frontal cortex and importantly the frontal cortex sends activation back to the visual cortex producing a sustained recurrent activation of the firing produced by the initial visual processing. The results thus suggest that such sustained recurrent activation involving both the visual and frontal cortex is required for conscious visual perception, though the correct interpretation of the data remains a matter of ongoing debate, with some arguing that more local recurrent activation solely within the visual cortex may by itself suffice to produce conscious perception without any need for the added involvement of frontal processes.

Binocular rivalry experiments involve project different images to each of a subject's two eyes—perhaps a man's face is projected to one eye and a city scene to the other. The subject perceptually processes both images but can be consciously aware of only one at a time. Subjects report that the images alternate spontaneously, back and forth (Logothetis 199x). First they are aware of the man's face, then of the city scene, and then it flips again to the face. The flipping back and forth enables researchers to isolate the neural signature of conscious processing. Because faces and scenes are known to be processed in distinct and identifiable brain regions—the fusiform face area (FFA) and the parahippocampal place area (PPA)—fMRI monitors can track what changes when the subject flips from being aware of the man's face to being aware of the cityscape. With binocular rivalry, both of the specialized processing regions stay active to some degree, reflecting the fact that both the conscious and the unconscious images are being processed to a high level of analysis. However, the cortical processing region associated with the currently conscious percept—whether the face region or the scene region—shows an increased level of activation and more importantly does so in part by receiving reciprocal activations from other cortical regions including the frontal cortex. Alternating conscious perceptions of a face and of a scene correlate with alternating increases of activation in the associated cortical regions, the FFA and PPA, driven in part by increased integration

with activity in other cortical regions, especially frontal and parietal regions associated with attention.

Contrast studies of these two sorts and others, can be used to try and discover the so called "neural correlates of consciousness" (NCCs) as an essential step in trying to understand how consciousness might result from physical or neural processes. If we know *what* physical brain states *correlate* with consciousness and what is different physically in the brain when we are in a conscious state, then we may be able to use that knowledge to figure out *how* such physical states might be able to produce consciousness. However, there are some important limits to keep mind in basing any conclusions about the physical basis of consciousness on data from contrast studies about the NCCs. Most importantly, as the term itself implies, NCCs are merely *correlates* of consciousness, states that co-occur with the cases of conscious perception. From the fact that are present in the conscious cases, it does not follow that those neural features are the basis or neural substrate of consciousness. Alternative hypotheses could also account for the observed correlations. Rather than being identical with consciousness, the NCCs might be the *causal precursors* of consciousness, or they might be *downstream effects* of consciousness, such as those involved in linguistically reporting the occurrence of the state, which is often used a criterion for determining when a state is conscious. Such causes or effects of consciousness would also be picked out by contrast studies as differentially present only in the conscious cases. So additional argument is needed to move from correlational data about NCCs to any conclusion about the neural substrate of consciousness itself.

Using the contrast method, it has also been possible to learn a lot about the functional differences between conscious and unconscious mental states and processes, e.g., between conscious and unconscious visual perceptions. Information that is consciously perceived or remembered is available for a wider more flexible range of uses and applications and is integrated with a wider range of other information. Thus conscious information is in that sense *more unified*—both functionally and in terms of its content. Conscious information exists within a larger more interactive network of connections. Given that enhanced functional and informational integration is one of the key features that distinguishes conscious from unconscious states, it will be appropriate to focus the last section of this paper on the unity of consciousness.

One important answer to the "What question" about the principal features of consciousness is that consciousness is *unified* in various ways. We observe this by both first person and third person methods. Understanding consciousness in part involves understanding what those *various unities of consciousness* are, and how they are produced. Though consciousness is generally agreed to be unified in some important respect, there less clarity or consensus about the specific respects in which it is unified and to what degree. Competing theories differ both in what they take the relevant form(s) of unity to be and what their status is: Are they necessary features of consciousness? Sufficient for consciousness? Or merely associated with consciousness. There is also disagreement about the order of dependence between consciousness and unity. Is the relevant information *conscious because it is unified*? Or is the information *unified because it is conscious?* The first option treats the relevant sort of unity as *constitutive* of conscious; it is being unified in the required way that the makes the state (or its informational content) conscious. Consciousness just is unity of the requisite type. On the second option, consciousness if regarded as having an essential nature that is distinctive from unity but which produces unity. On this second view, unity is an effect of consciousness rather than its constitutive essence.

The thesis that consciousness is unified has a long history in philosophy. In the seventeenth century René Descartes wrote, "For in truth, when I consider the mind, that is, when I consider myself in so far only as I am a thinking thing, I can distinguish in myself no parts, but I very clearly discern that I am somewhat absolutely one and entire; and although the whole mind seems to be united to the whole body, yet, when a foot, an arm, or any other part is cut off, I am conscious that nothing has been taken from my mind; nor can the faculties of willing, perceiving, conceiving, etc., properly be called its parts, for it is the same mind that is exercised in willing, in perceiving, and in conceiving, etc." (Descartes ,*Meditation VI*, section 1641).

There are many different types of conscious unity. One major division is between synchronic forms of unity, which involve unity at a single moment of time, and diachronic forms of unity which involve unity relations across extended periods of time. For example, the unity involved in perceiving an integrated real world scene seems primarily like a form of synchronic unity—involving unification of all the diverse contents that are represented by that momentary perceptual states. By contrast, the unity involved in episodic or

autobiographical memory seems to involve a diachronic form of unity, in so far as it involves an identification of past and presents selves as the single ongoing personal subject of experience. When I have such a memory, I appear to remember myself having that past conscious experience—I do not just remember that I drank coffee, I remember myself consciously drinking the coffee. Thus such memory depends in part on diachronic forms of unity as does very nature of personal identity itself.

Both synchronic and diachronic unity occur in many different specific forms, of which some of the most important include the following:

• Representational Unity—(coherent connections among contents)
• Object Unity—(representational unity of multiple features as present in a single object)
• Multi-Object/Scene Unity—representational unity of multiple features and objects as present in a single integrated scene or situation.
• World Unity—representational unity of multiple feature, objects, scenes and events existing as parts of a single unified world.
• Spatial Unity—representation of multiple, dimensions, and locations present as parts of a larger unified space.
• Multi-modal Unity—Unity in perceptual representation or awareness of information derived from multiple senses such as touch, vision and audition.
• Subject Unity—the unity of the conscious subject both at a time and across time.
• Introspective Unity—unity of the information made available by introspection from multiple aspects of the mind.
• Phenomenal Unity—the sort of unity present in phenomenal conscious experience, i.e., as a feature which is present phenomenally as part of our experience, one that we experience as a unity—what we might call *experience unity*.

The so called "binding problem" in visual perception illustrates several of these types of unity. The binding problem is in essence the question of explaining how the many local partial representations that are activated in specific regions of the visual cortex related to different aspects of the visual stimulus are bound into a single unified representation or precept. For example *shape, color* and *motion* are all separately computed by distinct local regions of the visual cortex. How are they bound into a single unified representtation of a quickly moving red circle? It is not done by sending

their results to some executive module—a "homunculus" —where a single combined representation is produced. The representations of color, shape and motion continue to be realized locally by neural activations in their distinct specialized cortical regions. Thus binding must involve some larger relation among those local activetions. Some dynamic relation need to be established among them to bind them together into a single integrated percept of the external object, scene and world. Some have proposed that they are bound together by synchronized oscillations, in particular of gamma oscillations in the 40 Hz band. The idea is that neuron groups that oscillate and fire together also represent together in a unified and integrated way. The hypothesis remains under active consideration and debate.

Many current theories of consciousness appeal to some form of integration or unification between diverse items of information, contents, functions or subsystems as a key feature of consciousness. They all explain the transition from merely unconscious information or mental states and to conscious ones as involving the addition of some larger unifying relation among local states that are narrower both in their contents and in their specific neural bases. However, they differ in about what they take that larger unifying relation to be.

The global neuronal workspace model (GNWS) was first proposed by psychologist Bernard Baars in the late 1980's (1988) and further developed by numerous researchers especially by Stanislas Dehaene (2014) and his research group in Paris. According to GNWS, the transition from unconscious to conscious state involves the incorporation of the relevant state and its content into global activation structure that the makes that information globally available to other modules for use and that is mediated in part by reciprocal relations between the local state and frontal and parietal attention mechanisms. Information or other mental contents become conscious when they enter a functionally defined "workspace" that makes them globally available to diverse systems and modules throughout the brain or mind. Conscious contents are thus widely available and highly integrated; all the contents present in the global workspace at a time are simultaneously available to many "consumers" throughout the mind, as well as being highly integrated with each other.

The neuroscientist Giulio Tononi (2008) has proposed a more general abstract theory of consciousness, Integrated Information Theory (IIT) that also defines consciousness in terms of integration

but does so mathematically in terms of an information theory based dimension he labels Φ. What makes a content conscious according to IIT is that is included in whatever subsystem has the highest Φ value, where Φ is jointly determined by the amount of information carried as well as the degree of integration or interdependence among those items of information.

Coming at the problem from a more philosophical and first-person perspective, the philosopher Tim Bayne (2010) has offered a theory of what he calls phenomenal unity, a distinctive form of unity present as a phenomenal feature of our experience. As Baynes writes, "Over and above these unities is a deeper and more primitive unity: the fact that these two experiences possess *conjoint experiential character*. There is something that is like to hear the rumba, there is something that it is like to see the bartender work, and there is something that it is like to hear the rumba while seeing the bartender work. Any description of one's overall state of consciousness that omitted the fact that these experiences are had together as components, parts or elements of a single conscious state would be incomplete." (Bayne 2010)

Phenomenal unity according to Bayne is a synchronic feature of all our experience and involves the sense in which we experience all our conscious states as occurring together as the states of a single unified self. Bayne's theory is thus directly relevant to the fact noted above about that subjectivity requires a subject. The unity of consciousness, at least Bayne's subjective phenomenal unity of consciousness, depends explicitly on its relation to the unified subject of experience. Bayne defines *phenomenal unity* as a relation between pairs or groups of experiences: Two experiences E1 & E2 are *phenomenally unified* just if they have a *conjoint phenomenal character*. The latter notion is in turn defined subjectively in terms of a special sort of 'what it is likeness': E1 & E2 have a *conjoint phenomenal character* just if there is *something that it is like to experience them together*–(where that involves *not merely the conjunction* of the what-it-is-likeness of having an experience of A and the what-it-is-likeness of having an experience of B, but also the what-it-is-likeness of *having them together* – hearing the rumba *and* tasting the coffee…).

With phenomenally unity thus defined, Bayne goes on to make a strong claim about the phenomenal unity of human experience. He asserts the Unity Thesis: *All the experiences that a conscious subject has at a time are phenomenally unified with each other*. It is a claim about *synchronic unity* and asserts that such unity is a universal (and perhaps

necessary) feature of all (human) consciousness. According to the Unity Thesis, all the experiences that a conscious subject has at a moment must be phenomenally unified with each other. Conscious subjects never have any experiences at a moment that fail to be unified with each other; they are all experienced by the subject as conjointly occurring together.

In assessing the Unity Thesis, a question immediately arises as to how to interpret the notion of a *subject* or *single subject*. The Unity Thesis is a claim about a relation holding among all the experiences of a given subject at a moment, but how do we individuate subjects? What must be true for a set of experiences at a moment to count as being *experiences of one single subject*? It is important to avoid turning the Unity Thesis into a tautology. Thus Bayne must not explain the relation of being *experiences of a single subject* in terms of being *experiences that are conjointly experiences*. Bayne rightly interprets and defines the notion of a *single subject* independently of experienced conjointness. For most of his 2010 book (chapters 1-11) Bayne treats the conscious subject as the *human organism*. So understood and substituting *organism* for *subject*, the Unity Thesis asserts that that all the experiences had by a human being *qua organism* at a time are phenomenally unified with each other.

The Unity Thesis is a factual empirical claim about the reality of actual human experience. How well supported is it by the empirical evidence, whether that of first person introspection or scientific third person investigation? Introspection may seem to show that all our experiences are phenomenally unified but perhaps that is simply an artifact of introspection. Whenever we look at two experiences to ask if they are conjointly experienced, of course we find they are. But perhaps that is true only because we are attending to them whenever we ask whether or not they are conjoint. Asking the question and attending to them may suffice to make them conjoint, but it does not settle the matter about all the experiences we have any moment to which are not attending and all the experiences we have when we are not asking about conjointness. Introspection's apparent support for the Unity Thesis could be just a case of a so called "refrigerator illusion." Whenever we look in the fridge, the light is on. Whenever we look for conjointness, we find it.

A variety of empirical cases may seem to contradict the Unity Thesis, especially ones that involve patients suffering from mental deficits that result in certain forms of disunity. Bayne considers many such cases and aims to answer each challenge. Among other

disorders, he considers anosagnosia (mental deficit together lack of awareness of one's deficit), simultanagnosia (inability to be simultaneously conscious of two distinct objects), multiple personality (dissociative personality) disorder and split brain cases. Bayne attempts to respond to all, but succeeds more convincingly with some than with others.

The ones that pose perhaps the most serious challenge to the Unity Thesis are the split brain cases. The description "split brain" refers to the fact that for medical reasons such patients have had surgery that severed the large bundle of neural fibers that connects their two cerebral hemispheres (the corpus callosum). After the surgery, they appear normal in everyday situations, but laboratory tests show that disconnecting their two hemispheres in fact produces an important degree of disunity in perception and action. For example, if differing visual stimuli are presented briefly to their respective visual fields, the information that reaches the visual processing areas of their two hemispheres also differs. Given the normal contralateral association between hemispheres and visual fields, the severed left hemisphere has access only to the information from the right visual field and the right hemisphere has access only to the information from the left visual field. In normal subjects visual about left and right fields is shared across the corpus callosum, but such sharing does not take place in the split brain patients.

The subjects are tested and asked to respond to what they briefly saw—e.g. by picking up the matching object (from a set of objects behind a screen), or pointing to a picture of an object that is related to what they saw. The subjects in these situations respond differently with their left and right hands. Indeed they sometimes do so simultaneously. Each hand is controlled by a single hemisphere (again the contralateral one—i.e. left hemisphere controls the right hand), and thus each hand acts in a way that reflects the information available to the hemisphere controlling it. The right hand responds to what was in the right visual field because that was the information available to the left hemisphere that controls it, and the left hand similarly responds to what was shown on the left.

Interestingly when asked to explain why they responded as they did, subjects can accurately explain why they did what they did with their right hand, but are typically unable to accurately explain why they did what they did with their left hand. When asked to explain, they either say they do not know or confabulate, i.e. make an a plausible but incorrect story. It is literally a case of "the right hand

not knowing what the left hand is doing." The subjects cannot say why they responded as they did with their left hands because language is typically lateralized to the left hemisphere, especially in right handed people. Given the severed corpus callosum, the left hemisphere regions that control speech do not have access to the right hemisphere's information about the left visual field or about the control of the left hand. The right hemisphere language regions can thus produce accurate answers only to questions about the subject's right hand responses.

Split brain cases seem to challenge the Unity Thesis. If we follow Bayne and define *single subject* as *single human organism*, then there is only one subject (one human organism) present in the split brain cases. Yet that single subject seems to have disunified experiences. In particular it seems that a split brain patient in the test situations has simultaneous visual experiences of his right and left visual fields—call them ER and EL—that are not phenomenally unified. ER and EL are not phenomenally conjoint; the split brain subject does not have any what-is-it-likeness of ER and EL co-occurring as part of his experience. Bayne offers a response, but one that is less than convincing. He proposes a *switching model* according to which consciousness rapidly switches back and forth between the left and right hemispheres in split brain cases. Only one hemisphere is conscious at a moment. Thus the switching model, if it were true, would give the Unity Thesis a way to avoid conflict with the split brain cases. If consciousness rapidly switches and only one hemisphere is conscious at a given moment, then the subject never has ER and EL simultaneously. If they are not simultaneous, then the fact that they are not phenomenally unified does not contradict the Unity Thesis. Thus the data about the split brain cases in itself does not refute the Unity Thesis, but Bayne's means of preserving it requires accepting the switch model, and that hypothesis does seem a bit ad hoc and not strongly motivated beyond its roles in saving the Unity Thesis. Indeed as noted above, split brain subjects sometimes seem to act simultaneously and differentially with their two hands. It is difficult to conclusively prove that very rapid switching does not occur in such cases, but in the absence of strong independent reason to believe in such switching, the split brain cases continue to pose an important open challenge to the Unity Thesis.

In the final chapter (chapter 12) of his book, Bayne considers a different notion of "conscious subject" that might yield a version of the unity thesis as an apriori necessary truth, one that might

explain *why* the Unity Thesis holds for humans. That notion of subject appeals the idea of the Self as a virtual structure—the point-of-view of the experiential subject akin to Daniel Dennett's notion of the self as "the center of narrative gravity, according to which the self is the *point of view of the serial narrative* that is created by the interpretative "stream of experience" (1992). Importantly on Dennett's view, the self is not the author of the narrative steam. The self does not create the story. It is the *story (the serial narrative)* that is primary, and the self is *implicit in that story*; the self is the point of view from which the story coheres. As such the self is an intentional structure implied by the story. But would such virtual selves be real enough as a theory of the self? Do we need a more robust notion of the self as real. Anticipating such criticism Baynes writes, "The worry is this: if the self is a 'merely intentional entity' then does it follow that it is unreal that selves don't really exist" (Bayne 2010, p.292). To this imagined worry he replies, "But there is no kind of real self with which our kind of selves could be contrasted, for it is in the very nature of selves to be virtual. The kind of selves that we possess are as real as selves get. This kind of reality might not be enough for some, but I think it provides all the reality that we might have reasonably hoped for here. Perhaps more importantly, it provides all the reality that we *need."* (Bayne 2010, p. 293)

Will virtual selves suffice? We may be reluctant to cocede that. We feel virtual selves cannot be the full story; there must be more to the self than Bayne's virtual phenomenalism asserts. How can the self be nothing more than an intentional entity, on a par with the character from whose point of view the narrative of a novel coheres? One intuitively objects, "Surely there must be more to the self than that." Even if the virtual self is part of the story, it does not seem as if it could be the full story of the self.

Can we adopt the notion of a virtual self but use it in a somewhat different way to construct a theory of the self that is more robustly realist about the self than Bayne's virtual phenomenalism? I believe may indeed be possible, and I offer in what follows a specific proposal that aims to do just that (Van Gulick 2105). Let us begin by returning to a point raised earlier, to which I said we would return: the fact that subjectivity requires a subject. Reference to the self is implicit in the very phenomenal and intentional structure of experience. Experiences function and represent *as experiences of a subject,* and thus their intentionality involves an inherent reference to a subject or self. Contrary to Humean atom-

ism, there cannot be a pain that no subject feels, nor an experience of red that is not a red *for* any subject. Similarly, Jean Paul Sartre (1943) argued that there can be no object of experience, no "*en soi*" (in itself) without a "*pour soi*" (for itself). There must be some subject or self *for whom* it is an experience and *to whom* the object of experience is present. Experience is always experience *for a self*. It is the self in this sense, the self as *pour soi,* that is implicitly present as a part of the phenomenal intentionality of any experience.

Building on this idea, we can begin to construct an alternative theory of the self that incorporates the following principles.

1. A conscious mental state CM (or experience E) can exist at time t only if there is "something that it is like" to be in the state (have that experience) at t.

2. There can be something that it is like to be in CM (or have experience E) at t only if *there is some self or subject for whom it is like some way to be in CM (or have E) at t.*

So the next question is: What must true in for there to be such a self or subject for whom it is someway to be in CM (or to have E)?

It is here that the notion of a virtual self gets applied but in a way somewhat different way from Bayne or Dennett. The basic idea is that a mental state M can exist as a conscious mental state CM (or experience E) only if it is contained within a set of representations whose contents are integrated or unified in a way that implies the existence of a single self or subject. The individual experience must occur within that larger intentional structure that implicitly defines the perspective of a virtual self. The proposal is to identity the self with the total system of experiences when they cohere "as from the point of view of a single self." The virtual self is the point of view defined by such a coherent set of experiences, and it is only when a set of experiences within a system defines such a coherent point of view, i.e. only when it defines a virtual self, that the system of experiences itself constitutes a self. The self needs to be constructed, and constructing it is in part of a matter of producing a set of experiences that cohere from the point of view of a virtual self. But when the process succeeds, the self that is produced is a real self.

Finally let us return to our initial question. What are our prospects for dispelling the mystery of consciousness? Will we able to explain consciousness, especially subjective phenomenal what-

it-likeness consciousness, in terms of neural or other objective physical processes? Will we be able to solve the Hard Problem? At present the question remains open. No adequate explanation as yet exists that lets see with intuitive understanding *how the trick is done.* But there is reason to believe we are making progress and thus for optimism about our eventual success. Theories like the global workspace model offer promising models of how many features of consciousness might be related to underlying neural substrates. Such theories are most promising in dealing with the functional aspects of consciousness but less so with subjective phenomenal consciousness. However, even with regard to subjectivity there is reason to believe that more research will lead to a successful constructive theory. And in pursuing that goal we should keep in mind, what we just seen so clearly illustrated just above: subjectivity requires a subject, and if you want a constructive theory of subjective consciousness, you need to figure out how to construct a subject or self out of parts that are not. Perhaps by the end of this century, the mystery of consciousness will be solved and the Gordian knot untied. Or perhaps not. We will have to wait and see.

References

Baars, Bernard (1988). *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press.

Bayne, Tim (2010). *The Unity of Consciousness.* Oxford: Oxford University Press.

Chalmers, David (1996). *The Conscious Mind.* New York: Oxford University Press.

Churchland, Patricia. (1994). "Can neurobiology teach us anything about consciousness?." Presidential Address to the American Philosophical Association, Pacific Division. In: *Proceedings and Addresses of the American Philosophical Association*. Lancaster, PA: Lancaster Press. 67-4: 23-40.

Dehaene, Stanislas (2014). *Consciousness in the Brain.* New York: Penguin.

Descartes, René (1641). *Meditations on First Philosophy.* Paris.

Dennett, Daniel (1992). *Consciousness Explained.* Boston: Little Brown.

Flanagan, Owen (1992). *Consciousness Reconsidered.* Cambridge: MIT Press.

Gertler, Brie (2007). In efense of mind body Dualism." (2007) InJ. Feinberg and R. Shafer-Landau, eds., *Reason and Responsibility*, 13th edition. Belmont, CA: Wadsworth

Haldane, John B. S. (1913). *Mechanism, Life and Personality: An Examination of the Mechanistic Theory of Life and Mind.* London: John Murray.

Huxley, Thomas (1866). *Lessons on Elementary Physiology* 8. London: MacMillan.

McGinn, Colin (1989). "Can we solve the mind body problem?" *Mind* 98:391, pp. 349-366.

McGinn, Colin (2000). *The Mysterious Flame: Conscious Minds in a Material World.* New York: Basic Books.

Nagel, Thomas (1974) "What is it like to be a bat." *The Philosophical Review,* 83:4, pp. 435–450.

Prinz, Jesse (2012). *The Conscious Brain: How Attention Engenders Consiousness.* New York: Oxford University Press.

Sartre, Jean Paul (1943). *L'Être et le néant : Essai d'ontologie phénoménologique Being and nothingness: An essay on phenomenological ontology).* Paris: Gallimard.

Searle, John (1997). *The Mystery of Consciousness.* New York: The New York Review of Books.

Strawson, Galen, (2006). *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?* Thorveton, UK: Imprint Academic

Tononi, Giulio. (2008). Consciousness as integrated information: a provisional manifesto. *Biological Bulletin* 215: 216–42.

Van Gulick, Robert (2005). "Consciousness." *Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/consciousness/*

Van Gulick, Robert (2015). "E pluribus unum: Rethinking the unity of consciousness." In C. Hill and D. Bennett (eds.): *Sensory Integration and the Unity of Consciousness.* Cambridge, MA: MIT Press, 375-92

Wittgenstein, Ludwig (1953). *Philosophical Investigations.* Oxford: Basil Blackwell Ltd.