The Function of Conscious Experience: An Analogical Paradigm of Perception and Behavior

Steven Lehar

slehar@cns.bu.edu

Submitted to Consciousness and Cognition

Abstract

The question of whether conscious experience has any functional purpose depends on a more fundamental issue concerning the nature of conscious experience. In particular, whether the world of experience is the external world itself, as suggested by direct realism, or whether it is merely a virtual-reality replica of that world in an internal representation, as in indirect realism, or representationalism. There is an epistemological problem with the notion of direct realism, for we cannot be consciously aware of objects beyond the sensory surface. Therefore the world of experience can only be an internal replica of the external world. This in turn validates a phenomenological approach to studying the nature of the perceptual representation in the brain. Phenomenology reveals that the representational strategy employed in the brain is an analogical one, in which objects are represented in the brain by constructing full spatial replicas of those objects in an internal representation.

Introduction

The question of the functional role of conscious experience is currently an active area of debate. On the one side there are those like Dennett (1988, 1991) who argue that consciousness is an epiphenomenon, with no direct functional value. Others, such as Humphrey (1999 p. 250) argue that consciousness must have some adaptive value on evolutionary grounds, for nothing can evolve by natural selection unless it has some effect on behavior. The question is a paradigmatic one in the Kuhnian sense (Kuhn 1970), because the differences of opinion on the function of conscious experience reflect deeper differences on the more fundamental question of what consciousness itself actually is. I propose that the debate over the ontological status of conscious experience in turn rests on the epistemological question of whether the world we see around us is the real world itself, or whether it is merely a virtual-reality replica of that external world in an internal representation. Until this central issue is resolved on sound logical grounds, the sciences of psychology and consciousness studies are condemned to remain in a pre-paradigmatic state, with opposing camps arguing at cross-purposes due to lack of consensus on the foundational issues of the science.

If we accept the modern materialist view of mind as the operation of the physical brain, then the epistemological question is not open; there is only one reasonable interpretation of the ontology of conscious experience, i.e. that consciousness is in fact an internal replica of the external world rather than

the world itself. This in turn validates a *phenomenological* approach to the study of conscious experience, i.e. to examine the world around us not as a scientist examining an objective external world, but as a *perceptual* scientist examining a rich and complex internal representation. I will show how the phenomenological approach can be employed to examine both the structure of conscious experience, and also the detailed workings of the computational strategy or algorithm that guides behavior. This approach to the study of conscious experience clearly demonstrates that consciousness is not an epiphenomenon, but serves an essential functional role, which is to provide an analogical representation of the external world, that operates in conjunction with an analogical computational strategy that guides behavior. In other words perception and behavior are intimately coupled through the agency of conscious experience, and careful examination of the properties of that experience offers insights into the nature of both perception and of behavior.

The Epistemological Divide

The debate over the nature of conscious experience is confounded by the deeper epistemological question of whether the world we see around us is the real world itself, or merely an internal perceptual copy of that world generated by neural processes in our brain. In other words this is the question of *direct realism*, also known as *naive realism*, as opposed to *indirect realism*, or *representationalism.* Although this issue is not much discussed in contemporary psychology, it is an old debate that has resurfaced several times, but the continued failure to reach consensus on this issue continues to bedevil the debate on the functional role of conscious experience. The reason for the continued confusion is that both direct and indirect realism are frankly incredible, although each is incredible for different reasons.

Problems with Direct Realism

The direct realist view (Gibson 1972) is incredible because it suggests that we can have experience of objects out in the world directly, beyond the sensory surface, as if bypassing the chain of sensory processing. For example if light from this paper is transduced by your retina into a neural signal which is transmitted from your eye to your brain, then the very first aspect of the paper that you can possibly experience is the information at the retinal surface, or the perceptual representation that it stimulates in your brain. The physical paper itself lies beyond the sensory surface and therefore must be beyond your direct experience. But the perceptual experience of the page stubbornly appears out in the world itself instead of in your brain, in apparent violation of everything we know about the causal chain of vision. The difficulty with the concept of direct perception is most clearly seen when considering how an artificial vision system could be endowed with such external perception. Although a sensor may record an external quantity in an internal register or variable in a computer, from the internal perspective of the software running on that computer, only the internal value of that variable can be "seen", or can possibly influence the operation of that software. In exactly analogous manner the pattern of electrochemical activity that corresponds to our conscious experience can take a form that reflects the properties of external objects, but our consciousness is necessarily confined to the experience of those internal effigies of external objects, rather than of external objects themselves. Unless the principle of direct perception can be demonstrated in a simple artificial sensory system, this explanation remains as mysterious as the property of consciousness it is supposed to explain.

Problems with Indirect Realism

The indirect realist view is also incredible, for it suggests that the solid stable structure of the world that we perceive to surround us is merely a pattern of energy in the physical brain, i.e. that the world that appears to be external to our head is actually inside our head. This could only mean that the head we have come to know as our own is not our true physical head, but is merely a miniature perceptual copy of our head inside a perceptual copy of the world, all of which is completely contained within our true physical skull. Stated from the internal phenomenal perspective, out beyond the farthest things you can perceive in all directions, i.e. above the dome of the sky and below the earth under your feet, or beyond the walls, floor, and ceiling of the room you perceive around you, beyond those perceived surfaces is the inner surface of your true physical skull encompassing all that you perceive, and beyond that skull is an unimaginably immense external world, of which the world you see around you is merely a miniature virtual-reality replica. The external world and its phenomenal replica cannot be spatially superimposed, for one is inside your physical head, and the other is outside. Therefore the vivid spatial structure of this page that you perceive here in your hands is itself a pattern of activation within your physical brain, and the real paper of which it is a copy it out beyond your direct experience. Although this statement can only be true in a topological, rather than a strict topographical sense, this insight emphasizes the indisputable fact that no aspect of the external world can possibly appear in consciousness except by being represented explicitly in the brain. The existential vertigo occasioned by this concept of perception is so disorienting that only a handful of researchers have seriously entertained this notion or pursued its implications to its logical conclusion. (Kant 1781/1991, Koffka 1935, Köhler 1971, Russell 1927, Boring 1933, Feigl 1958, Smart 1959, Smythies 1989, 1994, Harrison 1989, Hoffman 1998, Lehar 2002)

Another reason why the indirect realist view is incredible is that the observed properties of the world of experience when viewed from the indirect realist perspective are difficult to resolve with contemporary concepts of neurocomputation. For the world we perceive around us appears as a solid spatial structure that maintains its structural integrity as we turn around and move about in the world. Perceived objects within that world maintain their structural integrity and recognized identity as they rotate, translate, and scale by perspective in their motions through the world. These properties of the conscious experience fly in the face of everything we know about neurophysiology, for they suggest some kind of three-dimensional imaging mechanism in the brain, capable of generating three-dimensional volumetric percepts of the degree of detail and complexity observed in the world around us, that are able to rotate and translate freely relative to the space in which they appear. No plausible mechanism has ever been identified neurophysiologically that exhibits this incredible property. The properties of the phenomenal world are therefore inconsistent with contemporary concepts of neural processing, which is exactly why these properties have been so long ignored.

Problems with Projection Theory

There is a third alternative besides the direct and indirect realist views, and that is a *projection theory,* whereby the brain does indeed process sensory input, but that the results of that processing get somehow projected back out of the brain to be superimposed on the external world (Ruch 1950 quoted in Smythies 1954, O'Shaughnessy 1980 pp 168-192, Velmans 1990, Baldwin 1992). According to this view, the world around us is part real, and part perceptual construction, and the two are spatially superimposed. However no physical mechanism has ever been proposed to account for this external projection (Smythies 1954). The problem with this notion becomes clear when considering how an artificial intelligence could possibly be endowed with this kind of external projection. Although a sensor may record an external quantity in an internal register or variable in a computer, there is no sense in which

that internal value can be considered to be external to that register or to the physical machine itself, whether detected externally with an electrical probe, or examined internally by software data access. Unless the principle of external projection can be demonstrated in a simple artificial sensory system, this explanation too remains as mysterious as the property of consciousness it is supposed to explain.

Selection from Incredible Alternatives

We are left therefore with a choice between three alternatives, each of which appears to be absolutely incredible. Contemporary neuroscience seems to take something of an equivocal position on this issue, recognizing the epistemological limitations of the direct realist view and of the projection hypothesis, while being unable to account for the incredible properties suggested by the indirect realist view. However one of these three alternatives simply must be true, to the exclusion of the other two. And the issue is by no means inconsequential, for these opposing views suggest very different ideas of the function of visual processing, or what all that neural wetware is supposed to actually *do*. Therefore it is of central importance for psychology to address this issue head-on, and to determine which of these competing hypotheses reflects the truth of visual processing.

The problem with the direct realist view is of an epistemological nature, and is therefore a more fundamental objection, for direct realism is nothing short of magical, that we can see the world out beyond the sensory surface. The projection theory has a similar epistemological problem, and is equally magical and mysterious, suggesting that neural processes in our brain are somehow also out in the world. Both of these paradigms have difficulty with phenomena of dreams and hallucinations (Revonsuo 1995), which present the same kind of phenomenal experience as spatial vision, except independently of the external world in which that perception is supposed to occur in normal vision. It is the implicit or explicit acceptance of this naive concept of perception that has led many to conclude that consciousness is deeply mysterious and forever beyond human comprehension. For example Searle (1992) contends that consciousness is impossible to observe, for when we attempt to observe consciousness we see nothing but whatever it is that we are conscious of; that there is no distinction between the observation and the thing observed. This is also the "*Problem of Transparency*" in Tye's (1995) *Ten Problems of Consciousness.*

The problem with the indirect realist view on the other hand is more of a technological or computational limitation, for we cannot imagine how contemporary concepts of neurocomputation, or even artificial computation for that matter, can account for the properties of perception as observed in visual consciousness. It is clear however that the most fundamental principles of neural computation and representation remain to be discovered, and therefore we cannot allow our currently limited notions of neurocomputation to constrain our observations of the nature of visual consciousness. The phenomena of dreams and hallucinations clearly demonstrate that the brain is capable of generating vivid spatial percepts of a surrounding world independent of that external world, and that capacity must be a property of the physical mechanism of the brain. Normal conscious perception can therefore be characterized as a guided hallucination (Llinás & Paré 1991, Revonsuo 1995), which is as much a matter of active construction as it is of passive detection. If we accept the truth of indirect realism, this immediately disposes of at least one mysterious or miraculous component of consciousness, which is its unobservability. For in that case consciousness is indeed observable, contrary to Searle's contention, because the objects of experience are first and foremost the product or "output" of consciousness, and only in secondary fashion are they also representative of objects in the external world. Searle's difficulty in observing consciousness is analogous to saying that you cannot see the moving patterns of glowing

phosphor on your television screen, all you see is the ball game that is showing on that screen. The indirect realist view of television is that what you are seeing is first and foremost glowing phosphor patterns on a glass screen, and only in secondary fashion are those moving images also representative of the remote ball game.

The choice therefore is that either we accept a magical mysterious account of perception and consciousness that seems impossible in principle to implement in any artificial vision system, or we have to face the seemingly incredible truth that the world we perceive around us is indeed an internal data structure within our physical brain. If science is to triumph over mysticism, we are compelled to accept the latter view, and accept the reality of conscious experience as a direct manifestation of neurophysiological processes within our physical brain. This in turn validates a *phenomenological* approach to the study of conscious experience, i.e. to examine the world around us not as a scientist examining an objective external world, but as a *perceptual* scientist examining a rich and complex internal representation. I will show how phenomenological observation can be used to determine the dimensions of conscious experience, and what the structural form of conscious experience tells us about the representational strategy used in the brain. I will then show how the phenomenological technique can also be employed to determine the functional properties of the conscious experience, or how the information encoded in perception is used to guide behavior.

A History of the Epistemological Debate

The question of the epistemology of conscious experience has a long and colorful history in psychology and philosophy. The epistemological question is intimately related to the issue of mind-body dualism, because all of the problems inherent in the naïve realist view simply evaporate if we allow for the existence of an immaterial soul, whose function is not entirely dependent on the mechanical functioning of the physical brain. For in visual perception we do indeed feel as if our consciousness extends outwards beyond the confines of our physical head, to make direct sensory contact with remote objects in the world. But while the external nature of perception has been recognized into ancient times, so too has the causal role of the sensory organs in perception, for the world goes dark when we close our eyes, and disappears altogether when we lose our eyes. We have therefore two fundamental and apparently contradictory observations on the nature of vision, one that appears to lead inwards from the world through the eye to the brain as revealed by physiology, while the other appears to lead outward from the mind to the external visual world as revealed by phenomenology. With the rise of the materialistic world view ushered in by Newton's mechanical universe, there has been ever increasing pressure to formulate a materialistic explanation of the mystery of visual consciousness. As we review the history of this problem, a gradual progression is observed, as ever more of the properties of the phenomenal world, initially assumed to be identical to the external world, are attributed instead to processes internal to the brain and body. This leaves the phenomenal portions of the external world in a peculiar kind of limbo, being observed external to the body while being attributed to processes within the body. The ultimate solution to this paradox is only to be found by a complete indirect realist inversion, whereby everything that appears to be external is finally attributed unambiguously to internal processes in the brain, a view which forever closes the naïve window of direct observation onto the external world.

The first stage of the epistemological inversion was proposed by Descartes. To the modern reader Descartes' view is remarkable in its dualism, i.e. its continued commitment to the immaterial soul of ancient religion. But to the philosophers of his day, the most remarkable aspect of the Cartesian view was

how much of the perceptual process he attributed to the physical body, and how far the immaterial soul had retreated within it. For Descartes proposed that all sensory information is transmitted by the nerves to a central "theatre", where the soul makes contact with the physical body, a mind / body interface which Descartes located in the pineal gland. The prevailing view at the time was that the soul extended throughout the body and beyond, and this accounted for the external nature of perception. (Pastore 1971, p. 19) Descartes' rejection of this thesis was based on physiological grounds. For anatomy reveals the sensory organs connected to the brain by nerves, which suggests a propagation of sensory information inwards from the external world to the brain. But the full implication of Descartes theory would have been that the world we see around us is the picture in the pineal gland projected by the nervous system to the soul, for now the soul is shut in to the physical head, and can only see outwards indirectly by way of the senses. However Descartes was not prepared to accept the full implications of his own theory, so he appealed to the mystery of the soul to account for the external nature of perception, proposing that as the soul receives the sensory signals at the pineal gland, it instantaneously becomes aware of the external objects which are the ultimate source of the sensory signals. The role of the sensory organs is thereby somewhat ambiguous, being required to transmit sensory information from the world to the soul, but the soul does not even "see" this internal picture, but instead sees the world itself directly.

Subsequent explanations of the epistemological question were equally confused and self-contradictory. Malebranche (1674) proposed that sensory qualities, such as color, taste, and sounds, are not part of the external object but are internal phenomenal properties. The mind "becomes colored" when we view a colored object, although the object on which that color is perceived is not an internal replica, but is seen directly in the world where it lies. A perceived object therefore is partially an internal and partially an external phenomenon, with shape and volume existing out in the external world itself, but the color clothing the object being something in the mind. John Locke (1690) took the epistemological inversion one step farther to propose that even shape and volume were sensory properties, and therefore they too, like color, taste, and sound are properties of the mind rather than of the world itself. However Locke retained a residual aspect of direct perception in his distinction between the *primary* and *secondary* qualities of perception. The primary qualities, which include shape, size, and number, are, he claimed, a veridical representation of the actual properties of the objects themselves, while the secondary qualities such as color, smell, taste, warmth or coldness, or pain, are also produced by stimulation of sense but they have no resemblance to the corresponding qualities in the objects. (Pastore 1971, p. 64-66) The isolation of the soul from direct contact with the external world led Bishop Berkeley to question the very existence of the external world, for since we cannot see it directly, how can we know that it even exists? Berkeley therefore professed a form of idealism, whereby mind is all that really exists, as captured by Berkeley's dictum that "to be is to be perceived." As long as the choice is restricted to epistemological monist alternatives, the choice is between two incredible possibilities, either naive realism or idealism.

Immanuel Kant was the first to propose the full epistemological inversion. Kant proposed that there are in fact two worlds of reality, the *phenomenal world* of conscious experience and the *nouminal world* of objective external reality. The only way that we can observe the nouminal world is by its effects on the phenomenal world. But Kant's epistemology was confused by his acceptance of Berkeley's idealism, whereby the external world is not the material world familiar to modern science, but is itself a world of mind, although it is the mind of God, and thereby external to the mind of the individual percipient. But appeals to God and to the immaterial soul are ultimately unsatisfactory for a scientific theory because spiritual entities are in principle beyond the bounds of scientific investigation. Therefore the modern epistemological debate only began in earnest after the materialistic view of mind had been generally

accepted, following the French materialists La Mettrie (1748) and Cabanis (1802) (see Boring 1950, p. 212-216) who first extended to man Descartes' notion that animals are automata.

The followers of Kant abandoned his epistemological dualism, and focused instead on Kant's idealism, claiming that mind is all that is knowable to man. This unfettered idealism eventually triggered a realist backlash by the American realist, or *neo-realist* philosophers (Holt *et al.* 1912) who professed what is plainly evident to the common man, that the external world *is* knowable through consciousness. But the neo-realists also advocated an epistemological monism whereby the world observed in conscious experience is the external world itself, and the epistemological debate had gone full circle back to the naive realism familiar to the common man. For the neo-realists claimed that even the secondary qualities such as color, smell, and taste, are objective properties of the external objects themselves, and are observed out in the external world, superimposed on the external objects to which they belong. But a naive realist philosophy cannot survive long against the obvious objections, and once again the same progression was replayed as one by one, ever more of the properties of the external world of perception were attributed instead to processes internal to the brain and body.

A more sophisticated epistemology was offered by the subsequent *critical realist* movement (Sellars 1916, Russell 1921, Broad 1925, Drake *et al.* 1920) who, like Malebranche and Locke, claimed that some aspects of the perceived object such as its color are in fact subjective and inhere in the mind, while other aspects such as its shape are properties of the objects themselves, and are observed directly out in the world on the objects themselves. But this explanation, which is intermediate between epistemological monism and dualism, led to the same kind of confusions as its earlier incarnation under Malebranche and Locke. For the perceived shape and perceived color appear superimposed in the same apparently external space, although one is supposedly a property of the mind which is presumably in the brain, whereas the other is a property of the external world which is outside the brain. The critical realists were vaguely aware of some kind of difficulty here, as they spent considerable effort agonizing over this muddled concept. For example Sellars (1916 p. 58) argued that "*if realism is to be saved, it must disembarrass itself of its immediatism, i.e. the physical object can no longer be regarded as immediately present in perception.*" This statement is consistent with the epistemological dualism of Kant. However later in the same book Sellars (1916 p. 244) argues that "*consciousness is not extended after the manner of a physical thing for the very simple reason that it is not a physical thing. Let all this be granted; yet in a very real sense consciousness is extended. As a variant of the brain it is in the brain, not as an ivory sphere is encapsulated in another in those curious products of Chinese patience which we see in museums, but in a unique way which it requires reflection to make clear.*" But no amount of reflection can clear up this mess, as seen in Sellars' conclusion (p. 246) that "*Consciousness in the brain is not the relation of one thing to another, but the immanence of that part of reality which is our changing field of experience to the rest of the same existential part of the physical world. Unfortunately, there is no adequate word to express what we think.*" The obvious confusion in these statements about the relation of mind to brain is a direct consequence of a confused epistemology whereby the mind is in the brain, and yet at the same time it is somehow *not* in the brain.

Even more problematic was the question of perceived shape. For this supposedly objective property of external objects nevertheless has a subjective aspect due to perspective. For example a penny viewed in perspective appears subjectively as an ellipse, and the aspect ratio of that elliptical percept depends on the angle from which the penny is viewed, so different viewers of the same object see it in different shapes. The critical realists offered all kinds of improbable explanations to account for this troublesome aspect of perception. One solution was the concept of *sense data*, or *sensa*, whereby one object can

produce a multitude of different sensa in different observers. (Broad 1925, p. 170 - 180) The sense data are neither physical nor mental, but "*particular existents of a peculiar kind; they are not physical, ... and there is no reason to suppose that they are either states of mind or existentially mind-dependent. In having spatial characteristics ... they resemble physical objects ... but in their privacy and their dependence on the body ... of the observer they are more like mental states.*" (Broad 1925, p. 181) Bertrand Russell (1921, p. 97- 98) made the most concerted effort to formulate the concept of sense data as independent of the individual observer by suggesting that every possible sensum exists as a potential perceptual experience as if there were a subject there to perceive it. For example when viewed face-on, a penny projects a circular aspect whose size diminishes with distance from the penny, while when viewed edge-on the penny projects a linear or rectangular aspect whose length diminishes with viewing distance, and intermediate to the face- on and edge-on views are an infinite set of elliptical aspects of that same penny through the full range from linear to circular, and through a range of sizes, depending on viewing angle and distance. All of these sensa, according to Russell, are objective external existents which are part of the penny whether it is being viewed by a percipient or not, and that is how different observers can see different aspects of thesame objective entity. The whole set of these sensa are to be taken as actually *being* the perceived penny, which is itself neutral between different observers. Curiously, since the penny cannot be viewed from inside its own volume, Russell proposed that the sense-data which *are* the penny do not extend to within its own volume, so the penny exists everywhere in space except at that location where common sense would indicate it to be. Again, the ineffable relation between the object and its infinite set of sensa follows directly from the epistemological confusion whereby the sensa are both in the brain and subjective, and yet at the same time they are out in the external world and somehow objective.

The critical realists all agreed on the fact that the sense data are independent of the object of perception itself. However the question of whether the sense data are part of the mind, or whether they are aspects of the external object is one on which critical realists differed in subtle ways. In a book on critical realism by a consortium of authors (Drake *et al.* 1920) Lovejoy, Pratt, and Sellars claim that the sensa are completely "*the character of the mental existent ... although its existence is not given*" whatever that might possibly mean, while Drake, Rogers, Santayana, and Strong agree that the data are characteristic of the apprehended object, although "*the datum is, qua datum, a mere essence, an inputed but not necessarily actual existent. It may or may not have existence.*" (p. 20-21 footnote), whatever that might possibly mean! So the critical realists solved the problem of sense-data by defining a unique kind of existent which may either be part of the external object, or of the internal mental state, but in any case it has a status of quasi-existence which supposedly escapes the problems inherent in identifying it explicitly as either an external or internal entity. Epistemological confusion inevitably leads to a confused philosophy.

Bertrand Russell, originally himself a critical realist, eventually discovered the resolution to this quandary with a realist version of Kant's epistemological dualism. What finally convinced Russell was consideration of the causal chain of vision. Light from an object in the world enters the eye, where it is transduced to a neural signal in the optic nerve, from whence it is eventually transformed into a pattern of activation in the visual cortex. There are two aspects of that perceptual activity, an electrophysiological aspect measurable by cortical electrodes, and a phenomenal or experiential aspect in the form of the percept itself. But the two are different manifestations of the same underlying structure, and therefore if the first is located within the physical brain, then the second must also be in the brain. (Russell 1927 p. 137-143) Russell observed that a potent source of confusion in this matter is a confusion of physical

space with perceptual space. For although our percept of the external world appears external to our head, it is not external to our true physical head, but only to our perceptual head in perceptual space. All of our perceptual space, including the externally perceived world, is inside our physical head in physical space. (Russell 1927) This explanation of perception finally resolved all of the epistemological problems inherent in naive realism and in idealism without resort to any supernatural gods or mystical souls. It accounts for the fact that the perceived world appears external although we know it to be internal, by the fact that the external world of perception is internal to our physical brain. It accounts for the realism known to common sense, by the fact that the phenomenal world, while truly internal and shut-in within the physical brain, nevertheless accurately reflects certain geometrical aspects of the external world, which is thereby knowable indirectly through its perceptual replica. It accounts for the fact that different individuals each have their own unique perspective on a commonly viewed object by the fact that each individual percipient has his own perceptual copy of the commonly viewed object. And it does away with the incredible proliferation of infinite sets of different perspectives for every object in the world, as well as with notions of phenomenal sense data which are experienced but which do not or may not actually exist. Bertrand Russell's epistemological dualism and causal theory of perception should therefore have resolved the epistemological question once and for all. But curiously it did not, and the reason why it has failed to do so is almost as interesting and significant as the epistemological question itself.

The epistemological debate highlights the very powerful human inclination to favor a naive realist view. After all we are all born naive realists, and only a few in each generation ever come to see through the grand illusion of conscious experience. Russell's causal theory of perception has never been refuted, and yet it continues to be simply ignored or misunderstood, although each of the alternatives to epistemological dualism have been repeatedly shown to be fatally flawed. The chief problem is that indirect realism seems so incredible that it is most often not even considered as a serious alternative. As Russell himself said, "*Perhaps there is nothing so difficult for the imagination as to teach it to feel about space as modern science compels us to think ... This question is very important, and must be understood if metaphysics is ever to be got straight. The traditional dualism of mind and matter, which I regard as mistaken, is intimately connected with confusions on this point. So long as we adhere to the conventional notions of mind and matter, we are condemned to a view of perception which is miraculous. We suppose that a physical process starts from a visible object, travels to the eye, there changes into another physical process, causes yet another physical process in the optic nerve, finally produces some effect in the brain, simultaneously with which we see the object from which the process started, the seeing being something 'mental', totally different in character from the physical processes which precede and accompany it. This view is so queer that metaphysicians have invented all sorts of theories designed to substitute something less incredible. But nobody notices an elementary confusion*". (Russell 1927, p. 137 - 143)

Lovejoy's (1930 p. 227-249) response to Russell's epistemological dualism is typical. Lovejoy does not argue on logical grounds that percepts cannot be in our heads, but argues instead that there are alternative explanations which, however improbable, at least leave the door of doubt open a tiny crack. One such possibility is a projection theory, that the patterns of electrochemical activity in our brain get somehow projected back out into the world. Another possibility is that although perceived objects are not in physical space, "*it does not follow that they are in our heads; they might ... be neither in our heads nor where the 'scientific objects' are, but in some other situation in physical space.*" (Lovejoy 1930 p. 228) His third argument is that it has never been proven that `being known' is necessarily equivalent to `being in our heads'; and finally Lovejoy argues that the question whether percepts are in our heads is not the

same as the question whether perceiving and awareness are physical processes, and that he rejects the former but accepts the latter. What is curious about these arguments is that Lovejoy feels no need to commit to any one of his proposed alternatives. Lovejoy does not profess a projection theory, but merely argues that it is not self-evident that it is untrue. If percepts are not in our heads, Lovejoy does not propose where else they might be. He does not explain how anything can be known that is not explicitly represented in our physical brain, nor how the physical processes underlying perception and consciousness could be anywhere other than in our brain. This therefore is not a refutation of Russell's causal theory of perception, but merely an expression of Lovejoy's opinion that Russell's theory seems so incredible to him, as to be on a par with those other incredible hypotheses.

Contemporaneous with Bertrand Russell in philosophy, the Gestalt movement in psychology was arriving independently at the same epistemological conclusion. Although initially Gestalt theory began as a theory of perception, Wertheimer clearly recognized the epistemological issue at the heart of the matter, and Wolfgang Köhler and Kurt Koffka motivated their presentation of Gestalt theory explicitly on epistemological dualist grounds. Curiously the Gestaltists made no reference to Kant as the originator of this idea, possibly because of the confusion caused by his idealist position, nor to Bertrand Russell's parallel arguments in the philosophical world. One of the most controversial and pivotal aspects of Gestalt theory is Wertheimer's principle of isomorphism, elaborated by Köhler (1924) as the hypothesis that every perceptual experience is "*not only blindly coupled to its corresponding physiological processes, but is akin to it in essential structural properties.*" This theory is a direct consequence of the indirect realist foundations of Gestalt theory, whereby phenomenal experience is a direct manifestation of neurophysiological processes in our physical brain, and therefore it cannot help but be similar in structure, since they are identical in ontology.

But despite the clear and cogent arguments made by the Gestaltists in defense of indirect realism, this idea was never widely accepted, although it was rarely challenged directly. It was however challenged by J. J. Gibson, who is one of the few people to explicitly defend a naïve realist view of perception. But Gibson's challenge did more to highlight the shortcomings of direct realism than of representationalism, because Gibson was forced to make all kinds of implausible assumptions about the perceptual process in defense of his naïve realist position. Gibson denied, for example, that the retinal image is anything like an image, and he denied the general materialist view that the sensory organs transmit sensory information into the brain, where neurophysiological processes compute a perceptual representation of the external world. Instead Gibson suggested that perception occurs somehow out in the world itself, rather than in the physical brain. Exactly how this occurs, or what this actually means however, he could never explain to any satisfaction. Once again Gibson demonstrates how a confused epistemology leads to a confused psychology. Significantly, neither Gibson nor anybody else has ever shown indirect realism to be untenable, only that it seems (to them) incredible. But despite the absence of challenge to the Gestalt arguments for indirect perception, and despite Gibson's failure to provide a plausible direct perception alternative, the issue has simply dropped out of the collective consciousness, and naïve realism once again by default rules the day in psychology and philosophy.

Naive Realism in Contemporary Psychology

Modern critics of Gestalt theory do not challenge the epistemological argument of indirect realism, but rather they focus their criticism on specific details of Gestalt theory, although careful examination reveals that their criticism is only meaningful when viewed from a naïve realist perspective. This is a classic case of a paradigmatic debate where the opposing camps argue at cross purposes, due to different

foundational assumptions. I present for illustration a prominent modern vision scientist, Richard Gregory, a man who has written many insightful books and papers on a wide variety of subjects in visual perception. However Gregory also reveals a fundamental naïve realism, of which he is almost certainly unaware, and I suspect that Gregory would be deeply offended at being accused of naïve realism. However Gregory's views on the critical epistemological issues are representative of the prevailing consensus in contemporary psychology, and therefore Gregory's arguments are illustrative of the common errors of contemporary psychology.

In his discussion of Immanuel Kant, Gregory (1981 p. 335) rejects epistemological dualism but makes no mention of the deep logical problems in the monist alternatives that forced Kant to this incredible conclusion. Gregory challenges Kant's argument that perception involves an *a priori* understanding of the nature of space, i.e. that it is three-dimensional and Euclidean, by pointing out that we *can* conceive of non-Euclidean spaces as seen for example in Riemannian geometry. However the only way we can conceive of a Riemannian space is as a warped or distorted Euclidean space. If we could not conceive of Euclidean space, we would have great difficulty conceiving a Riemannian space, let alone of the principles of Euclidean geometry. Gregory misses the more fundamental point that perception involves a representation, and a representation has a pre-ordained dimensionality. For example the retina is a representation, and it has two dimensions, so it is impossible for the retina itself to encode a three-dimensional image. The cortex is also a representation, but it too is limited, this time to representing three spatial dimensions (and time). Even if there were four-dimensional objects to be found in our world, we would never be able to perceive them as four-dimensional, all we would ever see is the peculiar morphing of a three-dimensional shape.This is not to deny that the external world also has (at least) three spatial dimensions, but those dimensions would be completely invisible to us were we not also equipped with a three-dimensional perceptual representation. Kant was right therefore, that the dimensionality of the phenomenal world is a property first and foremost of the representational mechanism of the brain, and only in secondary fashion is it also representative of some of the dimensions of external reality.

As to the Gestalt principle of isomorphism Gregory (1981, p. 368) finds "*a deep logical problem here concerning identity and analogy.*" According to Gestalt theory, when we experience a continuous spatial structure in phenomenal experience, that percept is a direct manifestation of some continuous spatial structure in the brain. Even if this were true, Gregory asks how are these perceptual structures themselves conscious? How can a physical process in the brain somehow become conscious of its own spatial structure? Gestalt theory invokes emergent processes such as the soap bubble, or electric fields, as the principle behind the spatial structures observed in perception. But Gregory objects that if the structures in the brain corresponding to perceptual experiences are somehow conscious of themselves, then similar emergent structures such as soap bubbles and electric fields, must also necessarily be somehow conscious. This conclusion is to Gregory so absurd as to preclude its antecedent. But if Gregory denies the possibility of physical processes ever becoming conscious, that precludes any kind of materialist or scientific explanation of consciousness in the brain. For if we take seriously the materialist thesis that mind is a physical process taking place in the physical mechanism of the brain, then that by itself is already an acceptance of the fact that a physical process taking place in a physical mechanism can under certain circumstances be conscious. Furthermore, even if Gregory's objection were valid, it would necessarily apply also to non-spatial or symbolic representations in the brain. But Gregory raises this objection only against spatial or perceptual structures, thus revealing that his objection is not to representations in the brain as such, or their ability to become conscious, but only to spatial

representations.

Gregory argues that it is absurd to believe that a part of your brain becomes colored when you perceive a colored surface, or that the percept of a complex three-dimensional object like a house is represented by a three-dimensional model house in the brain. This too is a common criticism of the principle of isomorphism, but it is only intelligible from a naïve realist perspective. For from the indirect realist perspective it is even more absurd to think that the color you are seeing is the external color itself, entering consciousness directly, somehow bypassing the perceptual representations in the brain. The indirect realist view recognizes that the perceived color is not a property of physical light itself, but only of its perceptual replica in the brain. We have no idea what physical color is really like, we can only know how a color is represented in the brain, which by definition is identically equal to the way we see it in phenomenal experience. It is true that the phenomenal experience of red does not mean that the corresponding portion of our brain would be observed to be red under microscopic examination, nor that it would appear red to a micro-electrode implanted in that portion of the brain. But that does not make it any the less red when experienced phenomenally, or any the less an intrinsic property of the physical mechanism of the brain. Similarly, if we deny that the brain is capable of fabricating a three-dimensional spatial replica of a house, then by definition we could never have the spatial experience of such a house. For the spatial structure of the house cannot enter consciousness directly, only by way of its explicit representation in the physical brain. Gregory (1981, p. 370) argues that there are many different ways to represent information; for example the number six can be expressed as the digit "6" or the word "six", or by the presence of six dots as on dice. "*Why then*" asks Gregory "*did the Gestalt psychologists choose isomorphic representation? This is only one of an infinite set of kinds of representation.*" The answer, Dr. Gregory, is to be found by inspection, for the representation employed in the brain is exactly as it appears in conscious experience.

Richard Gregory's comments on the problem of the inverted retinal image are also representative of the prevailing view in contemporary psychology. Ever since Johannes Kepler identified the retina as the sensory organ of the eye onto which the lens projects an inverted image, the question arose why we do not see the visual world as inverted. A number of philosophers have weighed in on this issue over the centuries, but to this day the question remains unresolved, although very few are even aware of it, for this too is a manifestation of indirect realism in perception that is inconsistent with naive realism. Gregory's view, which again is representative of the prevailing view in psychology, is that the issue is a pseudoproblem which is actually no problem at all. Gregory argues that " *it does not in the least matter that [the retinal images] are upside-down with respect to the objects that they image - for there are, so to say, no eyes looking at the retinal images*". In other words Gregory raises the oft-refuted "homunculus objection" (of which more below) to argue that there are no spatial representations in the brain, and non-spatial or symbolic representations have no particular orientation. While this argument might seem plausible to a naïve realist who identifies the spatial world of conscious experience with the external world itself, even a naïve realist must recognize that a retinal after-image, seen after exposure to a bright light or camera flash, is a phenomenon internal to the eye rather than out in the world. And the spatial structure of the retinal after-image is clearly evident, and no internal homunculus is required to view it. Furthermore, the retinal after-image is clearly erect relative to the rest of the perceived world, when viewed "internally" by phenomenological observation, although the corresponding patch of over-exposed retina is obviously inverted by the lens in the eye, relative to the external world. The inescapable conclusion is that the retinal image is in register with our perceptual replica of the external world, both of which are inverted relative to the external world itself. So the reason why we do not see the world

inverted is not because there are no spatial representations in the brain, but because we cannot see any trace of the external world directly, and therefore we cannot notice that it is inverted relative to the phenomenal world that we do see.

What is most interesting about Gregory's viewpoint is the fact that a man with such extensive knowledge of the physiology, psychology, and phenomenology of vision, does not seem to fully comprehend the position he contests, so his criticisms inevitably miss their intended targets. It is equally interesting that Gregory never strikes at the core issue hidden behind the debate, i.e. the indirect nature of perception. Nor would Gregory explicitly defend naive realism, a position which he would almost certainly reject, although that philosophy unwittingly pervades his arguments on the other peripheral issues. This schizophrenic attitude intermediate between direct and indirect realism underlies much of the confusion over the core issues of perception and consciousness in the contemporary literature.

While the epistemological question is generally not discussed directly in modern psychology, occasionally the issue comes to a head in particular subject domains. For example the phenomenon of hemi-neglect (McFie & Zangwill 1960, Heilman & Watson 1977, Heilman et al. 1985) initially caused a great stir in psychological circles because it appeared to be concrete evidence for an explicit spatial representation in the brain (Denny-Brown & Chambers 1958, de Renzi 1982, Bisiach & Luzzatti 1978, Bisiach *et al.* 1981). It is curious that half of phenomenal space should have to disappear for psychologists to take account of its existence in the first place! But after the initial excitement, the naïve realists quickly marshalled their defenses with an array of arguments which many believe to dispose of the troublesome issue of hemi- neglect. Some argue that hemi-neglect is not a failure of spatial representation, but rather an imbalance of attention, or `orienting response', i.e. that half of phenomenal space does not actually disappear, but that the neglect patient is merely inclined to ignore its presence. (Heilman & Watson 1977, Heilman *et al.* 1985, Kinsbourne 1987, 1993) But even if these arguments are valid, they do not account for the presence in visual consciousness of the spatial structure of the phenomenal world whenever it is *not* being ignored or neglected; they merely offer a convenient escape clause to make neglect syndrome seem no *more* mysterious than normal spatial perception. Others argue that the phenomenon of hemi-neglect fractionates to a number of distinct patterns of impairment (Vallar 1998 p. 88). For example many neglect patients can describe the global gestalt of a figure, but when copying its local features, leave those on the left side out (Marshall & Halligan 1995). Present accounts of the multiple forms of neglect refer to several spatial maps and their interaction (e.g. Ladavas et al. 1997). This highlights a conflict between the phenomenal and neurophysiological evidence, the former presenting a unified spatial structure in visual experience, while the latter suggests discrete mechanisms in different cortical areas. To the naive realist this suggests that the spatial percept must be somehow illusory, which thereby supposedly relieves neuroscience from any obligation to account for its manifest properties. What is curious about the debate over neglect is the passion that it engenders. The evidence presented by each side never seems to convince the opposition, because the debate is not really about neglect, but about its implications for perceptual representation, and that issue is not so much a matter of experimental evidence but of the interpretation of that evidence, or the foundational assumptions with which one comes to the debate in the first place.

A similar paradigmatic question underlies the contemporary debate on the nature of mental imagery. Ever since the behaviorist movement mental imagery had been strictly off-limits to psychology until the cognitive revolution revived interest in mental phenomena. A number of ingenious psychophysical experiments have since provided quantitative evidence in support of mental imagery (Kosslyn *et al.* 1995). For example Shepard and Metzler (1971) show that the time required to rotate a mental image is

in linear relation to the angle through which it is rotated. Stephen Kosslyn and his coworkers established that the time it takes to scan between two points on the mental image of a memorized map is in direct proportion to the distance between the points, just as with a real map. (Kosslyn *et al.* 1978) In addition, it takes subjects approximately the same amount of time to scan a real map as it takes them to scan their visual image of a map memorized from a verbal description (Denis & Cocude 1989). Kosslyn (1975) showed that mental images take a finite time to zoom up or down to a different size, as when comparing the mental image of an elephant immediately following an image of a rabbit. As incredible as these properties might seem in the context of contemporary concepts of neurocomputation, these properties are perfectly consistent with the subjective experience of mental imagery as observed phenomenologically. Visual pathologies generate further striking evidence for the existence of mental images. Bisiach and Luzzatti (1978) and Bisiach *et al.* (1979) showed that patients suffering from hemispatial neglect sometimes exhibit a parallel neglect of one side of their imaged visual field. Llinás and Paré (1991) observe that neglect patients report a similar lack of perception in their dreams, and that the people inhabiting the dreams of prosopagnosic subjects are faceless. Levine *et al.* (1985) found that patients who had lost the ability to perceive either shape or location, had corresponding difficulties in mental imagery tasks. Mellet *et al.* (2000) show that imagery based on verbal descriptions can recruit cortical regions known to be engaged in high-order visual processing. As with the debate over neglect syndrome, the imagery debate generates more heat than light, as the opposing camps never seem to be persuaded by the evidence presented by the other side, due to the paradigmatic issue hidden at the core of the debate which remains unaddressed. For if the subjective experience of visual consciousness is considered to be a valid source of evidence for the nature of the representation in the brain, then the existence and spatial nature of mental imagery can be confirmed by inspection, making all that psychophysical verification actually unnecessary.

Ned Block (1981*a*) provides an excellent summary of the imagery debate, with chapters contributed by both supporters and opponents of mental imagery. The arguments rallied by the opponents of mental imagery clearly reveal their underlying naive realist assumptions. Ned block himself (1981*a*, p. 2) and Daniel Dennett (1981*a* p. 53) raise the oft-refuted homunculus objection, of which more below. Block also raises the neurophysiological objection, that "*brain scientists have found no pictures in the brain*" revealing that he subscribes to the belief that neurophysiological evidence (or the absence of it) trumps phenomenal evidence, i.e. that we cannot be sure that our phenomenal experience is "real" until it is confirmed neurophysiologically. Jerry Fodor (1981*a* p. 63) simply announces that he considers the idea incredible: "*I am, in fact, strongly inclined to doubt the very intelligibility of the suggestion that there is a stage at which cognitive processes are carried out in a medium that is fundamentally nondiscursive*" and leaves it at that. Robert Schwartz (1981*a* p. 122) complains that the principle of isomorphism between mental imagery and brain processes is "*vacuous*" because "*the notion of resemblance is too vague*". Zenon Pylyshyn (1981*a* p. 153) provides the most blatant naive realist argument that what people report as properties of mental images, for example their spatial extension, are properties of the objects they represent, not of the images themselves. Pylyshyn argues as if the quality of spatial extent can somehow bypass the representational mechanism and penetrate directly into the experience of the imaging subject, without leaving an impression in the subject's brain.

The naive basis of Pylyshyn's concept of visual processing is made abundantly clear in his "Theory of Visual Indexes" (Pylyshyn 1998). According to that theory, an early preattentive stage in visual perception involves a resource-limited mechanism for individuating a small number (four to six) of visual tokens in the stimulus. These tokens are then indexed, like books in a library, and the index

remains attached to its object as the object changes its retinal location or other properties. So far, this description is not objectionable, and is consistent with both direct and indirect realism. But where is the sensory image in which these tokens are tracked? Is it registered in an image-like sense-data array? Not according to Pylyshyn, who speaks as if that "image" is out in the world itself, for, Pylyshyn argues, the external world can be accessed as if it were an internal memory, as suggested by O'Regan (1992). Pylyshyn writes (p. 217) "*it may no longer be necessary for the perceptual representation itself to have metrical or pictorial properties.*" Of course since the phenomenal world clearly *does* have both metrical and pictorial properties, Pylyshyn must be confusing the phenomenal world for the external world. The principal argument that Pylyshyn (1981*b*, 1988) raises against mental imagery is that much of mental imagery is "cognitively penetrable", i.e. it can be influenced by cognitive considerations. Pylyshyn argues that "*a process that is sensitive to the logical contents of beliefs must itself contain at least some inferential ... rule-governed process.*" But this argument hangs on the assumption that inferential rule-governed processes cannot be analogical, they can only be symbolic and non-spatial. I will show later how computational devices can be designed to incorporate inferential rule-governed processes implemented as analog field-like forces, which are in fact "cognitively penetrable", i.e. they can be influenced or modulated by manipulating a few "cognitive variables". The mental image medium *does* change its properties to correspond to what subjects believe about the world, as Pylyshyn himself points out, with mental imagery tasks such as imagining the color produced by overlapping yellow and blue color filters, or imagining dropping two objects of different weights. The outcome of these mental images will depend on one's knowledge of color mixing theory, and the theory of gravity respectively, so a physicist's mental image will naturally be more physically correct than that of a non-scientist. This, according to Pylyshyn, demonstrates that the answer cannot be produced by imagery mechanisms, but must involve non-spatial symbolic operations. But as is often the case with paradigmatic issues, this self-same evidence can be interpreted as demonstrating that mental imagery is both analogical, and it is "cognitively penetrable". In fact, the very act of willing a mental image into existence is itself a "cognitive penetration", for the mental image medium would be useless unless it can be controlled cognitively to depict whatever visual problem is required to be solved, when that solution is needed. Pylyshyn's definition of what can be validly considered to be mental imagery excludes mental imagery as it is observed.

Pylyshyn (1981*b*, 1988) complains that adding "free variables" such as voluntary control of the rate of rotation, translation, or scaling of mental imagery, amounts to "*ad hoc contrivances*" that detract from the theory's explanatory power. Pylyshyn insists that an analog model should be both "*principled and constrained*", by which he means that the rate of rotation etc. must be fixed. But Pylyshyn has some *ad hoc* explanations of his own. His own explanation for the linear time-to-rotate data for mental imagery (Shepard & Metzler 1971) is that the subjects know that rotation is constant in the real world (is it?) and therefore they "*make it be the case*", i.e. they hold back their keypress response until the mental image *would have* stopped rotating, *if* they were using a mental image, except that they *arn't*. This is about as *ad hoc* an explanation as any I have heard! As to the model remaining "principled and constrained", we must not lose sight of the fact that the *explanandum* in this case is the cognitive mechanism of the human mind, an extraordinarily complex and sophisticated function, and clearly far more "cognitively penetrable" than the lower level perceptual function. A model of mental imagery that is arbitrarily constrained to fixed rates of rotation and scaling etc. is guaranteed from the outset to fall far short of an adequate model of mental imagery. To the contrary - what is required are more sophisticated and complex models with a great deal of cognitive control to account for this most complex and malleable mental phenomenon.

But perhaps the most surprising aspect of the imagery debate is the fact that even the proponents of mental imagery seem to hold back from a full acceptance of mind-brain identity, and the validity of phenomenological observation of neurophysiological entities. In the words of Ned Block (1981*a*, p. 2) "*no one writing in this book (nor any other serious participants in the debate) thinks that people can literally see and manipulate real internal pictures.*" This opinion is echoed by perhaps the most ardent proponent of mental imagery, Stephen Kosslyn (1981*a*, p. 207) "*Although no serious researcher today maintains that images are actual pictures in the head, some still find it reasonable to posit quasi-pictorial representations that are supported by a medium that mimics a coordinate space.*" I find this statement puzzling, since the notion of quasi-pictorial representations supported by a medium that mimics a coordinate space is exactly what I consider to be "actual pictures in the head". It seems that even the proponents of mental imagery have not entirely abandoned all vestiges of their native naive realism. Daniel Dennett (1981*a*, p. 88) makes the insightful observation that although this issue remains unresolved, and that nobody really knows with any certainty whether there are pictures in the brain, "*A curious feature of the debate is the passion it evokes, which is unlike the normal passion of scientific controversy ... everyone, it seems, has a fiercely confident opinion about the nature and existence of mental images. This manifests itself in remarkable ways: in unhesitating predictions of the results of novel psychological experiments, in flat disbelief in the integrity of recalcitrant experiments, in gleeful citation of 'supporting' experimental evidence, coupled with bland imperviousness to contrary evidence.*" In their summary of the debate to date, Kosslyn *et al.* (1981*a* p. 132) offer the observation that... "*Not surprisingly neither arguments nor counter-arguments have been definitive, and neither seems to have had enough force to sway most people from whatever position they found most congenial in the first place.*" This is a sure sign of a paradigmatic debate!

Naive Realism in Contemporary Philosophy

Of all the branches of human knowledge, philosophy might be expected to be the best inoculated against the naïve realist error, since the issue of the epistemology of conscious experience is a central focus of philosophy. However modern philosophy is just as rife with naïve realists as are modern psychology and neuroscience. As in psychology there is a recurring pattern of the occasional visionary who points out the fallacy of the naïve view, interspersed with long periods of enthusiastic support for the latest naïve inspired view, although again the issue is generally not addressed directly but only peripherally, as it is hidden in the details of various theories. Jaegwon Kim (1998) traces the origins of the mind-body identity theory to C. D. Broad (1925) and Herbert Feigl (1958). Broad, according to Kim, was the original exponent of this theory, although Feigl was more influential. But a careful analysis of Broad's theory reveals the residual naïve realism of his critical realist philosophy. Broad argues that our mind is the object of our introspection, just as physical things are the objects of our perceptions. However Broad does not recognize the perception of external objects to be introspection, but `inspection' (i.e. exteroception). He denies that `sense data' such as the perceived shape and color of a perceived object can be counted as states of our mind, because "*if this were so, the colour could not pervade the external place*" (Broad 1925 p. 177). This confusion of external and internal entities lead to endless confusion in Broad's concept of perception.

Feigl (1958) on the other hand does present a cogent case for the identity of mind and brain in its entirety, and Feigel even cites Bertrand Russell and Wolfgang Köhler among those who preceded him in this belief. On isomorphism Feigl explains (1958, p. 79) "*the states of direct experience - are identical with certain (presumably configurational) aspects of neural processes - [so that] what is had in*

experience - is identical with - what the science of neurophysiology describes - as processes in the central nervous system, perhaps especially in the cerebral cortex." He then systematically lists and refutes all of the common objections to mind-brain identity. Feigl concludes (1958 p. 41) that "*our arguments have -disproved- the Cartesian contention that the mental is non-spatial. To put it very strongly, mental events as directly experienced and phenomenally described* ***are*** *spatial.*" Similar compelling arguments are presented by Boring (1933) and Smart (1959). This then should have put the final nail in the coffin of naïve realism. But it was not long before identity theory was attacked on two fronts, which Kim (1998) calls the *multiple realizations* argument (Putnam 1968, 1975), and the *anomalous monist* argument (Davidson 1970, 1980).

Putnam objects that mental states can and do have vastly diverse physical or biological realizations in different species and structures. For example both humans and mollusks presumably experience pain, but that experience is mediated by entirely different neural mechanisms. Therefore no mental state can be identified with any single physical or biological state. Turning the argument the other way around, therefore seems to suggest that if pain that we experience as humans could potentially be implemented in a number of different ways, then there is little we can say about the neurophysiological correlate of our experienced pain. This supposedly absolves the philosopher of the need to worry about neurophysiology at all, and permits us to treat the phenomenal and the physiological as two separate domains. In fact Putnam claims that mental kinds and properties are functional kinds of a higher level of abstraction than physicochemical or biological kinds, and that therefore mental or cognitive properties are a distinctive domain to be investigated independently of their physical / biological interpretations. That mental properties are realized or implemented in physical properties, although, Putnam insists, they are neither identical to, nor are they reducible to them. But although this argument may have some validity when applied to basic constituent qualia such as color and pain, it does not apply to the dimensionality of the perceptual representation or its information content, i.e. the manner in which those qualia are arranged or structured to represent external reality. For whatever the neurophysiological mechanism underlying spatial perception, the principle of isomorphism suggests that it must be a spatial mechanism of three spatial dimensions, because its spatial nature is clearly evident phenomenologically. Now there are multiple possible realizations of a spatial representation, so the mere fact of the spatiality of phenomenal experience cannot of itself select between different spatial realizations. However phenomenology can exclude the possibility of non-spatial representations to account for the spatial nature of perception. While this argument might seem controversial when applied to spatial perception, the same argument is widely accepted in the case of color perception. For color experience has been shown to encode three dimensions of color value, i.e. hue, intensity, and saturation. So although we cannot yet with certainty identify the neurophysiological mechanism responsible for encoding color experience, we can say with certainty that that mechanism must encode at least three dimensions of color information. Different realizations of color experience in different phyla and species may well correspond to different qualia of color experience, and the subjective aspect of those experiences in other species is perhaps forever unknowable to us (Nagel 1974, Chalmers 1995). However psychophysical experiments in humans and animals can and have determined the dimensions of color experience, and that in turn offers concrete information about color representation in the brain. There is no reason why this argument should not hold also for spatial experience.

Davidson's anomalous monist argument (Davidson 1970, 1980) suggests that the mental domain, on account of its essential anomalousness and normativity cannot be the object of serious scientific investigation, because the mental is on a wholly different plane from the physical. This argument sounds

like the ontological dualism of Descartes that disconnects mind from brain entirely, except that Davidson qualifies the previous statement with the monistic proviso that each mental event is connected with specific physical events (in the brain), although there are no laws connecting mental kinds with physical kinds. In other words there are no events which have only mental properties, and this rescues the thesis from ontological dualism. Kim (1998) points out however that this is a negative thesis, for it tells us only how the mental is *not* related to the physical, it says nothing about how they *are* related. As such this is more an article of faith rather than a real theory of any sort, and in the context of the history of the epistemological debate this can be seen as a last desperate attempt to rescue naïve realism from its own logical contradictions. This kind of physicalism has been appropriately dubbed `token physicalism', for it is indeed a token admission of the undeniable link between mind and the physical brain, without admitting to any of its very significant implications. It is Davidson's attempt to have his cake and eat it too. In order to rationalize this view of the mind-brain relation Davidson introduces the peculiar notion of *supervenience*, a one-way asymmetrical relation between mind and brain that makes mind dependent on the brain, but that forever closes the possibility of phenomenological observation of brain states. The notion of supervenience is so peculiar that mind and brain are supposedly the only example in the universe of entities joined by this paradoxical relation. However as Kim observes, this is not really a theory of mind- body relation for it is silent on the nature of the dependence relation that might explain *how* or *why* the mental supervenes on the physical. Mind-body supervenience *states* the mind-body problem, it is not a solution to it, for the central paradox of consciousness is preserved in the paradox of supervenience.

Dretske (1995) also defends Putnam's multiple realizations argument and Davidson's concept of supervenience, although Dretske explicitly declines to consider the epistemological dualist position even as a theoretical possibility to be refuted. (Dretske 1995, p. 127) "*I will have nothing to say to those who reject all forms of externalism ... I'm trying to win a battle, not the war.*" But the war is already lost, although Dretske has not yet heard the news, so winning the battle will turn out to be a futile gesture. The position Dretske defends is reminiscent of the confused epistemology of the critical realists, which makes one wonder whether Dretske has ever heard of Bertrand Russell's causal theory of perception. For example Dretske (1995, p. 44) describes introspection as "*an instance in which an experience (of blue, say) is conceptually represented as an experience of blue via a sensory representation not of the experience, but of some other object. One comes to know ... that one is experiencing blue by experiencing, not the experience of blue, but some displaced object ... i.e. the blue object one sees.*" So once again the blueness of the subjective experience is not a blueness in the percipient's mind, but is supposedly a property of the external object in which that blueness is perceived. Dretske takes on the issue raised by Lovejoy (1930) about whether `being known' is necessarily equivalent to `being in our heads', but he does so by way of an invalid analogy. Dretske argues (p. 38) that "*the mind isn't in the head any more than stories ... are in books.*" But what makes this analogy invalid is that the story printed in a book is incomplete. It does not even become a story of people and events until a human mind is present to read the patterns of ink on the page and transform them into an internal representation of people and events. A better analogy is that of an artificial intelligence or robot, whose representation of the external world does not require the intervention of a human intermediary. No matter how complex and intelligent such a robot might be built, there is no way that the sensory or perceptual information encoded by the robot can be considered to be located anywhere other than in the robot's "brain", or its explicit physical mechanism of representation.

Despite the deep and fundamental problems with these non-reductive physicalist theories, they have

been, and still are the most influential metaphysical position on the mind-body problem (Kim 1998). It is interesting how this fits in with the historical pattern of the epistemological debate. For neither Putnam nor Davidson nor Dretske propose to challenge the core issues of epistemology directly, but merely present a convenient escape clause by which philosophers can evade the issue altogether while supposedly preserving their scientific integrity. This is the same kind of evasion of the fearsome facts of epistemology as was offered by the dualist account that pushed the problem into the domain of God or Spirit, and the behaviorist solution that pushed the problem of consciousness off-limits to scientific scrutiny, and the critical realist solutions that invoke semi-existent spatial entities located in non-space. And as with those earlier movements in philosophy, all that was required was to open the door of doubt a tiny crack to unleash a stampede of popular support all desperately seeking a respectable justification for their own naïve realist intuitions. There are various other strategies to be found in contemporary philosophy to evade the epistemological issue, such as the modern dualism of *nonnaturalism* (Popper & Eccles 1977, Swinburne 1984, Adams 1987) which holds that consciousness is not a natural phenomenon and is therefore closed to scientific scrutiny; and *anticonstructive naturalism* (McGinn 1989) which suggests that consciousness is terminally mysterious, and all attempts to resolve the mind-body problem are doomed to failure from the outset; and *eliminative naturalism* (Churchland 1981, Churchland 1983) which suggests that consciousness is a natural phenomenon, but that there is a sense in which it cannot be explained, for the concept of consciousness is simultaneously too simplistic, too vague, and too historically embedded in false and confused theory to denote a phenomenon in need of explanation. But the deeply mysterious aspect of consciousness which motivates these pessimistic analyses is mostly involved in its confusion with the external world. Once we recognize that `external' consciousness is in fact an internal representation, and that it takes the form of a spatial structure in the brain, the mystery is transformed from a deep logical paradox to a neurophysiological or neurocomputational problem, i.e. the question of what kind of neurophysiological or computational principle could possibly account for the emergence of dynamic spatial structures in the brain.

Another somewhat different challenge to the mind-brain identity theory is posed by Smythies (1994, 1999). While Smythies explicitly refutes naïve realism (Smythies & Ramachandran 1998) and the resultant confusion of the phenomenal body or `body image' with the objective physical body, (Smythies 1953) Smythies argues that "*the brain, as a machine, is simply the wrong sort of machine to be able to actually construct the visual field and other components of phenomenal consciousness.*" (Smythies 1994, p. 311) Smythies cites Leibnitz's principle that for two entities to be identical, they must have identical properties. "*Since events in the sensory brain and events in our sensory fields in consciousness have clearly distinct properties, - the theory fails in principle.*" (Smythies 1999, p. 168) Here Smythies puts his finger on the principal motivation for naive realism in modern philosophy, which is the glaring disparity between phenomenology and contemporary neuroscience. For modern neuroscience tells us that the brain is composed of innumerable discrete neurons, interconnected in a network of synaptic connections. It is hard to resolve this discrete or quantized concept of brain physiology with the continuous, unitary, and field-like character observed in the phenomenal world, an issue which is sometimes called the "grain problem" (Wilfred Sellars 1963). This either means that consciousness is a complete illusion that bears no resemblance to its corresponding neurophysiological state, or that contemporary neuroscience is in a state of serious crisis, for it offers no evidence for the continuous field- like pictorial representations that we know to be present in the brain. It is not so surprising for a neuroscientist to favor the former eliminative alternative, having more inherent faith in his own method of investigation. It is surprising however that the modern philosopher most often defers to neuroscience whenever it is in conflict with the observed properties of the mind, for it is the mind, and not the brain,

which is, or at least should be the primary object of philosophical inquiry. A philosopher who cannot trust his observations of the properties of the mind unless or until they are confirmed by neurophysiology, would do better to abandon philosophy altogether as a futile enterprise, and switch to the more certain science of neurophysiology. Smythies' own solution is to propose that consciousness may be concealed in one of the hidden dimensions of reality which are sometimes proposed in modern cosmology. In some sense this is reminiscent of the semi-existent entities proposed by the critical realists for sense-data, which are spatial structures but not to be found in physical space. However by identifying the hidden dimensions of physical reality as the locus of these hidden percepts, Smythies moves the theory into scientific territory where those hidden dimensions should at least in principle be accessible to scientific scrutiny. But until modern science can actually confirm the existence of those hidden dimensions, and demonstrate how information can be stored in, and retrieved from them by physical processes, this is a speculative hypothesis that remains to be confirmed. Unlike the critical realists however, Smythies would presumably allow that the properties of those hidden dimensions of the universe are accessible phenomenologically, so Smythies' theory is an identity theory, that places sense data within the physical brain, albeit in a hidden dimension not readily accessible to scientific scrutiny except by way of conscious experience.

The continued dominance of the naive realist view in contemporary philosophy is highlighted by the fact that books like Tye's (1995) "Ten Problems of Consciousness" pass largely unchallenged, even though several of the fundamental problems of consciousness identified by Tye disappear altogether when viewed from the indirect realist perspective. For example Tye's problem number eight, the *Problem of Transparency*, is the same issue raised by Searle (1992), that we cannot distinguish consciousness itself from the object of which we are conscious. But consciousness is only transparent to those who fall prey to the naive illusion, and believe they are viewing the world itself directly, as if by magic. Tye's problem number nine, the *Problem of Felt Location*, is the question of external perception, which is also resolved by the dualist epistemology which reveals that the space of our phenomenal experience is a representation inside our head, and the paradox disappears. While indirect realism does not resolve all of Tye's ten problems as easily as it does these two, it does cast the problem in an entirely new light, whose implicatons deserve at least as much scrutiny as the naive view has been afforded, to see if this unexplored alternative might finally release the problem of consciousness from its current paradoxical impasse. But what is interesting in this case is that Tye does not consider indirect realism even as an alternative to be discussed and refuted, it is simply ignored altogether as if it had no place in the debate.

Flanagan's (1992) *Consciousness Reconsidered* comes closer to expressing the true mind-brain relation, for Flanagan argues that mind is a natural phenomenon that can be investigated by blending insights from neuroscience, psychology, and phenomenology. Thus it would seem that Flanagan's "constructive naturalism" is a kind of mind-brain identity theory. But Flanagan provides subtle clues that he still clings to a few last vestiges of naive realism. Although Flanagan approves of the investigation of brain processes through phenomenology, he also claims (p. 12) that "*Phenomenology alone has been tried and tested. It does not work. ... Phenomenology alone never reveals anything about how `seemings' are realized, nor can it reveal anything about the mental events and processes involved in conscious mental life.*" But phenomenology has already revealed that perception involves a three-dimensional volumetric spatial representation, and that is a very significant fact of conscious mental life, with direct implications for how spatial "seemings" are realized in the brain. The only way that this most obvious representational fact could possibly have escaped Flanagan's notice was if he mistook this phenomenal world for reality itself. Significantly Flanagan never discusses the epistemological issue itself directly, or its very

significant implications for all of the other issues of perception and consciousness, as if Flanagan were totally unaware of the existence of this theoretical alternative.

If there is anything to be learned from the long history of the epistemological debate, it is that the issue is by no means simple or trivial, and that whatever is ultimately determined to be the truth of epistemology, we can be sure that it will do considerable violence to our common-sense view of things. This however is nothing new in science, for many of the greatest discoveries of science seemed initially to be so incredible that it took decades or even centuries before they were generally accepted. But accepted they were, eventually, and the reason why they were accepted was not because they had become any less incredible. In science, irrefutable evidence triumphs over incredibility, and this is exactly what gives science the power to discover unexpected or incredible truth. Ultimately, therefore, the most convincing argument for epistemological dualism is the fact that its monistic alternatives have all been refuted on sound logical grounds, which leaves epistemological dualism as the only viable alternative. Until this most basic fact of conscious experience finally triumphs over our naive realist inclinations, philosophy and psychology are doomed to an endless and futile recapitulation of the ancient epistemological debate.

The Dimensions of Conscious Experience

Once we accept the fact that the world of visual consciousness is a pattern of energy in our physical brain, we can begin to examine that conscious experience to see what it might tell us about its neurophysiological correlate. The practice of phenomenology for investigating mental function was more popular before modern neuroscience introduced a new concept of neurocomputation that seems inconsistent with phenomenological observation. (Vernon 1937, 1952, Gregory 1981, Ramachandran & Blakeslee 1998, Smythies 1953, 1988, 1994, 1999, Koffka 1935, Köhler 1924) The most basic and salient fact of visual consciousness is that it appears as a three-dimensional spatial structure (Vernon 1952, p. 81- 92). More specifically, the phenomenal world is composed of solid volumes, bounded by colored surfaces, embedded in a spatial void. Every point on every visible surface is perceived at an explicit spatial location in three-dimensions (Clark 1993), and all of the visible points on a perceived object like a cube or a sphere, or this page, are perceived simultaneously in the form of continuous surfaces in depth. The perception of multiple transparent surfaces, as well as the experience of empty space between the observer and a visible surface, reveals that multiple depth values can be perceived at any spatial location. The information content of perception can therefore be characterized as a three-dimensional volumetric data structure in which every point can encode either the experience of transparency, or the experience of a perceived color at that location. Since perceived color is expressed in the three dimensions of hue, intensity, and saturation, the perceived world can be expressed as a six-dimensional manifold (Clark 1993), with three spatial and three color dimensions.

The Cartesian Theatre and the Homunculus Problem

This "picture-in-the-head" or "Cartesian theatre" concept of visual representation has been criticized on the grounds that there would have to be a miniature observer to view this miniature internal scene, resulting in an infinite regress of observers within observers. However this argument is invalid, for there is no need for an internal observer of the scene, since the internal representation is simply a data structure like any other data in a computer, except that this data is expressed in spatial form. If the existence of a spatial data structure required a homunculus to view it, the same objection would also apply to symbolic or verbal information in the brain, i.e. *epistemic* as opposed to *sensory* perception, which would also

require a homunculus to read or interpret that data. In fact any information encoded in the brain needs only to be available to other internal processes rather than to a miniature copy of the whole brain. To deny the spatial nature of the perceptual representation is to deny the spatial nature so clearly evident in the world we perceive around us. To paraphrase Descartes, it is not only the existence of myself that is verified by the fact that I think, but when I experience the vivid spatial presence of objects in the phenomenal world, those objects are certain to exist, at least in the form of a subjective experience, with properties as I experience them to have, i.e. location, spatial extension, color, and shape. I think them, therefore they exist. All that remains uncertain is whether those percepts exist also as objective external objects as well as internal perceptual ones, and whether their perceived properties correspond to objective properties. But their existence in my internal perceptual world is beyond question if I experience them, even if only as a hallucination.

The Neuroreductionist Objection

A number of theorists have proposed (Dennett 1991, 1992, O'Regan 1992, Pessoa *et al.* 1998) that consciousness is an illusion, and that in fact the conscious experience is considerably more impoverished than it appears subjectively. For example the loss of resolution in peripheral vision is not immediately apparent to the naïve observer. However the objective of phenomenology is not to quantify the casual experience of the naïve observer, but the careful observation of the critical observer. For the loss of acuity in peripheral vision is plainly evident under phenomenological observation, and can be easily verified psychophysically, and therefore this should also be reflected in the perceptual model. Dennett argues that visual information need not be encoded explicitly in the brain, but merely implicitly in some kind of compressed representation. For example the percept of a surface with uniform color could be abbreviated to a kind of edge image, with a single value to encode the color of the whole surface, as is the practice in image compression algorithms. This notion appears to be supported by neurophysiological studies of the retina which show that ganglion cells respond only to spatial or temporal discontinuities of the brightness profile, with no response within regions of uniform color or brightness. Dennett argues that the experience of a filled-in field of color in uniform fields, and in the blind spot, does not suggest an explicit filling-in mechanism in the brain, but that the color experience is encoded by "ignoring an absence" (Dennett 1991,1992). However an absence can only be ignored from a representation that already contains something in the place of the ignored item, otherwise one would experience nothing at all, rather than a spatially continuous field of color. In fact the experience of the retinal blind spot, or a uniformly colored surface, produces a distinct colored experience at every point throughout the colored region to a particular spatial resolution as a spatial continuum, and the informational content of that experience is greater than that in a compressed representation. If it is true that the retinal image encodes only brightness transitions at visual boundaries, then some other mechanism higher up in the processing stream must perform an explicit filling-in to account for the subjective experience of the filled-in surface. In fact the many illusory filling-in phenomena such as the Kanizsa illusion implicate exactly this kind of mechanism in perception. If visual information were indeed expressed in a compressed neurophysiological code, then our subjective experience of that information would have to also be correspondingly compressed or abstracted, as is the case for example with an experience of a remembered or imagined scene. The fact that our phenomenal experience is of a filled-in volumetric world is direct and concrete evidence for a volumetric filling-in mechanism in the brain.

An Analogical Paradigm of Representation

Once we recognize the world of experience for what it really is, it becomes clearly evident that the representational strategy used by the brain is an *analogical* one. In other words, objects and surfaces are represented in the brain not by an abstract symbolic code, as suggested in the propositional paradigm, nor are they encoded by the activation of individual cells or groups of cells representing particular features detected in the scene, as suggested in the neural network or feature detection paradigm. Instead, objects are represented in the brain by constructing full spatial effigies of them that appear to us for all the world like the objects themselves- or at least so it seems to us only because we have never seen those objects in their raw form, but only through our perceptual representations of them. Indeed the only reason why this very obvious fact of perception has been so often overlooked is because the illusion is so compelling that we tend to mistake the world of perception for the real world of which it is merely a copy. This is a classic case of not seeing the forest for the trees, for the evidence for the nature of perceptual representation in the brain has been right before us all along, cleverly disguised as objects and surfaces in a virtual world that we take to be reality. So for example when I stand before a table, the light reflected from that table into my eye produces an image on my retina, but my conscious experience of that table is not of a flat two-dimensional image, but rather my brain fabricates a three-dimensional replica of that table carefully tailored to exactly match the retinal image, and presents that replica in an internal perceptual space that includes a model of my environment around me, and a copy of my own body at the center of that environment. The model table is located in the same relation to the model of my body as the real table is to my real body in external space. The perception or consciousness of the table therefore is identically equal to the appearance of the effigy of the table in my perceptual representation, and the experience of that internal effigy is the closest I can ever come to having the experience of the physical table itself.

There is ample evidence suggestive of an analogical representation at least in the function of mental imagery. Kosslyn (1995) lists four computational functions of mental imagery, which are the ability to 1: generate images, 2: interpret the shapes in the images, 3: retain the image over time, and 4: to transform the image in some way. Pinker *et al.* (1988) present a computational model of the mental imagery medium which, although restricted to two dimensions, nevertheless employs an explicit spatial representation. Pinker (1980) shows how mental imagery phenomena extend also into the third dimension. Finke *et al.* (1989) show that, given suitable conditions, people can assign novel interpretations to ambiguous images which have been constructed out of parts or mentally transformed. For example, when asked to imagine the letter "D" on its side, affixed to the top of the letter "J", subjects spontaneously report "seeing" an umbrella. Kosslyn *et al.* (1995) show that mental imagery plays a role not only in memory and spatial reasoning tasks, but in fact imagery also plays a role in abstract reasoning, skill learning, and language comprehension. Mellett *et al.* (2000) present regional cerebral blood flow (rCBF) data which provide strong evidence that imagery based on verbal descriptions can recruit cortical regions known to be engaged in high-order visual processing.

The Function of Conscious Experience

The concept of perceptual representation developed above relates directly to the issue of the function of conscious experience, and whether it is an epiphenomenon that has no direct functional value. For once we accept the inescapable fact that the brain is capable of generating three-dimensional volumetric spatial structures in perception, the functional purpose of that spatial representation becomes clear. It is to provide an internal replica of the external world in order to guide our behavior through the world, for otherwise we would have no knowledge of the structure of the world, or of our location within it. Exactly

how behavior is guided by conscious experience can also be determined by phenomenological observation. What that observation reveals is an analogical paradigm of behavioral computation that is quite unlike the analytical symbolic paradigm of computation embodied in the digital computer. In order to illustrate the functional principle behind this unique computational strategy I will present a spatial analogy that operates on the same essential principle as human behavioral computation, although in a much simplified form. I will then present the phenomenological evidence that implicates that same principle of spatial computation in human behavior.

The Plotting Room Analogy

During the Battle of Britain in the second world war, Britain's Fighter Command used a plotting room as a central clearinghouse for assembling information on both incoming German bombers, and defending British fighters, gathered from a variety of diverse sources. A chain of radar stations set up along the coast would detect the range, bearing, and altitude of invading bomber formations, and this information was continually communicated to the Fighter Command plotting room. British fighter squadrons sent up to attack the bombers reported their own position and altitude by radio, and squadrons on the ground telephoned in their strength and state of readiness. Additional information was provided by the Observer Corps, from positions throughout the British Isles. All of this information was transmitted to the central plotting room, where it was collated, verified, and cross-checked, before being presented to controllers to help them organize the defense. The information was presented in the plotting room in graphical form, on a large table map viewed by controllers from a balcony above. Symbolic tokens representing the position, strength, and altitudes of friendly and hostile formations were moved about on the map in order to maintain an up-to-date graphical depiction of the battle as it unfolded. I propose that the functional principle behind this concept of plotting information is directly analogous to the strategy used for perceptual representation in the brain.

From Perception to Behavior

Now the plotting room analogy diverges from perception in that the plotting room does indeed have a "homunculus" or homunculi, in the form of the plotting room controllers, who issue orders to their fighter squadrons based on their observations of the plotting room map. However the idea of a central clearinghouse for assembling sensory information from a diverse array of sensory sources in a unified representation is just as useful for an automated system as it is for one designed for human operators. The automated system need only be equipped with the appropriate spatial algorithms to make use of that spatial data. In order to clarify the meaning of a spatial algorithm that operates on spatial data, I will describe a hypothetical mechanism designed to replace the human controllers in the Fighter Command plotting room. The general principle of operation of that mechanism, I propose, reflects the principle behind human perception and how it relates to behavior. Let us begin by designing a mechanism to command a squadron of friendly fighters to close with an enemy formation depicted on the plotting room map. This objective could be expressed in the plotting room model as a force of attraction, like a magnetic or electrostatic force, that pulls the fighter squadron token in the direction of the approaching bomber formation token on the plotting room map. However the token cannot move directly in response to that force. Instead, that attractive force is automatically translated into instructions for the squadron to fly in the direction indicated by that attractive force, and the force is only relieved or satisfied as the radio, radar, and Observer Corps reports confirm the actual movement of the squadron in the desired direction. That movement is then reflected in the movement of it's token on the plotting room map. The

force of attraction between the squadron token and that of the bomber formation in the plotting room model represents an analogical computational strategy or algorithm, designed to convert a perceptual representation, the spatial model, into a behavioral response, represented by the command for the squadron to fly in the direction indicated by the force of attraction. The feedback loop between the perceived environment and the behavioral response that it provokes, is mediated through actual behavior in the external world, as reflected in sensory or "somatosensory" confirmation of that behavior back in the perceptual model.

To demonstrate the power of this kind of computational strategy, let us delve a little deeper into the plotting room analogy, and refine the mechanism to show how it can be designed to be somewhat more intelligent. When intercepting a moving target such as a bomber formation in flight, it is best to approach it not directly, but with a certain amount of "lead", just as a marksman leads a moving target by aiming for a point slightly ahead of it. Therefore the bomber formation is best intercepted by approaching the point towards which it appears to be headed. This too can be calculated with a spatial algorithm by using the recent history of the motion of the bomber formation to produce a "leading token" placed in front of the moving bomber token in the direction that it appears to be moving, advanced by a distance proportional to the estimated speed of the bomber formation.The leading token therefore represents the presumed future position of the moving formation a certain interval of time into the future. The fighter squadron token should therefore be designed to be attracted to this leading token, rather than to the token representing the present position of the bomber formation itself. But in the real situation the invading bombers would often change course in order to throw off the defense. It was important therefore to try to anticipate likely target areas, and to position the defending fighters between the bombers and their likely objectives. This behavior could be achieved by marking likely target areas, such as industrial cities, airports, or factories etc. with a weaker attractive force to draw friendly fighter squadron tokens towards them. This force, in conjunction with the stronger attraction to the hostile bombers, will induce the fighters to tend to position themselves between the approaching bombers and their possible targets, or to deviate their course towards those potential targets on their way to the attacking bombers, and then to approach the bombers from that direction. Additional forces or influences can be added to produce even more complex behavior. For example as a fighter squadron begins to exhaust its fuel and / or ammunition, its behavior pattern should be inverted, to produce a force of repulsion from enemy formations, and attraction back towards its home base, to induce it to refuel and re-arm at the nearest opportunity. With this kind of mechanism in place, fighter squadrons would be automatically commanded to approach the enemy, attack, and return to base, all without human intervention.

The mechanism described above is of course rather primitive, and would need a good deal of refinement to be at all practical, to say nothing of the difficulties involved in building and maintaining a dynamic analog model equipped with spatial field-like forces. But the computational principle demonstrated by this fanciful analogy is very powerful. For it represents a parallel analogical spatial computation that takes place in a spatial medium, a concept that is quite unlike the paradigm of digital computation, whose operational principles are discrete, symbolic, and sequential. There are several significant advantages to this style of computation. For unlike the digital decision sequence with its complex chains of Boolean logic, the analogical computation can be easily modified by inserting additional constraints into the model. For example if the fighters were required to avoid areas of intense friendly anti-aircraft activity, this additional constraint can be added to the system by simply marking those regions with a repulsive force, that will tend to push the fighter squadron tokens away from those regions without interfering with their other spatial constraints. Since the proposed mechanism is parallel and analog in nature, any

number of additional spatial constraints can be imposed on the system in similar manner, and each fighter squadron token automatically responds to the sum total of all of the analog forces acting on it in parallel. In an equivalent Boolean system, every additional constraint added after the fact would require re- examination of every Boolean decision in the system, each of which would have to be modified to accommodate every combination of possible contingencies. In other words adding or removing constraints after the fact in a Boolean logic system is an error-prone and time consuming business requiring the attention of an intelligent programmer, whereas in the analogical representation spatial constraints are relatively easy to manipulate independently, while the final behavior automatically takes account of all of those spatial influences simultaneously.

Analogical v.s. Sequential Logic

Despite the advantages inherent in the analogical paradigm, there are cases in which a Boolean or sequential component is required in a control system. For example if it is required to direct a squadron to proceed to a point B by way of an intermediate point A. This kind of sequential logic can be incorporated in the analogical representation by installing an attractive force to point A that remains active only until the squadron token arrives there, at which point that force is turned off, and an attractive force is applied to point B instead. Or perhaps the attractive force can fade out gradually at point A in analog fashion as the squadron token approaches, while a new force fades in at point B, allowing the squadron to cut the corner with a smooth curving trajectory instead of a sharp turn, or to adapt the curve of their turn to account for other spatial influences acting on it at that time. The analogical paradigm therefore can be designed to subsume digital or sequential functions, while maintaining the basic analogical nature of the elements of that logic, thereby preserving the advantages of a parallel decision strategy within sequentially ordered stages of processing.

Internal v.s. External Representation

The analogical spatial strategy presented above is reminiscent of the kind of computation suggested by Braitenberg (1984) in his book *Vehicles.* Braitenberg describes very simple vehicles that exhibit a kind of animal-like behavior by way of very simple analog control systems. For example Braitenberg describes a light-powered vehicle equipped with two photocells connected to two electric motors that power two driving wheels. In the presence of light, the current from the photocells drives the vehicle forward, but if the light distribution is non-uniform and one photocell receives more light than the other, the vehicle will turn either towards or away from the light, depending on how the photocells are wired to the wheels. One configuration produces a vehicle that exhibits light-seeking behavior, like a moth around a candle flame, whereas with the wires reversed the same vehicle will exhibit light-avoiding behavior, like a cockroach scurrying for cover when the lights come on. The behavior of these simple vehicles is governed by the spatial field defined by the intensity profile of the ambient light, and therefore, like the analogical paradigm, this type of vehicle also performs a spatial computation in a spatial medium. However in the case of Braitenberg's vehicles, the spatial medium is the external world itself, rather than an internal replica of it. Rodney Brooks (1991) elevates this concept to a general principle of robotics, whose objective is "intelligence without representation". Brooks argues that there is no need for a robotic vehicle to possess an internal replica of the external world, because the world can serve as a representation of itself. O'Regan (1992) extends this argument to human perception, and insists that the brain does not maintain an internal model of the external world, because the world itself can be accessed as if it were an internal memory, except that it happens to be external to the organism. Nevertheless

information from the world can be extracted directly from the world whenever needed, just like a data access of an internal memory store. (See also Pylyshyn 1998)

However there is a fundamental flaw with this concept of perceptual processing, at least as a description of human perception. For unless we invoke mystical processes beyond the bounds of science, then surely our conscious experience of the world must be limited to that which is explicitly represented in the physical brain. In the case of Braitenberg's vehicles, that consciousness would correspond to the experience of only two values, i.e. the brightness detected by the two photocells, and the conscious decision-making processes of the vehicle (if it can be called such) would be restricted to responding to those two values with two corresponding motor signals. These four values therefore represent the maximum possible content of the vehicle's "conscious experience". The vehicle has no idea of its location or orientation in space, and its complex spatial behavior is more a property of the world around it than of anything going on in its "brain". In the case of human perception, our consciousness would be restricted to a sequence of two-dimensional images, as recorded by the retina, or pairs of images in the binocular case. However our experience is very different from the retinal representation. For when we stand in a space, like a room, we experience the volume of the room around us as a simultaneously present whole, every volumetric point of which exists as a separate parallel entity in our conscious experience. Braitenberg's vehicles can be programmed to go to the center of a room by placing a light at that location, but the vehicle cannot conceive of the void of the room around it or the concept of its center, for those are spatial concepts that require a spatial understanding. The world of visual experience therefore clearly demonstrates that we possess an internal map of external space like the Fighter Command plotting room, and the world we see around us is exactly that internal representation.

Symbol Grounding by Spatial Analogy

The analogical spatial paradigm offers a solution to some of the most enduring and troublesome problems of perception. For although the construction and maintenance of a spatial model of external reality is a formidable computational challenge, the rewards that it offers makes the effort very much worth the trouble. The greatest difficulty with a more abstracted or symbolic approach to perception has always been the question of how to make use of that abstracted knowledge. This issue was known as the *symbol grounding problem* (Harnad 1990) in the propositional paradigm of representation promoted by the Artificial Intelligence (AI) movement. The problem of vision, as conceptualized in AI, involves a transformation of the two-dimensional visual input into a propositional or symbolic representation. For example an image of a street scene would be decomposed into a list of items recognized in that scene, such as "street", "car", "person", etc., as well as the relations between those items. Each of these symbolic tags or labels is linked to the region of the input image to which it pertains. The two-dimensional image is thereby carved up into a mosaic of distinct regions, by a process of segmentation (Ballard & Brown 1982, p. 6-12), each region being linked to the symbolic label by which it is identified. Setting aside the practical issues of how such a system can be made to work as intended, (which itself turns out to be a formidable problem) this manner of representing world information is difficult to translate into practical interaction with the world. For the algorithm does not "see" the street in the input image as we do, but rather it sees only a two-dimensional mosaic of irregular patches connected to symbolic labels. Consider the problem faced by a robotic vehicle designed to find a mail box on the street and post a letter in it. Even if an image region is identified as a mail box, it is hard to imagine how that information could be used by the robot to navigate down the street to the mail box avoiding obstacles along the way. What is prominently absent from this system is a three-dimensional consciousness of the

street as a spatial structure, the very information that is so essential for practical navigation through the world.

An analogical representation of the street on the other hand would involve a three-dimensional spatial model, like a painted cardboard replica of the street complete with a model of the robot's own body at the center of the scene. It is the presence of such a three-dimensional replica of the world in an internal model that, I propose, constitutes the act of "seeing" the street. Setting aside the issue of how such a model can be constructed and updated from the two-dimensional sensory image, (which is also a formidable problem) making practical use of such a representation is much easier than for a symbolic or abstracted representation. For once the mailbox effigy in the model is recognized as such, it can be marked with an attractive force, and that force in turn draws the effigy of the robot's body towards the effigy of the mailbox in the spatial model. Obstacles along the way are marked with negative fields of influence, and the spatial algorithm to get to the mailbox is to follow the fields of force, like a charged particle responding to a pattern of electric fields.

An essential component of this analogical concept of perceptual processing therefore is a spatial replica of the percipient's own body in the spatial replica of his environment. This perceptual "homunculus" is not to be confused with the internal observer of the "homunculus objection", for this body replica is not the observer of the internal scene, but is merely another object in the perceived world, for that world would be incomplete without a representation of the percipient's own body as an object in the scene. The analogy of the fighter command plotting room reveals the function of this body percept, i.e. that it is required to perform the spatial computations that elicit the appropriate behavioral response to that environment. The fact that human perception employs such an internal "body image" is plainly evident by inspection of our own apparent body at the center of our phenomenal world. But this fact is only evident to the indirect realist, for the naive realist mis-identifies his perceived body with his actual physical body in physical space, a paradigmatic error which has been shown to be untenable many times over. (Russell 1927, Schilder 1942, 1950, Smythies 1953, Ramachandran & Blakeslee 1998, Smythies & Ramachandran 1998)

The analogical paradigm can also be employed to compute the more detailed control signals to the robot's wheels. The forward force on the model of the robot's body applies a torque force to the model wheels, but the model wheels cannot respond to that force directly. Instead, that torque in the model is interpreted as a motor command to the wheels of the larger robot to turn, and as the larger wheels begin to turn in response to that command, that turning is duplicated in the turning of the model wheels, producing behavior as if responding directly to the original force in the model world. Side forces to steer the robot around obstacles can also be computed in similar fashion. A side force on the model robot should be interpreted as a steering torque, like the torque on the pivot of a caster wheel. That pivoting torque in the model is interpreted as a steering command to pivot the larger wheels, and the steering of the larger wheels is then reflected in the steering of the model wheels also. The forces impelling the model robot through the model world are thereby transformed into motor commands to navigate the real robot through the real world, and the physical response of the robot to those commands is in turn communicated back into the model world to keep it aligned with events in the external world.

The idea of motor planning as a spatial computation has been proposed in *field theories* of motor control, (Gibson & Crooks 1938, Koffka 1935, Lewin 1969) in which the intention to walk towards a particular objective in space is expressed as a field-like force of attraction, or *valence*, between a model of the body, and a model of the target, expressed in a spatial model of the local environment. The target is

marked with a positive valence, while obstacles along the way are marked with negative valence. When we see an attractive stimulus, for example a tempting delicacy in a shop window at a time when we happen to be hungry, our subjective impression of being physically drawn towards that stimulus is not only metaphorically true, but I propose that this subjective impression is a veridical manifestation of the mental mechanism that drives our motor response. For the complex combination of joint motions responsible for deviating our path towards the shop window are computed in spatial fashion in a spatial model of the world, exactly as we experience it to occur in subjective consciousness. Indeed the spatial configuration of the positive and negative valence fields evoked by a particular spatial environment can be inferred from observation of their effects on behavior, in the same way that the pattern of an electric field can be mapped out by its effects on moving charged particles. For example the negative valence field due to an obstacle such as a sawhorse placed on a busy sidewalk can be mapped by observing its effect on the paths of people walking by. The moving stream of humanity divides to pass around the obstacle like water flowing around a rock in a stream in response to the negative valence field projected by that obstacle. Although the influence of this obstacle is observed in external space, the spatial field that produces that behavioral response actually occurs in the spatial models in the brains of each of the passers-by individually.

Another example of a spatial computational strategy can be formulated for the problem of targeting a multi-jointed limb, i.e. specifying the multiple angles required of the individual joints of the limb in order to direct its end-effector to a target point in three-dimensional space. This is a complex trigonometrical problem that is underconstrained. However a simple solution to this complex problem can be found by building a scale model of the multi-jointed limb in a scale model of the environment in which the limb is to operate. (McIntyre & Bizzi 1993) The joint angles required to direct the limb towards a target point can be computed by simply pulling the end-effector of the model arm in the direction of the target point in the modeled environment, and recording how the model arm reacts to this pull. Sensors installed at each individual joint in the model arm can be used to measure the individual joint angles, and those angles in turn can be used as command signals to the corresponding joints of the actual arm to be moved. The complex trigonometrical problem of the multi-jointed limb is therefore solved by analogy, as a spatial computation in a spatial medium.

There is psychophysical evidence to suggest that this kind of strategy is employed in biological motion. For when a person reaches for an object in space, their body tends to bend in a graceful arc, whose total deflection is evenly distributed amongst the various joints to define a smooth curving posture, i.e. the motor strategy serves to minimize a configural constraint expressed in three-dimensional space, thus implicating a spatial computational strategy. The dynamic properties of motor control are also most simply expressed in an external spatial context. For the motion of a person's hand while moving towards a target describes a smooth arc in space and time, accelerating uniformly through the first half of the path, and decelerating to a graceful stop through the second half. (Bizzi *et al.* 1995, 2000, Dornay *et al.* 1993, Hollerback 1990) In other words the observed behavior is exactly as if the person's body were indeed responding lawfully to a spatial force of attraction between the hand and the target object in three-dimensional space, which in turn suggests that a spatial computational strategy is being used to achieve that result. Further evidence comes from the subjective experience of motor planning, for we are unaware of the individual joint motions when planning such a move, but rather our experience is more like a force of attraction that seems to pull our hand towards the target object, and the joints in our arm seem to simply follow our hand as it responds to that pull. This computational strategy generalizes to any configuration of limbs with any number of joints, as well as to continuous limbs like a snake's body or an

elephant's trunk.

Conclusion

The study of mind and its place in nature has always been a principal focus of philosophy. But from the outset the investigation has been plagued by a fundamental confusion over the object of this inquiry, i.e. which aspects of phenomenal experience are a manifestation of mind as distinct from the world beyond the mind. The debate began three centuries ago with a choice between the two epistemological monist alternatives of realism and idealism; either the world we perceive around us is the real world itself, or we cannot see the world directly, all we can ever experience is our own mind. The problem is that the phenomenal world shows evidence of both origins. It appears to be the world external to our mind because it seems to have an objective external existence independent of our mental states. And yet there is also clear evidence that the world of experience is a product of mind, as seen for example in visual illusions, and even more clearly in the case of dreams and hallucinations. But if the mind is but the activity of the brain, how can a product of our brain, which is in our head, escape the confines of our head to appear in the world around us? Both of these epistemological monist alternatives are inconsistent with the facts of the causal chain of visual processing.

An extraordinary variety of intermediate theories have been proposed over the centuries, in an attempt to place phenomenal experience partially inside, and partially outside the head, but in neither place explicitly. From Descartes' dualist mind as a non-spatial entity with no defined location in space, to Malebranche's perceived colors which are in the mind, but also somehow in the external object, to the sense data of the critical realists that are experienced, but which do not, or may not exist, to Davidson's notion of supervenience, of which the mind / brain relation is the only example in the known universe. All of these explanations propose to make an exception in the laws of nature just to accommodate the special case of conscious experience. The only alternative which does not entail suspension of the normal laws of nature is epistemological dualism. This theory explains how the phenomenal world can appear external to the body while at the same time actually being in the head. It explains how different individuals can each have their own unique perspective on a commonly viewed object. And it offers the only plausible explanation for those most troublesome phenomena of dreams and hallucinations, as well as for the data of mental imagery and neglect syndrome, which no longer require heroic efforts of *denial* to account for their manifest properties. All of these phenomena follow naturally from the indirect view of perception.

But the indirect realist solution comes at a cost. In return for resolving the epistemological question, indirect realism opens a new paradox, and that is a glaring disparity between two primary sources of knowledge, phenomenology and neurophysiology. Phenomenology presents the mind as a three-dimensional colored structure or analogical representation, while neurophysiology presents the brain as an assembly of billions of discrete quasi-independent local processors interconnected in a massively parallel network. Where in that mass of neural circuitry are the three-dimensional volumetric real-time moving pictures that we know so well in conscious experience? The brain just seems to be the wrong kind of device to create that kind of representation. Is consciousness therefore an illusion with no direct neurophysiological correlate? Or is there something fundamentally wrong with our understanding of neurophysiology?

The information that phenomenal experience gives us about the external world is known to be somewhat uncertain, as we are easily fooled by illusions, and occasionally by outright hallucinations. But when the

object of our phenomenological investigation is conscious experience itself, our knowledge of that particular entity is very certain. In fact our knowledge of our own conscious state is more certain and reliable than any other knowledge we can possibly have, even when our conscious experience is itself only a hallucination. Neuroscience on the other hand is a science very much in its infancy, and is rife with uncertainty. In fact the "dirty little secret" of neuroscience, as Searle (1997, p. 198) calls it, is that the central principles of representation and computation in the brain remain to be discovered. Very little is known with any real certainty about how perceptual or cognitive information is encoded in the brain, or what kind of computation the brain actually performs in perception. And there are several prominent aspects of brain activity whose functional significance remains almost entirely obscure, such as the synchronous oscillations observed between neurons in remote cortical areas, and the global oscillations of the brain as a whole as seen in Electroencephalogram (EEG) recordings. The phenomenological inspection of conscious experience therefore offers more reliable and certain knowledge of the essential principles of mental representation and function than anything that modern neuroscience has yet to offer, because it gives us direct access to the massive quantities of information encoded in the brain, presented in a form that is immediately meaningful to us. If our observations of the nature of phenomenal experience are in conflict with contemporary concepts of neurocomputation, it is our neurophysiological theories which are in urgent need of revision in order to bring them in line with observed phenomenologically. For a neuroscience which explains everything about the brain except for how it generates the mind, is a neuroscience which essentially explains nothing, because it is the mind that makes the brain interesting in the first place.

So if we identify the world of experience as an internal spatial model, what does that tell us about the function of conscious experience? In the first place it tells us that one of the most significant functions of conscious experience is to serve as a structural model of the external world in an internal representation. This function is completely transparent, or invisible, as long as consciousness is viewed from the naïve realist perspective, which is why this most obvious fact of perception has gone unnoticed for so long. Now it might be argued that this is not a function of conscious experience itself, but only of perception. For example one could still imagine a hypothetical zombie that behaves in every way like a conscious human being, but supposedly lacks all conscious experience. However if perception is indeed indirect, and if behavior is governed by analogical forces, this means that the zombie would also have to be equipped with a volumetric spatial model of external reality and an analogical computational strategy in order to duplicate human behavior. And the zombie must also be able to report the colors of the surfaces in that internal model, all in the absence of conscious experience. This description of unconscious experience comes so close to a description of consciousness itself as to leave very little real distinction, because the structural and representational aspects of consciousness are every bit as much an essential part of visual consciousness as is the experiential, or "what it is like" aspect. In any case, whether or not hypothetical zombies can have an internal spatial model without a conscious experience of it, we know for a fact that we ourselves *do* have an internal spatial model, and that in our case we are also conscious of it, which makes the whole question of zombies somewhat moot with respect to human consciousness.

Information theory offers an interesting new angle on the problem. For information cannot exist without some physical medium, or carrier, because the information is encoded as modulations of that carrier. In the brain the carrier is some kind of electrochemical state, and the information encoded in the brain is presumably expressed as modulations of that electrochemical state across space and time. A similar information theoretic organization is observed on the subjective side of the mind / brain barrier. Every point in the three-dimensional matrix of phenomenal space can express every color in the gamut of

phenomenal color experience, including the experience of transparency, or of empty space. Conscious experience is expressed in perception as patterned modulations of those color qualia across space and time. Information theory therefore suggests that the qualia, such as the primal experience of color and space, are themselves the carrier, or the mechanism by which experience is represented or expressed in the brain, whereas the spatial and temporal modulations of those basic qualia across the volume of phenomenal space represent the information content of the representation, i.e. the perceptual scene that is being currently portrayed. For the most part perception is indirect, we view the world through the medium of conscious experience. But there is one, and only one entity that we do see directly, and that is the representational mechanism itself, the inside of our own brain. The volume of space we perceive around us is a data structure in our physical brain, and the primal color qualia with which that world is painted are different states of the physical mechanism of our own physical brain. That does not mean that those parts of the brain would actually appear colored to a micro-electrode inserted into that part of the brain, nor would they appear colored under microscopic examination. But that does not make them any the less colored, or any the less an intrinsic property of the physical brain.

This description of conscious experience applies only to the sensory component of perception, i.e. the non-epistemic experience of vivid colored surfaces embedded in phenomenal space. Epistemic perception, which involves an understanding, or recognition of patterns of experience, is not expressed in the form of patterns of vivid modal experience, and therefore requires a different explanation.

But how does the brain make use of this structural information? Who is the observer of this internal scene? Well for one thing it is not the "homunculus" if by that is meant a miniature copy of the entire brain. The data of consciousness need only be available to other internal processes and mechanisms designed to read and interpret that data, and to generate an appropriate behavioral response. And once we recognize conscious experience for what it really is, we can employ phenomenological observation to determine not only the structure of conscious experience, but also the principles of its function, i.e. the principle by which fears, urges, pains, and desires, often stimulated by recognized patterns present in the conscious representation, are expressed as forces in our perceptual space that seem to thrust us away from aversive patterns while drawing us towards attractive ones. And the object on which these synthetic forces act in perception is a different kind of "homunculus" at the center of our perceptual representation, the item known in psychology as the "body image", most frequently mis-identified as our own physical body. It is the arms and legs and torso that we perceive to sprout outward from the egocentric point of our private representational space. It is this perceptual replica of a human body that feels the influence of the analog forces that appear in perceived space in response to perceived objects. And the forces that act on the body image are interpreted as motor commands to the larger external body which is beyond our direct experience. As the greater external body moves in response to these internal commands, the body image mirrors those movements in the internal replica. The external physical body therefore moves in the world exactly as if it were responding to analog field forces in the external environment directly, although in fact it is responding indirectly to the miniature forces in the internal replica. At the same time the subjective experience of consciousness gives us the impression of being a free agent in an external world, although in fact our conscious experience is forever entombed within the walls of our own physical skull. Until this most basic fact of conscious experience is generally accepted as an essential fact of nature, philosophy will be condemned to a view of consciousness as something that is deeply mysterious, and forever beyond the capacity of human comprehension. The indirect realist perspective reveals that in fact it is the remote external world which is forever beyond human comprehension, and that consciousness is perhaps the *only* thing we can *ever* fully comprehend.

References

Adams R. M. (1987) Flavors, Colors, and God. In The Virtue of Faith and Other Essays in Philosophical Theology. New York: Oxford University Press.

Baldwin T. (1992) The Projective Theory of Sensory Content. In: T. Crane (Ed.) The Contents of Experience: Essays on Perception. Cambridge UK: Cambridge University Press, 177-195.

Ballard D. H. & Brown C. M. (1982) "Computer Vision". Prentice-Hall, Englewood Cliffs, NJ.

Bisiach E. & Luzzatti C. (1978) Unilateral Neglect of Representational Space. Cortex 14, 129-133.

Bisiach E. Luzzatti C., & Perani D. (1979) Unilateral Neglect, Representational Schema and Consciousness. Brain 102, 609-618.

Bisiach E., Capitani E., Luzzatti C., & Perani D. (1981) Brain and Conscious Representation of Outside Reality. Neuropsychologia 19, 543-552.

Bizzi E., Mussa-Ivaldi F. A. (1995) "Toward a Neurobiology of Coordinate Transformations." IN:The Cognitive Neurosciences. M. S. Gazzaniga (Ed.) Cambridge MA: The Mit Press, 495-506.

Bizzi E., Mussa-Ivaldi, F. A., Ghez, C., Krakauer J. W. (et al). (2000) Motor systems. In: The New Cognitive Neurosciences (2nd ed.). M. S. Gazzaniga (Ed.) Cambridge, MA: The MIT Press, 489- 618.

Block N. (1981) Introduction. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press, 1-18

Boring E. G. (1933) The Dimensions of Consciousness. New York: Dover.

Boring E. G. (1950) A History of experimental psychology. New York: Appleton-Century-Crofts.

Braitenberg (1984) Vehicles: Experiments in Synthetic Psychology. Cambridge MA: MIT Press.

Broad C. D. (1925) The Mind and Its Place In Nature. London: Routledge & Kegan Paul.

Brooks R. A. (1991) Intelligence Without Representation. Artificial Intelligence 47 (1/3) p. 139.

Cabanis P. J. G. (1802) Rapports du Physique et du Moral de L'Homme. In G. S. Brett (Ed. 1921) History of Psychology II, p. 375-382.

Chalmers, D. J. (1995) Facing Up to the Problems of Consciousness. Journal of Consciousness Studies 2 (3) 200-219. Reprinted in "Toward a Science of Consciousness II, The Second Tucson Discussions and Debates". (1996) S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.) Cambridge MA: MIT Press, 5-28.

Churchland P. M. (1981) Eliminative Materialism and the Propositional Attitudes. Journal of Philosophy 78, 67-90.

Churchland P. S. (1983) Consciousness: The Transmutation of a Concept. Pacific Philosophical Quarterly 64, 80-93.

Clark A. (1993) Sensory Qualities. Oxford: Clarendon Press.

Davidson D. (1970) Mental Events. Oxford: Oxford University Press.

Davidson D. (1980) Essays on Actions and Events. Oxford: Oxford University Press.

Denis M. & Cocude M (1989) Scanning Visual Images Generated from Verbal Descriptions. European Journal of Cognitive Psychology 1, 293-307.

Dennett D. C. (1981) The Nature of Images and the Introspective Trap. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press. 51-61.

Dennett D. (1988) Quining Qualia. In Consciousness in Contemporary Science (Marcel A. J., & Bisiach E. Eds.) Oxford: Clarendon Press, pp 42-87.

Dennett D. (1991) Consciousness Explained. Boston, Little Brown & Co.

Dennett D. (1992) `Filling In' Versus Finding Out: a ubiquitous confusion in cognitive science. In Cognition: Conceptual and Methodological Issues, Eds. H. L. Pick, Jr., P. van den Broek, & D. C. Knill. Washington DC.: American Psychological Association.

Denny-Brown D. & Chambers R. A. (1958) The Parietial Lobe and Behavior. A. Res Nerv Ment Dis 36, 35-117.

de Renzi E. (1982) Disorders of Space Exploration and Cognition. New York: John Wiley.

Dornay, M., Mussa-Ivaldi F. A., McIntyre, J., Bizzi E. (1993) Stability Constraints for the Distributed Control of Motor Behavior. Neural Networks 6 (8), 1045-1059.

Drake D., Lovejoy A. O., Pratt J. B., Rogers A. K., Santayana G., Sellars R. W., & Strong C. A. (1920) Essays in Critical Realism: A co-operative study of the problem of knowledge. New York: Gordian Press.

Dretske F. I. (1995) Naturalizing the mind. Cambridge MA: MIT Press.

Feigl H. (1958) The `Mental' and the `Physical'. Minnesota Studies in the Philosophy of Science, Vol. 2. H. Feigl, G. Maxwell, & M. Scriven (Eds.), Minneapolis: University of Minnesota Press.

Finke R. A., Pinker S., & Farah M. J. (1989) Reinterpreting Visual Patterns in Mental Imagery. Cognitive Science 13, 51-78.

Flanagan O. (1992) Consciousness Reconsidered. Cambridge MA: MIT Press.

Fodor J. A. (1981) Imagistic Representation. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press, 63-86.

Gibson J. J. & Crooks L. E. (1938) A Theoretical Field-Analysis of Automobile Driving. The American Journal of Psycholgy 51 (3) 453-471.

Gibson J. J. (1972) A Theory of Direct Visual Perception. In: The Psychology of Knowing. (J. R. Royce & W. W. Rozeboom (Eds.), Gordon & Breach.

Gregory R. L. (1981) Mind in Science. Cambridge UK: Cambridge University Press.

Harnad S. (1990) The Symbol Grounding Problem. Physica D. 42, 335-346.

Harrison S. (1989) A New Visualization on the Mind-Brain Problem: Naive Realism Transcended. In J. Smythies & J. Beloff (Eds.) The Case for Dualism. Charlottesville: University of Virginia.

Heilman K. M. & Watson R. T. (1977) The Neglect Syndrome - A unilateral defect of the orienting response. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.) Lateralization in the Nervous System. New York: Academic Press.

Heilman K. M., Watson R. T. & Valenstein E. (1985) Neglect and Related Disorders. In K. M. Heilman & E. Valenstein (Eds.) Clinical Neuropsychology. New York: Oxford University Press.

Hoffman D. D. (1998) Visual Intelligence: How We Create What We See. New York: W. W. Norton.

Hollerbach J. M. (1990) Fundamentals of Motor Behavior. In: D. N. Osherson & S. M. Kosslyn (Eds) Visual Cognition and Action: An invitation to cognitive science, Vol. 2. Cambridge MA: The Mit Press 153-182

Holt E. B., Marvin W. T., Montague W. P., Perry R. B., Pitkin W. B., & Spaulding E. G. (1912) The New Realism. New York: The MacMillan Co.

Humphrey N. (1999) The Privatization of Sensation. In: S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.) Toward a Science of Consciousness II the Second Tucson Discussions and Debates, Cambridge MA: MIT Press, 247-257.

Kant I. (1781 / 1991) Critique of Pure Reason. Vasilis Politis (Ed.) London: Dent.

Kinsbourne M. (1987) Mechanisms of Unilateral Neglect. In M. Jeannerod (Ed.) Neurophysiological and Neuropsychological Aspects of Spatial Neglect. Amsterdam: North-Holland.

Kinsbourne, M. (1993). Orientational bias model of unilateral neglect:Evidence from attentional gradients within hemispace. In I.H. Robertson & J.C. Marshall (Eds), Unilateral neglect: clinical and experimental studies (pp. 63-86). Hove, UK: Erlbaum.

Kim, J. (1998) Mind in a Physical World: An essay on the mind-body problem and mental causation. Cambridge MA: MIT Press.

Koffka K, (1935) Principles of Gestalt Psychology. New York, Harcourt Brace & Co.

Köhler W. (1924) Die physischen Gestalten in Ruhe und im stationären Zustand: Eine naturphilosophische Untersuchung. Erlangen: Verlag der Philosophichen Akademie.

Köhler W. (1971) A Task For Philosophers. In: The Selected Papers of Wolfgang Koehler, Mary Henle (Ed.) Liveright, New York. pp 83-107.

Kosslyn S. M. (1975) Information Representation in Visual Images. Cognitive Psychology 7, 341-370.

Kosslyn S. M. (1981) The Medium and the Message in Mental Imagery. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press, 207-244.

Kosslyn S. M. (1995) Mental Imagery. In: S. M. Kosslyn & D. N. Osherson (Eds.) Visual Cognition: An invitation to cognitive science, Vol 2 (2nd ed.) Cambridge MA: MIT Press, 267-296.

Kosslyn S., Ball T., & Reisner B. (1978) Visual Images Preserve Metric Spatial Information: Evidence of

Studies from Image Scanning. Journal of Experimental Psychology: Human Perception and Performance 4, 47-60.

Kosslyn S. M., Behrmann M., & Jeannerod M. (1995) The Cognitive Neuroscience of Mental Imagery. Neuropsychologia 33 (11), 1335-1344.

Kuhn T. S. (1970) The Structure of Scientific Revolutions. Chicago: Chicago University Press.

Làdavas E., Berti A., Ruozzi E., & Barboni F. (1997) Neglect as a Deficit Determined by an Imbalance Between Spatial Representations. Experimental Brain Research 116, 493-500.

La Mettrie (1748) L'Homme Machine. In G. S. Brett (Ed. 1921) History of Psychology II, p. 358.

Lehar S. (1992) Gestalt Isomorphism and the Primacy of the Subjective Conscious Experience: A Gestalt Bubble Model. Submitted: Behavioral and Brain Sciences September 1999, under review. Available at http://cns-alumni.bu.edu/pub/slehar/webstuff/bubw2/bubw2.html.

Levine D. N., Warach J., & Farah M. J. (1985) Two Visual Systems in Mental Imagery: Dissociation of "what" from "where" in imagery disorders due to bilateral posterior cerebral lesions. Neurology 35, 1010-1018.

Lewin K. (1969) Principles of Topological Psychology. New York: McGraw-Hill.

Llinás R. R. & Paré D. (1991) Of Dreaming and Wakefulness. Neuroscience 44 (3), 521-535.

Locke J. (1690) An Essay Concerning Human Understanding. 2 vols. A. C. Fraser (Ed.) from 3rd edition, 1694. Oxford: Clarendon 1894. (Reprinted New York: Dover, 1959.

Lovejoy A. O. (1930) The Revolt Against Dualism: An inquiry concerning the existence of ideas. W. W. Norton & Co.

Malebranche N. (1674) De la Recherche de la Verité. Vol. 2 (Edition of Bouillier 1880) Paris: Garnier.

Marshall J. C. & Halligan P. W. (1995) Seeing the Forest but only Half the Trees? Nature 373, 521-523.

McFie J. & Zangwill O. L. (1960) Visual-Constructive Disabilities Associated with Lesions of the Left Cerebral Hemisphere. Brain 83, 243-260.

McGinn C. (1989) Can We Solve the Mind-Body Problem?" Mind 98, 349-366.

McIntyre J. & Bizzi E. (1993) "Servo Hypotheses for the Biological Control of Movement." Journal of Motor Behavior 25 (3):193-202.

Mellet E., Tzourio-Mazoyer N., Bricogne S., & Mazoyer B. (2000) Functional Anatomy of High-Resolution Visual Mental Imagery. Journal of Cognitive Neuroscience 12 (1), 98-109.

Nagel T. (1974) What Is It Like to Be a Bat? Philosophical Review 83 435-450

O'Regan K. J., (1992) Solving the `Real' Mysteries of Visual Perception: The World as an Outside Memory. Canadian Journal of Psychology 46 461-488.

O'Shaughnessy B. (1980) The Will: A Dual Aspect Theory. (2 volumes) Cambridge UK: Cambridge University Press.

Pastore N. (1971) Selective History of Theories of Visual Perception: 1650-1950. New York: Oxford University Press.

Pessoa L., Thompson E., & Noë A. (1998) Finding Out About Filling-In: A guide to perceptual completion for visual science and the philosophy of perception. Behavioral and Brain Sciences 21, 723-802.

Pinker S. (1980) Mental Imagery and the Third Dimension. Journal of Experimental Psychology: General. 109, 354-371.

Pinker S. (1988) A Computational Theory of the Mental Imagery Medium. In M. Denis & J. Engelkamp (Eds.) Cognitive and Neuropsychological Approaches to Mental Imagery. Dordrecht, Netherlands: Martinus Nijhoff, 17-32.

Popper K. & Eccles J. (1977) The Self and Its Brain. New York: Springer-Verlag.

Putnam H. (1968) Psychological Predicates. In Putnam H. (Ed.) Collected Papers II. Cambridge UK: Cambridge University Press.

Putnam H. (1975) The Nature of Mental States. In Putnam H. (Ed.) Collected Papers II. Cambridge UK: Cambridge University Press.

Pylyshyn Z. (1981a) The Imagery Debate. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press. 151-206.

Pylyshyn Z. (1981b) The Imagery Debate: Analog Media Versus Tacit Knowledge. Psychological Review 88 (1), 16-45.

Pylyshyn Z. (1988) The Imagery Debate: Analog Media Versus Tacit Knowledge. In: A. M. Collins & E. E. Smith (Eds.) Readings in Cognitive Science: A perspective from psychology and artificial intelligence. San Mateo CA: Morgan Kaufmann Inc. 600-614.

Pylyshyn, Zenon. (1998) Visual indexes in Spatial Vision Imagery. In: Richard D. Wright, (Ed.) Visual Attention. New York, NY Oxford University Press, 215-231.

Ramachandran V. S. & Blakeslee S. (1998) Phantoms in the Brain: Probing the Mysteries of the Human Mind. New York: William Morrow & Co.

Revonsuo A. (1995) Consciousness, Dreams, and Virtual Realities. Philosophical Psychology 8 (1) 35-58.

Ruch (1950) In J. F. Fulton (Ed.) Textbook of Physiology, 16th Ed. Philadelphia, p. 311. Pertinent passage quoted in Smythies (1954).

Russell B. (1921) The Analysis of Mind. New York: MacMilan; London: George Allen & Unwin Ltd.

Russell B. (1927) Philosophy. New York: W. W. Norton.

Schilder P. (1942) Mind: Perception and thought in their constructive aspects. New York: Columbia University Press.

Schilder P. (1950) The Image and Appearance of the Human Body; Studies in the constructive energies

of the psyche. New York: International University Press.

Schwartz R. (1981) There's More To It Than Meets the Eye. In N. Block (Ed.) Imagery, Cambridge MA: MIT Press, 109-130.

Searle J. R. (1992) The Rediscovery of Mind. Cambridge MA: The MIT Press.

Searle J. R. (1997) The Mystery of Consciousness. New York: New York Review.

Sellars R. W. (1916) Critical realism: Astudy of the nature and conditions of knowledge. New York: Russell & Russell.

Sellars W. (1963) Philosophy and the Scientific Image of Man. In: Science, Perception, and Reality. W. Sellars (Ed.) New York: The Humanities Press.

Shepard, R.N. and Metzler J.(1971) Mental rotation of three-dimensional objects. Science. 171(972):701-3.

Smart J. J. C. (1959) Sensations and Brain Processes. Philosophical Review 68, 141-156.

Smythies J. R. (1953) The Experience and Description of the Human Body. Brain 76, 132-148.

Smythies J. R. (1954) Analysis of Projection. British Journal for the Philosophy of Science 5, 120-133.

Smythies J. R. (1989) The Mind-Brain Problem. In: J. R. Smythies & J. Beloff (Eds) The Case For Dualism. Charlottesville: University of Virginia Press.

Smythies J. R. (1994) Requiem for the Identity Theory. Inquiry 37, 311-329.

Smythies J. R. (1994) The Walls of Plato's Cave: the science and philosophy of brain, consciousness, , and perception. Aldershot UK: Avebury.

Smythies J. R. (1999) Consciousness: Some basic issues - A neurophysiological Perspective. Consciousness & Cognition 8, 164-172.

Smythies J. R. & Ramachandran V. S. (1998) An Empirical Refutation of the Direct Realist Theory of Perception. Inquiry 40, 437-438.

Swinburne R. (1984) Personal Identity: The Dualist Theory. In S. Shoemaker & R. Swinburne (Eds.) Personal Identity. Oxford: Blackwell.

Tye M. (1995) Ten Problems of Consciousness: a representational theory of the phenomenal mind. Cambridge, MA: MIT Press.

Vallar G. (1998) Spatial Hemineglect in Humans. Trends in Cognitive Sciences, 2 (3) 87-96.

Velmans M. (1990) Consciousness, Brain and the Physical World. Philosophical Psychology 3 (1) 77-99.

Vernon M. D. (1937) Visual Perception. Cambridge UK: Cambridge University Press.

Vernon M. D. (1952) A Further Study of Visual Perception. Cambridge UK: Cambridge University Press.