The Journal of Mind and Behavior
Winter 2006, Volume 27, Number 1
Pages 43–56
ISSN 0271–0137

The Only Objective Evidence for Consciousness

Fred Kuttner and Bruce Rosenblum

University of California, Santa Cruz

We describe what seems to be the only *objective* evidence for the existence of consciousness as an entity beyond its neural correlates. We display this evidence, the nature of observation in quantum mechanics, with a theory-neutral version of the archetypal demonstration of quantum phenomena, the two-slit experiment. This undisputed empirical result provides objective evidence for consciousness, the straightforward alternative being the assumption of not only a completely deterministic world, but a conspiratorial one as well. The objection to this evidence for consciousness, that a not-conscious robot could be the observer, is examined.

Keywords: consiousness, quantum, evidence

Since our title indicates a controversial thesis, we start with our definitions. By "objective evidence" we mean third-person evidence that can be displayed to essentially all observers. This differs from first-person introspection (I know I have consciousness) or second-person reports (you say you have consciousness). Objective evidence in this sense is the normal requirement for establishing the reliability of a scientific theory. The sense in which we use "consciousness" is closely related to "awareness" or "subjective experience." It certainly includes the impression of free will. Ultimately, the term is best defined by its use in the experiments we describe. Our use of "consciousness" is the one commonly used in the literature of the quantum measurement problem. Though the consciousness we will speak of involves phenomena for which evidence can be objectively displayed, we in no way imply that studies based on introspection or on second-person reports are not valuable.

For comments on the subject and on a draft of this paper we thank Leonard Anderson, Phyllis Arozena, Donald Coyne, Freda Hedges, Alex Moraru, and Andrew Neher. Requests for reprints should be sent to Fred Kuttner, Ph.D., Department of Physics, University of California, Santa Cruz, California 95064.

However, the very existence of a consciousness displayed only in first-person or second-person reports can be denied. The claim is in fact frequently made that there is *no* such entity beyond the neural correlates of consciousness, the electrochemical signals that can be correlated with behavior. For example, Crick identifies electrochemical activity as being all there is to our subjective experience:

> . . . "You," your joys and sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. (1994, p. 3)

In contrast, David Chalmers (1996) claims that studies of electrochemistry can never explain subjective experience. He defines such studies as among the "easy problems" of consciousness. The explanation of conscious *experience* he calls the "hard problem" of consciousness, one that will require "psychophysical principles" beyond today's science.

To contest Crick's claim, and perhaps support the existence of the "hard problem," we will describe a demonstration of the unresolved "measurement problem" of quantum mechanics.[1] The empirical facts we report are completely undisputed, and their connection with consciousness has been discussed for decades. The demonstration can be considered *objective* evidence for the existence of consciousness as an entity *beyond* its neural correlates.

The objective evidence we will present is a version of the archetypal quantum mechanical demonstration, the so-called "two-slit experiment." In it a conscious choice is one hundred percent correlated with a physical situation *that would have been different* had an alternate choice been made. The experimental results described are generally accepted as a demonstration that a physical situation is created by its observation. A leading quantum mechanics text (Griffiths, 1995) emphasizes this by quoting a founder of the theory, Pascual Jordan: "Observations not only *disturb* what is to be measured, they *produce* it" (p. 3). The conventional interpretation of this, the usual version of the "Copenhagen interpretation," assumes that, for all *practical* purposes, such "observation" is performed by a *non*-conscious measuring device. However, over the decades, deeper versions of the Copenhagen interpretation have treated observation as enigmatically requiring a *conscious* observer (e.g., Stapp, 2004; von Neumann, 1932/1955; Wigner, 1961/1983).

In contrast to the usual *theory*-based treatments of the involvement of consciousness in quantum mechanics, ours is a theory-*neutral* description of a

[1]Chalmers (1996) in fact suggests a connection of the measurement problem of quantum mechanics with the hard problem of consciousness. The last chapter of his book, *The Conscious Mind*, is titled "The Interpretation of Quantum Mechanics."

quantum experiment.[2] (While no demonstration can be completely theory-neutral, we make no reference to the quantum theory.) Only facts observable by anyone constitute the evidence we cite for the direct involvement of consciousness in a physical phenomenon. We do not go beyond reporting those facts to speculate on the *nature* of the involvement of consciousness. We point to a footprint at the crime scene without suggesting a culprit.

The evidence for the involvement of consciousness in physical phenomena that is provided by the quantum experiment is circumstantial, meaning that one fact is used to infer another fact. Circumstantial evidence more readily admits different interpretations than does direct evidence. (It can nevertheless be convincing. It can legally secure a conviction.) But the logic involved in circumstantial evidence can be circuitous. Therefore, to illustrate the logic of the undisputed quantum demonstration presented later, we first tell a story, a parable, that is closely analogous to the quantum demonstration but in which the evidence for the physical involvement of consciousness beyond its neural correlates is direct rather than circumstantial. The demonstration of the parable cannot actually be done. But were that demonstration possible, it would be *direct* evidence for the physical involvement of consciousness beyond its neural correlates rather than the circumstantial evidence presented by the actual quantum demonstration. The point of the parable is merely to illustrate the chain of reasoning.

A *Consciousness Parable*

Dr. Elbe claims to demonstrate that a physical phenomenon external to the body can be brought about by conscious mental effort alone, without any physical mediation. Dr. Elbe displays a large number of box pairs. She instructs you, in your first experiment, to determine which box of each pair holds a marble by opening the boxes of a pair *in turn*. Opening the boxes sequentially, about half the time you find a marble in the first box and half the time in the second.

Presenting a second set of box pairs, Dr. Elbe notes that each marble can come apart into white and black hemispheres. She instructs you, in a second experiment, to determine which box of each pair contains the white hemisphere and which the black by opening both boxes of each pair *at about the same time*. Opening the boxes simultaneously, you always find a white hemisphere in one of the boxes and a black in the other box of that pair.

---

[2]Since the outcomes of the quantum "experiments" we will describe are all well known, the term "demonstrations" would be equivalent.

Now presenting you with further sets of box pairs, Dr. Elbe suggests that for each set you freely choose *either* of the two previous experiments. That is, you may open the boxes either sequentially *or* simultaneously. Allowed to repeat the experiment *of your choice* as many times as you wish, whenever you decide to open the boxes sequentially, you find the marble wholly in a single box; whenever you decide to open the boxes simultaneously, you find the marble distributed over both boxes of the pair.

Puzzled by the fact that the condition of the marble seems to depend on the way you choose to open the boxes, you challenge Dr. Elbe: "Obviously, some of your sets of box pairs had a whole marble in a single box, while other sets contained half a marble in each box. But how did I always get a result corresponding to the opening method I chose? After all, before I opened the boxes each marble had to have been either wholly in a single box or else have its parts distributed over both boxes of the pair. When you presented me with a set of box pairs, how did you know which experiment I would then choose?" Dr. Elbe responds: "I did *not* know which experiment you would choose. Your conscious choice *created* the particular situation of the marble in its box pair. You have just seen consciousness displayed as a physically efficacious entity beyond its neural correlates, what we call psychokinesis."

You are sure there's trickery involved. After all, Dr. Elbe's demonstration involved more than your conscious intent. Perhaps the mechanical opening of the box pairs, either sequentially or simultaneously, somehow physically put the marble wholly in a single box or spread it over two boxes. Therefore, with your unlimited resources, you bring in a broad-based team of scientists and magicians (illusionists) to investigate Dr. Elbe's demonstration. However, after their investigations, which you accept as exhaustive, they report there to be no trickery and that no *physical* explanation could be found for your method of opening the boxes to affect what was in them.

Psychokinesis is presumably impossible. Dr. Elbe's demonstration cannot actually be done. But if (*if!*) it could, you would be compelled to accept it as at least objective *evidence* that conscious choice itself could affect a physical situation, that consciousness existed as an entity beyond its neural correlates. In the actual quantum demonstration, the argument involved in the simultaneous opening of box pairs is a bit trickier, but it is almost as compelling.

The Quantum Demonstration

The two-slit interference experiment, for which our parable was an analogy, is described in every quantum physics textbook. It is often done as a lecture demonstration. In the usual two-slit experiment, a stream of small objects impinges on a diaphragm containing two openings. Most commonly

the objects are photons, electrons, or atoms, but quantum theory places no limit on the size of objects used.[3] To be general, we speak of "objects."

You could choose to observe individual objects, see through which opening each object came, and thus show that each object came through a single opening. On the other hand, by allowing the objects to pass through the openings without being observed, it is possible to show that each and every object came simultaneously through *both* openings.

The two-slit experiment is the *archetypal* quantum demonstration of "superposition," an object existing in two seemingly contradictory situations at the same time. However, almost *every* application of quantum mechanics exhibits this feature. Lasers have atoms simultaneously in two energy states. Transistors have electrons concentrated in one place and simultaneously spread throughout the crystal. MRI machines have protons with their north poles simultaneously pointing up and down. The quantum phenomena displayed in the two-slit experiment are ubiquitous.

We will present the empirical facts of the two-slit experiment in a theory-neutral manner. Again, by "theory-neutral" we mean that our description avoids any reference to the quantum *theory*. The point of the theory-neutral treatment is to emphasize that the objective evidence for consciousness can arise *directly* from empirically demonstrable facts. The usual treatment, introducing theoretical constructs such as the wavefunction, can mask this evidence.

We offer an intuitively compelling version of the two-slit experiment that is completely equivalent to the standard diaphragm-with-two-openings experiment referred to above. In this version, objects are sent one at a time to impinge on a "semi-transparent mirror," a sheet of material that has a fifty percent chance of allowing the object through and a fifty percent chance of reflecting it. A glass plate, for example, can be a semi-transparent mirror for light, which is a stream of photons. We can create semi-transparent mirrors for other objects.

A semi-transparent mirror, a fully reflecting mirror, and a pair of boxes are arranged as shown in Figure 1. Our objects are sent into this arrangement one at a time from the left, each object toward a new pair of boxes. What happens to objects on their encounter with the semi-transparent mirror? Do they sometimes go through toward the bottom box and sometimes get reflected off both mirrors toward the top? Or do they split at the semi-trans-

[3]In principle, quantum theory applies to baseballs as well as to atoms. For technical reasons demonstrations are limited to small objects. But the interference experiments we describe are today being done with increasingly large objects such as seventy-atom molecules. Similar quantum phenomena are now confirmed for structures consisting of millions of atoms.

parent mirror to go partly into each box? The experiments we will describe only involve the final situation of the objects in the boxes and do not involve observing their path. Therefore, in keeping with our theory-neutral description, what happens at the semi-transparent mirror need not be specified at this point. Knowing the speed of our objects, we know when each will enter the region of its boxes. The open doors of the boxes are then closed capturing the object.[4]

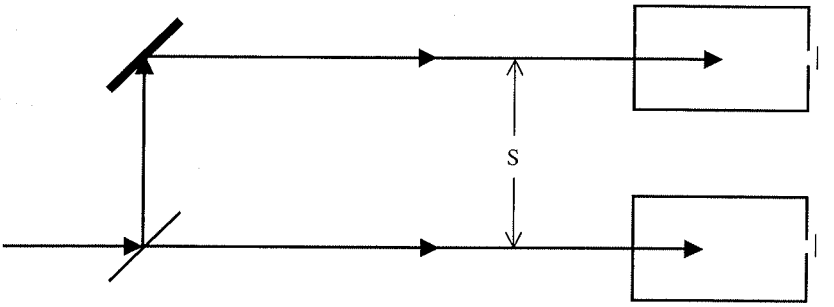

Figure 1: Schematic diagram of the box-pair experiment. The thin diagonal is the semi-transparent mirror. The bold diagonal is the fully reflecting mirror. "S" indicates the spacing between the boxes. Each object is sent in from the left, encounters the mirror arrangement, and moves to enter its box pair. The box-pair doors are closed at the time appropriate to trap the object.

We collect a set of box pairs, each pair containing a single object. In our first experiment (one analogous to Dr. Elbe's sequential-opening experiment) you choose to look in each box of a pair sequentially. Each time you find a whole object in one of the boxes of each pair, and you find the other box of that pair to be *completely empty*. You thereby demonstrate that each object had been *wholly* in a single box.

With another set of box pairs, you choose a different experiment (one analogous to Dr. Elbe's simultaneous-opening experiment). You choose to open both boxes approximately simultaneously. Here, however, the actual quantum experiment must differ from that of our parable. We must rely on circumstantial rather than direct evidence because we never see partial objects like Dr. Elbe's split marbles. Thus, to determine the situation of the object in its box pair, you simultaneously open small holes in the right-hand

[4]Holding our objects "gently" enough in physical boxes to accomplish the demonstration we describe is definitely possible, though difficult for objects other than photons. But talking this way is nice conceptually. In fact, our "boxes" need not be actual physical boxes; they need only be defined regions of space.

side of each box of each box pair to allow the object to emerge and impinge on, and stick to, a screen on which only *then* may it be observed.[5] You position subsequent box pairs identically and repeat the simultaneous openings.

When you examine where on the screen the objects have landed, you find that some places on the screen have many objects, and some have none. The objects are concentrated in bands. Moreover, by repeating the experiment with different spacings of the box pairs, you discover that *the spacing of the bands depends on the spacing between the box pairs* (the distance "S" in Figure 1). Each and every object coming from its box pair followed a rule allowing it to land only in certain places. Since that rule depended on the spacing of the box pairs, each object "knew" that spacing. (The spacing can thus be deduced from where the objects land.) Something of each object thus had to have been in each box of its pair.

By this second experiment you demonstrate that each object was *not* wholly in a single box. We emphasize that this display of what is called "interference" is accepted in physics as a demonstration that each object came from more than one source.[6] In the case of the standard two-slit experiment, each object came through both slits. In the case of our box pairs, it came out of both boxes; it thus had to have *been* in both boxes.

These two situations, each object wholly in a single box of its pair and each object spread over both boxes of its pair, are contradictory. Your conscious choice of what to demonstrate creates *either* of two contradictory prior physical conditions for the objects. Of course, in addition to your choice of experiment, an actual physical opening of the boxes, either sequentially or simultaneously, was required to produce the different physical situations. We must rule out the possibility that the particular method of opening the boxes exerted physical forces to bring about a particular situation. Could, for example, the opening of one box and, say, finding it empty exert a physical force *putting* the object wholly in the other box?

In fact, no investigation can distinguish the situation of the box found empty from the situation in which no object was sent into the box pair in the first place.[7] The search for a physical explanation, a physical force, would have results analogous to the investigations of Dr. Elbe's demonstrations.

---

[5]Observing an object on its path to the screen would be equivalent to the first quantum experiment.

[6]See the discussion of interference in any introductory physics text (e.g., Knight, 2004). Such an "interference experiment" has also been described in an earlier paper in this journal (Rosenblum and Kuttner, 1999).

[7]See the discussion of the "collapse" of the wavefunction in any quantum physics text (e.g., Griffiths, 1995).

As a rather dramatic example of the effect of the choice of what experiment to do, what knowledge to acquire (i.e., which box the object was in or the box-pair spacing) consider this experiment. With a set of box pairs, look in a single box of each pair. About half the time you will find an object. Discard those box pairs for which an object was found. With the remaining box pairs, for which the object was not physically disturbed, attempt an interference experiment. You will find no bands; the distribution of objects will be uniform. By having looked in the empty box of those box pairs, you acquired information that the object was in the other box. Acquiring which-box knowledge, *by any means whatsoever*, influences the behavior of the objects, even though the objects were not disturbed by any physical force.[8]

A comment on experimental methodology: it is, of course, possible that the reported conscious *intent* of which experiment to do did not correspond to the experiment actually done. (You pushed the wrong button.) In the case of those "mistakes," nothing is demonstrated, and such trials must be excluded from the data or treated as experimental error. The *reported* conscious intent might also have been a lie. No experimental result is ever immune from deceit.

Let us restate the problem with which we are left: every time you chose to open the boxes sequentially, you establish that each object had been wholly in a single box. Every time you chose to open the boxes simultaneously, you establish a *contradictory* situation, that each object had been spread over both boxes of its pair. If it was not the physical opening of the boxes that can explain the object's prior condition, what explains the strange correlation between the experiment you *chose* to do and the particular situation you demonstrate?

Here is a *conceivable* explanation, but one hard to accept. It is that our world is a totally deterministic one in which you did not have the free will to *choose* either one experiment or the other. Whenever a box-pair set whose objects were wholly in a single box was presented to you, you *had* to choose a sequential opening. And presented with a box-pair set with spread-out objects, you *had* to choose to open the boxes simultaneously. However, for this to work, the world had to conspire that your choices be correlated with the nature of the objects in the box pairs. A deterministic world is not enough. It must be a conspiratorial one.

---

[8]One problem with any such force is that it would have to propagate faster than the speed of light, in violation of special relativity. For example, opening one box and finding it empty would *instantaneously* ensure that the object was totally in the other box no matter how far apart the two boxes of the pair were — even though you presumably *could* have instead chosen to prove the object was distributed over both boxes.

An Objection: The Robot Argument

The most common objection to the quantum experiment as evidence for the involvement of consciousness is to claim that a *not*-conscious observer, a robot, could do the experiment as effectively as a human. The argument might go like this: with each set of box pairs, a robot could randomly do either a look-in-a-box experiment (sequential openings) or an interference experiment (simultaneous openings) and print out a report of its results telling whether the objects in a particular set of box pairs were each concentrated in a single box or were distributed over two.[9] Since the robot's printout would be indistinguishable from one presented by a conscious observer, the not-conscious robot qualifies as an observer.

Does this argument work? Let's consider the robot-performed experiments from a *human* perspective, our only meaningful perspective. You are given the robot's printout. It indicates, for example, that box-pair sets 2, 5, 7, 8, 11, and 13 contained objects wholly in a single box, and sets 1, 3, 4, 6, 9, 10, and 12 contained objects distributed over both boxes. That means the robot did a look-in-the-box experiment on the former group of box-pair sets and an interference experiment with the latter. In *itself*, the robot's printout displays no evidence for conscious involvement. Receiving the robot's printout, you could assume that the objects in the first-mentioned sets of box pairs were indeed wholly in a single box, and those in the other sets were distributed over both boxes. You could assume that the sets were *prepared* that way.

However, if the box-pair sets presented to the robot were indeed different in this way, how did the robot "decide" to do the *appropriate* experiment with each box-pair set? (For example, every time it did a look-in-the-box experiment it found an object wholly concentrated in a single box.) Since the robot had no information about the box pairs, its choice of experiment could be random. You investigate and find, indeed, that the robot's decision was made by a coin flip. Heads, it did the look-in-the-box experiment, tails, the interference experiment. However, the supposedly random landing of the coin mysteriously corresponded to the supposedly unknown nature of the object in the box pairs. You therefore replace the coin flip by the thing you are most sure is *not* determined by what is in the box pairs, your own free choice. You push a button telling the robot which experiment to do. You are now back to the original situation, and conscious choice is involved.

[9]To avoid the complication of the robot becoming entangled with conscious observers, we assume that, other than by its printout, the robot is isolated from the rest of the world.

The Quantum Theory and its Interpretations

We now briefly discuss the quantum theory relevant to the experiment we described, which until now we have treated from a theory-neutral viewpoint.[10] In quantum theory, an object is not only *described* as a wave, it *is* a wave, called a "wavefunction." No object in addition to its wavefunction is presumed to exist. Just as a light wave or a water wave can be split into two or more parts to exist in different regions, so can the wavefunction of a single object.[11] In the case of our mirrors and box pairs, the wavefunction of each object splits at the semi-transparent mirror and is captured in a box pair. In an interference experiment, the wavefunction comes simultaneously out of both boxes of each pair. Parts of the wavefunction from each box come together and "interfere" at the screen. That is, at some places on the screen, crests from one box arrive together with troughs from the other, canceling each other to produce regions of zero waviness. At other places on the screen, crests from both boxes arrive together reinforcing each other to produce regions of maximum waviness. Such interference is accepted as establishing wave phenomena.

With the quantum theory we can calculate the wavefunction for a given situation. But we never actually *see* a wavefunction. A crucial *postulate* connects the calculated wavefunction to what is observed. Namely, the waviness in a region is the probability of *finding* the whole object in that region.[12] Waviness is *not* the probability of the object having *been* there immediately before being observed there. "Finding" an object in, say, a particular box means experiencing evidence that its waviness is concentrated there — by bouncing light off it, for example. But *before* you bounced the light off it, its waviness, and thus the object itself, had to have been equally in *both* boxes simultaneously. You could have chosen to establish that fact by an interference experiment. That is, you could have chosen to establish either of two contradictory results. This dichotomy is accounted for in quantum theory by accepting that the choice of the type of observation *creates* the type of result

[10]Somewhat more extensive descriptions of the quantum theory specific to this experimental set-up are available (Rosenblum and Kuttner, 1999, 2002, 2006).

[11]Strictly speaking, the wavefunction is not a wave in ordinary three-dimensional space, but resides rather in a mathematical realm, a Hilbert space. But for the position wavefunction of a single object, a representation in ordinary space gives an adequate picture, and is the one generally presented in introductory quantum mechanics texts. The complete wavefunction of an object includes *all* its properties (velocity, spin, energy, etc.). We just discuss the part related to the object's position.

[12]Mathematically, what we here call "waviness" is the absolute square of the wavefunction. But "waviness," how high the crests and deep the troughs in a region, gives the reader the general idea.

observed. (This choice is the point at which the theory encounters the issue of consciousness.)

Even today, with quantum theory in its eighth decade, many practitioners of the theory admit that, taking what the theory says seriously, they find it hard to believe, or at the least they admit to not fully understanding its implications. We emphasize, however, that quantum theory is the most battle-tested theory in all of science. Never has a single one of its vast number of predictions been shown even the slightest bit in error. Some predictions have been shown accurate to parts in a billion. Quantum theory is the underlying basis of all physics. One third of our economy involves products requiring quantum theory in their design.

But a hard-to-believe theory requires interpretation. Today, contending interpretations try to tell what quantum mechanics reveals about the nature of our world. Interpretations of the theory often dismiss the *concern* with consciousness from the physics discipline. Such a dismissal, separation at least, is not inappropriate since physics seems to have come to a boundary of the discipline where the expertise of physicists is no longer uniquely relevant.

In the last few paragraphs discussing quantum theory, we have implicitly assumed the Copenhagen interpretation. It is the physics discipline's original and still-orthodox stance.[13] In its usual version, the world is divided into microscopic (atomic-scale) and macroscopic (human-scale) realms. Properties of microscopic objects, and thus the objects themselves, are not physically real until their observation. Quantum probability "collapses" to an observable classical actuality as soon as a microscopic property affects a macroscopic object.

Since for all practical purposes physicists need only report the behavior of their classical instruments, they can consider the objects of the microscopic realm as mere models whose strange behavior involving conscious choice need not be of concern. The Copenhagen interpretation was early on criticized by Einstein as being a tranquilizer, not a solution. Gell–Mann, in his Nobel Prize acceptance speech, claimed the Copenhagen interpretation brainwashed two generations of physicists into thinking the problem of conscious observation was solved. The interpretation has recently been summarized as "Shut up and calculate!" Nevertheless, it represents a convenient working attitude for all *practical* purposes, and essentially all physicists adopt it in our teaching and in our practical application of quantum theory.

In fact, however, in a mathematically rigorous treatment, sometimes considered a version of the Copenhagen interpretation, von Neumann (1932/1955) showed that no system obeying quantum theory could collapse a wavefunction.

[13]A good review is given by Stapp (1972).

He therefore concluded that an *ultimate* collapse, a probability becoming an actuality, could only take place at the point where quantum theory no longer applied — at conscious observation.

The Copenhagen interpretation depends on a clear boundary between the microscopic and macroscopic realms. Today the boundary blurs as interference is demonstrated with large molecules and quantum phenomena are displayed in structures involving millions of atoms. Interpretations of quantum theory competing with the Copenhagen interpretation proliferate.

To deal with the blurring of the boundary between the micro and macro realms, the process by which a wavefunction continuously distorts, or "decoheres," on contact with a macroscopic object is studied (Zurek, 1991). Since the possibility of displaying interference rapidly disappears with such contact, the situation appears classical, for all *practical* purposes. Therefore, though the question of the *ultimate* observer admittedly still remains (Zurek, 1999), physics need not worry about it.

The "many worlds" interpretation (Everett, 1957), which is also called "many minds," accepts quantum theory at face value. Looking into a box, you, and the rest of the world, bifurcate. In one world, one "you" is conscious of the object in the looked-in box. In another world, another "you" is conscious of that box being empty and the object being in the other box. Moreover, in this interpretation, you made *both* the choice of looking in the box *and* the choice of doing an interference experiment. In another sense, you made no choice at all; you actually did everything you possibly could have done.

David Bohm (Bohm and Hiley, 1993) developed an interpretation in which objects making up the world exist in *addition* to their wavefunctions. Objects are guided by a not-detectable "quantum potential" much as a ship is guided by a radio beacon. This interpretation presents a *completely* deterministic worldview, one that does not exclude the conscious observer, but avoids dealing with consciousness, for all *practical* purposes. In yet another recent interpretation, David Mermin (Mermin, 1998) has physics concerned only with "*physical* reality." Consciousness resides in a larger reality beyond this physical reality.

Conclusions

Physics' encounter with consciousness was recognized as unavoidable almost at the inception of the quantum theory (von Neumann, 1932/1955). In the theory's fourth decade, Eugene Wigner (1961/1983), a major contributor to the theory, claimed that it is ". . . not possible to formulate the laws of quantum mechanics in a fully consistent way without reference to the consciousness" (p. 169). Discussion of quantum theory's implications for consciousness increases today — and remains contentious.

The archetypal quantum experiment, the two-slit experiment, or the boxes version we discussed, demonstrate that a conscious choice can bring about *either* of two contradictory prior physical realities, and no physical force can be detected as responsible for bringing the selected one about. In accepting this as evidence for the existence of consciousness as an entity beyond its neural correlates, one assumes that there was the freedom to choose a particular observation method. However, rejecting this assumption of free will requires a conspiratorial world as well as a deterministic one. Extraordinary claims require extraordinary evidence. That consciousness has a direct physical effect would surely be an extraordinary claim. Whether or not the quantum demonstration is *sufficiently* extraordinary as evidence, it seems to be the *only* objective evidence for consciousness beyond its neural correlates.

In working with physics or teaching physics, physicists pragmatically accept "scientific realism," defined by the *Dictionary of the History of Science* (Bynum, Browne, and Porter, 1981, p. 362) as ". . . the thesis that the objects of scientific knowledge exist and act independently of the knowledge of them." The quantum experiment appears to tell us that the nature of reality is *not* that of scientific realism. Has it been discovered that the objects of scientific knowledge *do not* exist and act independently of the (conscious) knowledge of them? With quantum mechanics we have encountered something *beyond* the normal boundaries of our physics discipline. We have described a profound problem involving consciousness, the quantum enigma. We do not suggest a solution.

We find it hard to even imagine a solution. But we suspect that one would profoundly change the way we view the world — and our place within it. John Bell, perhaps the leading quantum theorist of the latter half of the twentieth century, wrote it is likely ". . . that the new way of seeing things will involve an imaginative leap that will astonish us" (1980, p. 27).

References

Bell, J.S. (1980). *Speakable and unspeakable in quantum mechanics*. Cambridge: Cambridge University Press.

Bohm, D., and Hiley, B.J. (1993). *The undivided universe*. London: Routledge.

Bynum, W.F., Browne, E.J., and Porter, R. (Eds.). (1981). *Dictionary of the history of science*. Princeton, New Jersey: Princeton University Press.

Chalmers, D.J. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

Crick, F. (1994). *The astonishing hypothesis*. New York: Scribner.

Everett, H. (1957). "Relative state" formulation of quantum mechanics. *Reviews of Modern Physics*, *29*, 454–462.

Griffiths, D.J. (1995). *Introduction to quantum mechanics*. Englewood Cliffs, New Jersey: Prentice Hall.

Knight, R.D. (2004). *Physics for scientists and engineers*. San Francisco: Addison Wesley.

Mermin, D. (1998). What is quantum mechanics trying to tell us? *American Journal of Physics*, 66, 753–767.

Rosenblum, B., and Kuttner, F. (1999). Consciousness and quantum mechanics: The connection and analogies. *The Journal of Mind and Behavior, 20*, 229–256.

Rosenblum, B., and Kuttner, F. (2002). The observer in the quantum experiment. *Foundations of Physics, 32*, 1273–1293.

Rosenblum, B., and Kuttner, F. (2006). *Quantum enigma: Physics encounters consciousness*. New York: Oxford University Press. (in press)

Stapp, H.P. (1972). The Copenhagen interpretation. *American Journal of Physics, 40*, 1098–1116.

Stapp, H.P. (2004). *Mind, matter and quantum mechanics*. New York: Springer Verlag.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton, New Jersey: Princeton University Press. (Originally published 1932)

Wigner, E. (1983). Remarks on the mind–body problem. Reprinted in J.A. Wheeler and W.H. Zurek (Eds.), *Quantum theory and measurement* (pp. 168–181). Princeton, New Jersey: Princeton University Press. (Originally published 1961)

Zurek, W.H. (1991). Decoherence and the transition from quantum to classical. *Physics Today, 44*, 36–44.

Zurek, W.H. (1999). Preferred states, predictability, classicality and the environment-induced decoherence. *Progress in Theoretical Physics, 88*, 282–312.

The Journal of Mind and Behavior
Winter 2006, Volume 27, Number 1
Pages 57–72
ISSN 0271–0137

Content Individuation in Marr's Theory of Vision

Basileios Kroustallis

Hellenic Open University

The debate concerning the individuating role of the external environment in propositional content has turned to Marr's (1982) computational theory of vision for either verification or disproof. Although not all the relevant arguments concerning the determining role of environmental constraints that Marr invokes in his visual account may succeed, the paper argues that Marr divides his computational explanation into two parts, the information processing "what" and the constraint introducing "why" aspect. It is the second part where separate claims concerning the necessity and sufficiency of constraints are advocated, and initiate a specific computational process. The above explanation becomes subordinate to a conception of inference that closely resembles deduction.

Recent advances in the explanation of visual processes (see for example, Biederman, 1993; Treisman, 1982; Treisman and Gelade, 1980; Ullman, 1979; Zeki, 1993) made visual theory a fruitful empirical field to test major controversies regarding content in the philosophy of mind, and more specifically the case for wide vs. narrow content. The existence and the nature of a propositional content of mental states is itself a debated issue, with different positions being articulated both from the side of intentional realists (Cummins, 1989; Dretske, 1981; Fodor, 1987; Millikan, 1984) and the antirealist camp (Churchland, 1981; Churchland, 1986; Stich, 1983). However, the counterfactual thought experiments of Putnam (1975) and Burge (1979) concerning concepts and their use in language have introduced a content of beliefs and desires which would presumably change as a result of environmental change ("wide content"), even though the intrinsic properties of an agent ("narrow content") would remain the same (cf. Bach, 1998).

Alternative proposals regarding a two-component (narrow and wide) theory of content (Block, 1986; Field, 1977) or a complete disagreement on

Requests for reprints should be sent to Basileios Kroustallis, Ph.D., 10, Agiou Dimitriou St., Corfu 49100, Greece. Email: basilisphil@yahoo.com

the challenge that counterfactual considerations are supposed to present (Crane, 1991) have also been expressed. Fodor (1987) tied his narrow content account with the current scientific practice, and argued that content determination by intrinsic causal powers is the best strategy for a science of psychology, since causation is a determining factor in other sciences as well, therefore the appeal to wide content and external environment would be proved unscientific. On the other hand, Burge (1986) invoked a prominent scientific work on perception, namely David Marr's (1982) computational theory of vision, which constructs vision as a series of visual representations from early image stages to object recognition. Burge claimed that this theory is intentional, and that it also individuates content by means of external environmental features and constraints, not by intrinsic causal properties.

It is Marr's theory of vision that the present paper examines regarding the existence and the nature of content attribution. The first section discusses already expressed arguments for and against the individuating function of visual computational states by means of their representational content, and concludes that while constraints are not purely expository and should play a role in individuation, a different approach is necessary. This is developed in the second section, where two explanatory parts are distinguished in Marr's computational level, the "what" and the "why" aspect: the second factor reveals his purpose that individuation of computational states according to environmental facts should be interpreted as the attempt to instantiate constraint necessity and sufficiency claims, and the latter claim defines a distinct computational process. The last section further examines this interpretation, and it is proposed that Marr's individuation account is not an outcome of his attempt to successfully mirror the veridical results of visual processes, but relies upon the interpretation of the inference from the visual image to the structure of the external world as a kind of deductive inference.

Content Attribution in Marr's Theory

Marr's computational theory of vision has been adequately described elsewhere (Gilman, 1996; Kitcher, 1988). Nevertheless, there are a few points that are worth repeating. First, it is important to note that, inside the information processing approach, vision is decomposed into a plurality of sequential tasks that lead to distinct data structures, called *sketches* (Marr, 1982, p. 42). There is a selection of sharp intensity discontinuity points in the computational image (called zero-crossings), which are considered to correspond mainly to object boundaries in the visual scene. Different patterns such as a row, a blob, etc., are located in the computational image array, and the whole scheme that emerges comprises what Marr calls "the raw primal sketch." A consequent grouping of those elements according to specific prop-

erties (such as similar size or local distance, etc.) results in the formation of larger unified areas with more prominent boundaries between them, a structure named *full primal sketch*, the end of image representation.

A series of parallel computational processes (using motion, depth and texture cues) operates on the previously created two-dimensional image representations, and computes the shape and orientation of the three-dimensional surface seen from a particular point of view. The resulting construction is the *$2^1/_2$ D sketch*, a viewer-dependent three-dimensional representation. A final stage is proposed to account for the further fact that observers have the ability to recognize an object shape independent of the particular viewing angle (Marr, 1982, p. 295ff.). The $2^1/_2$ D sketch is compared in that stage with an index of already memory-stored 3D object models. If that sketch agrees with a certain model in terms of similar distinct volumetric units (which represent specific object parts), then object recognition has been attained.

Although these consecutive stages usually denote the computational part of an information processing approach, Marr (1982, pp. 24–27) divides this approach into two levels of description, the computational and the algorithmic. The first level describes "what" is computed and "why," by specifying the relevant constraints that define a particular operation. Specific environmental facts work as constraints and determine each visual process, the way addition has particular rules that separate it from multiplication. On the whole, computational theory can supply the justification for the selection of the most appropriate algorithm among various candidates, which initiates a computational process and gives the desired computational outcome. The unfolding of those processes takes place in the so-called algorithmic level, where various computational algorithms are tested against the prescribed theoretical criteria, so that they will appropriately transform the input into the desired output. The effectiveness of computational processes is checked against another level of description, the implementation level, where the neurophysiological facts on vision are presented. Marr (1982, pp. 15, 336) warns that vision cannot be understood only on the level of neurons, although at the same time neurophysiology may be a separate, important part in any explanation of vision.

There seem to be two contrasting elements in the exposition of the computational level. On the one hand, Marr describes his theory in exceedingly precise information processing terms, in a way that does not allow this side of his computational theory to be easily dismissed:

> [In the computational theory] the performance of the [information processing] device is described as a mapping from one kind of information to another; the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task demonstrated. (Marr, 1982, p. 24)

> [T]he computational theory . . . is determined solely by the information processing task to be solved. (Marr, 1982, p. 337)

However, when Marr deals with the specific visual stages, he devotes much space and effort (1982, pp. 44–51) to analyze constraints derived from the visual environment, such as the existence and hierarchical organization of surfaces or spatial continuity (the fact that markings on a surface are often spatially organized into lines or other patterns). If there is presumably nothing more to be stated about the computational level other than a information processing account, then the above insistence on constraints seems either superfluous or in need of interpretation.

There are two kinds of proposals that have been expressed to address this case. Burge (1986) interprets this second part by claiming that Marr's computational theory is intentional. The purpose of positing this theory is not only to find a set of algorithms that operate in the computational image, but also — and more importantly — to describe the structure of the visual world by means of the content of computational representations. He further states (1986, p. 29) that this content is determined by means of "specific causal distal antecedents in the physical world" [e.g., edges], and it also includes assumptions about "contingent facts regarding the subject's physical environment" [constraints]. While the number of premises and the exact structure of the above argument has been a matter of discussion (see Bontly, 1998; Patterson, 1996; Shapiro, 1993), it seems nevertheless straightforward that Burge attributes to Marr's external constraints an individuating function that in turn determines the content of subsequent computational states.

On the other hand, Egan (1992, 1996) objects against the idea of distal environmental features and constraints in the computational level as individuative, and states that constraint mention is nothing but a "function–theoretic" description of the process, a "formal characterization of the function(s) computed by the various processing modules" (1996, p. 236). Her textual support involves Marr's discussion on edge detection as a mathematically-defined function (Marr, 1982, pp. 336–337). Egan (1992, p. 445) not only makes this whole process syntactic, but also proceeds to describe a "realization function fR which maps equivalence classes of physical features of a system to what we might call 'symbolic' features" (p. 445), where computational states supervene on brain features.

At the same time, Egan (1996) does not eliminate a role for content in Marr's theory, but supports the thesis that wide content assignment is an explanatory, or epistemic, and not an individuative factor. Accordingly, she proposes an "expository" notion of content, where content ascription may explain computational states in an information processing account, in the same way that models in the physical sciences explicate mathematically-

defined processes. This kind of exposition would also accompany the operation of computational processes by answering questions formulated in intentional terms, such as the problem of transition from a 2D image to a 3D object representation. Those considerations would constitute the function of content as explanatory or methodological (see also Francescotti, 1991; Bontly, 1998, for a similar characterization; or Morton, 1993, for the statement that semantic considerations answer the metatheoretical claim why a certain theory is successful).

Regarding the first part of Egan's account, consequent attempts to refute the syntactic character of computational states rely upon the role of neurophysiology in Marr's theory of vision, but their reasons for content attribution do not seem conclusive. Bermudez (1995) questions the sufficiency of the equivalence relation, and inquires whether all neural events may be adequately represented by computational structures. He claims that not all neural inputs in Marr's theory — such as visual noise — are causally demarcating, and not all causally demarcating properties (e.g., various neuroanatomical features) are functionally (computationally) relevant. He concludes that external constraints are invoked in order to remedy this computational inadequacy of neuronal operations. However, even if the relevant facts hold for Marr's theory, that conclusion does not necessarily follow. In the cases above, Marr thinks (1982, p. 336) the implementational level itself can determine which neuronal inputs are causally demarcating. Noise can be removed from computational outcomes on the algorithmic level by attending to what the retina does, not by appealing to the computational level. At the same time, the reason that some neuronal properties (such as light transduction) are causally demarcating but computationally irrelevant is that they cannot syntactically conform to an information processing account, without any evidence that lack of constraints is responsible for the exclusion.

Similarly, Shagrir's (2001) claim that appeal to content is needed to determine an otherwise underdetermined neuronal function seems to conflict with other elements in Marr's theory. Although this view correctly emphasizes the interaction of the implementational with the computational level towards the better explanation of specific computational processes (cf. Marr, 1982, pp. 17, 336), the preference over the implementation level does not have to stem from the fact that neuronal activity is functionally underdetermined: neuroscience can describe a single function that is attributed to neuronal activity, for example the function of being a hand-detector cell describes adequately that cell's activity (p. 15). There is no intrinsic neural problem that needs to be solved by an appeal to environmental features and constraints. There is a problem of algorithm determination (equated with precise task description), but this has to be solved in the computational level (pp. 122, 208), where the relevant constraints are located. Therefore, although

Marr might invoke content in the characterization of the computational states he describes in his theory, a different motivation is needed for that appeal.

In contrast, regarding the second part of Egan's account in which content is described as an expository construct, although it may be appropriate in general for an information processing account to be accompanied by an exegesis in non-computational terms, the presence of such a theory cannot be verified by examining Marr's work. The information processing account in the computational level *itself* explains what needs to be explained, namely neurophysiological function. There is no articulation of a second, intentional characterization of the explanatory, computational level. Mention of environmental features by itself does not imply the existence of a second interpretation of visual phenomena in intentional terms. Egan's second claim that content serves to bind different modules together in the 3D representation misses the fact that this is accomplished by the construction of the $2^1/_2$ D sketch, another outcome of the information processing approach. It is true that viewer-centered shape (surface orientation) has to be constructed out of shading, depth and other visual features. Nevertheless, in the processes that lead to the $2^1/_2$ D sketch there is neither a more frequent nor a more distinct mention of distal environmental features, any more than in the other computational stages, something that indicates that constraints might be systematically used for a quite different purpose.

Two Aspects of Computation and the Necessity and Sufficiency Claims

In the existing accounts of Marr's theory of vision the implicit assumption is that both information processing tasks and content individuation by means of constraints are incompatible. Butler (1996, p. 149) has argued that individuation by means of content may be supplementary to individuation by computational properties, though the question is how this is accomplished. However, Marr gives evidence concerning the independence of those two factors, which arise from the division of the computational level into two parts, the "what" and the "why" account. Information processing details determine what a certain algorithm does, and Marr's example of understanding the process of addition is mastering the relevant mathematical theory, that is, the specific kind of mapping from, for example, 3 and 5 to 8. But this part by itself (the "what" part) does not entail knowledge of any relevant constraints. These turn out to matter when there is the need to understand why the system computes addition and not multiplication, and the relevant rules (the rules for adding zero, communicativity and associativity) are invoked, so that "the rules we intuitively feel to be appropriate for combin-

ing the individual prices in fact define the mathematical operation of addition. These can be formulated as constraints . . ." (Marr, 1982, p. 22).

This early distinction between information processing theory and knowledge of constraints using the analogy of addition is reinforced in the explication of visual processes. When Marr describes the process from a complete image to an object surface, he distinguishes two "aspects" of the relevant computational theory — the one which transforms the input properties of the visual image into the output properties of a 3D surface (the equivalent of the mapping of a pair into another number in addition), and the discovery and employment of new physical world constraints that define uniquely the task in question. This distinction also explains the apparent inconsistency between the important role that constraints seem to play in edge detection and Marr's later review comment on that process, where mathematically computed functions are "from a computational point of view, [. . .] a precise specification of what the retina does" (1982, p. 337). Although Egan (1996, pp. 236–237) has taken this to imply that constraints do not function in an individuative way at all in edge detection, the "precise specification" of retinal function is not meant to exclude constraint individuation, but to work independently so as to disallow any potential description of neurophysiological functions (light transduction, fovea functions) that the human retina performs.

Nevertheless, even if the above considerations are correct — that external constraints intentionally characterize computational states in the "why" part of computational theory — this observation left unsupported could only entail that computational outcomes are externally individuated simply by methodological stipulation. Although it might be stated that constraints define uniquely a certain computational task, the logic and the application to the particular computational processes employed have to be shown, so that the distinction between the what and the why function may be consolidated.

For that purpose, Marr advances two separate claims concerning the latter function and the involvement of constraints in vision. He first advocates constraint necessity, which he finds a natural consequence of the reflection on visual processing, for example, every theory of vision should respect the fact that the perceptual world is composed of smooth surfaces (1982, pp. 44, 51, 115). It seems that Marr assumes that perceptual processes depend on environmental facts, and then appeals to a consequently assigned intentional character of any theory of vision — although it is not certain that he would also claim that the distal environmental conditions determine the retinal image formation, a kind of perceptual externalism (cf. Butler, 1998).

Nevertheless, necessity of constraints by itself implies dependence but no individuation of computational states by means of those constraints. So,

Marr advances constraint *sufficiency*, which is responsible for individuating computational states, since it leads to a distinct computational process determined completely by those constraints. His aim is reflected in the propositional form of the assumption he uses: if the relevant constraints are satisfied, then the process under examination is physically (perceptually) correct. Consequently, if there can be a distinct computational process defined in the computational level, which leads to a certain solution in accordance with the relevant constraints, then the assumption is true and the sufficiency claim will carry the weight of the individuating process.

A first example of a constraint-introduced process is the combination of zero-crossings from different filters, which gives, apart from the initial detection of intensity discontinuities, the computational result *edge*. Motivation for the initiation of this process comes initially from the thought that intensity changes must be somehow spatially localized. Marr (pp. 68–69) gives independent cases of ordinary perception, such as scratches or shadows, to prove the necessity of the spatial localization constraint. That constraint is transformed into the *spatial coincidence assumption:*

> If a zero-crossing segment is present in a set of independent $\nabla^2 G$ channels over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, orientation, depth, or surface orientation). [Marr, 1982, p. 70]

The assumption argues that postulation of constraints suffices to guarantee the presence of an intensity change in the visual scene. This invokes a computational process which takes place in the information processing part, a selection of zero values in different sized filters under certain rules (same position, orientation), which would represent the same intensity change. In effect, zero-crossings are grouped as edges, since they have the property of being in the same place in a combination of different filters. If the assumption of spatial coincidence (incorporating the constraint requirement of spatial localization) were absent, neither would that specific filter combination arise.

This seems to agree with Burge's (1986) counterfactual claim that if the subject were in an environment where "the properties and relations that normally caused visual impressions were different from what they are, the individual would obtain different information and have visual experiences with different intentional content" (p. 35). And, while it may be difficult to judge whether later processing stages are individuated according to environmental facts or not — for additional environmental assumptions have also affected prior stages of representation construction (Marr, 1982, pp. 104, 276) — the case of edge detection through zero-crossings seems to be the stage where

environmental constraints are used for the first time to determine a computational process.

Nevertheless, Segal (1989) argues, against Burge (1986), that the term *edge* in Marr's theory may potentially describe a multitude of physical phenomena, such as shadows or cracks. He assigns a form of an abstracted content (crackdow), which can denote either cracks or shadows, to describe these representational elements. According to this interpretation, Segal claims, no counterfactual case without constraint involvement may arise, for the term *edge* covers all possible physical phenomena. The only possibility would be a world in which edges were not perceptually present, but in this world the subject would forever suffer visual illusions.

The structure of the argument seems to be that, since the term *edge* does not differentiate between different physical phenomena, the resulting representational states will be the same, no matter what the environmental facts may be. This provides Segal's liberal interpretation: different physical phenomena are merged as a generic representational content. However, although the term *edge* is used to denote a variety of physical phenomena at different points in the image array, this does not necessarily imply that it is used to denote a variety of (or an abstraction over) physical phenomena *at the same image location*, and it is only in the latter way that individuation is attained. Computational states are considered representational in Marr's theory not because all intensity changes are called edges, but because all co-located intensity changes in different filters are called edges. It is exactly the case when zero-crossings cannot be found at the same image location in different filters, where merging of different physical phenomena is invoked, and no edges are detected. In a counterfactual world without the spatial coincidence assumption, while detection of zero-crossings would still be the first result, the resulting primal sketch, the whole pattern of edges and blobs would be different in different cases. Some of the sketch elements would be due only to a single physical phenomenon, while others might involve merging of different phenomena, and that would constitute an illusion. Consequently, vision would be less reliable in the counterfactual world but not totally illusory, as it would have been the case if absence of content coincided with absence of edge characterization.

Further evidence for the appeal to distinct computational processes in order to ground sufficiency of constraints comes from the middle part of Marr's visual theory, where various information sources (depth, motion, shading, etc.) help to recover shape information. Two main kinds of shape recovery that Marr himself distinguishes (p. 266) are stereopsis and structure from motion. Stereopsis denotes binocular depth information, which neurons and computational mechanisms recover by computing the relative difference in retinal/computational array object position between the two eyes (dispar-

ity). An exact point-to-point correspondence in the two images is a prior requirement for disparity estimation.

Marr invokes particular considerations, such as that matter is cohesive (so consequently object surfaces are generally smooth), to serve as constraints for correspondence, and proposes constraint necessity and sufficiency in order to solve the correspondence problem. The former claim, the fact that our visual perception obeys the relevant constraints, is "reasonable to infer" (1982, p. 115). On the other hand, the sufficiency claim states that if the correspondence process satisfies the three matching constraints employed, then that correspondence is physically correct, and the whole proposition is named "the fundamental assumption of stereopsis." Nevertheless, the latter claim seems less evident, and Marr seeks to verify it by reducing the problem to the one-dimensional case in his computational level, where all the relevant constraints have to be jointly and uniquely satisfied (p. 116). He states that his consequent proof completes the theory of stereopsis, and the specific algorithm used only applies this solution by creating excitatory connections in matching units from the two different images, and inhibitory connections where matching is not allowed.

Reflecting on the same process, Egan (1996, p. 242) claims that in environments where the constraints were not satisfied, the stereopsis module would still compute the same formally characterized function. This function, though, would not share the intentional description "depth from disparity," and that shows that constraints do not individuate computational states. Interestingly, Marr himself advances this comparison, not with regard to a counterfactual case, but addressing previous actual algorithms, in which some of the constraints were either absent or incompletely implemented. His comments initially seem to verify Egan's interpretation, stating that previously there were many *attempts* to compute correspondence, but "not one of them computed the right thing" (1982, p. 122). The above could denote that the formal characterization of the function (matching) remained the same, serving the same kind of computation, but the intentional interpretation (the right thing, depth) changed. However, the reason for that comment is not that previous accounts shared the same computational analysis and simply lacked appropriate constraint consideration. According to Marr, the specific matching they advocated was not the result of a previous computational analysis, in which constraint consideration would have to be prominent. Therefore their process did not even provide a formal characterization of the matching process, because no constraints were involved. In other words, Marr argues that either there is a computational analysis according to constraints or there is no computational analysis *at all*, that is why matching efforts are only the results of applying various algorithms, sometimes with partial success. For Marr, in a counterfactual world where no constraints operated, there would be

a process (matching) that could be described as a result of applying some algorithms, but not information processing, since this process does not have the prerequisites (external constraints) to be uniquely defined (that is, without alternative solutions) and, therefore, formally characterized.

The necessity and sufficiency of constraints claims are also present in the discussion of shape information from apparent motion. Marr makes use of the fact that most objects in the visual environment are rigid, something that works as a necessary constraint in deciding how to reconstruct object shape from motion. However, he also presents rigidity as sufficient for the reconstruction process by means of the following assumption, derived from Ullman (1979): "Any set of elements undergoing a two-dimensional transformation that has a unique interpretation as a rigid body moving in space is caused by such a body and hence should be interpreted as such" (as cited in Marr, 1982, p. 210). The way to prove that comes from Ullman's structure-from-motion theorem, which shows that if there is a rigid body in motion, there can be a computational way to find its three-dimensional structure from three distinct frames of the moving object. This proof is placed inside the information processing part of the computational account, and it is another instance of the fact that Marr employs a separate formal element (here, the structure-from-motion theorem) with direct influence to the corresponding algorithm to safeguard an assumption based on external constraints.

The Notion of Inference

The overall motivation for that particular strategy of content attribution still has to be addressed. Burge's argument (1986, pp. 32, 43–44) concerning the function of content in Marr's theory of vision is that his theory is success-oriented, and distal visual features and constraints are employed to explain the successful (veridical) interaction of agents with the world. However, as Patterson (1996, p. 260) notes, Marr only needs a general reliability of vision and not complete success to construct his theory, and this view seems to correspond to his references that the computational primitives of the image "have a high probability of reflecting physical reality directly" (1982, p. 71, cf. p. 99). Marr's attempts to minimize the epistemic status of the constraint necessity claim also echo this attitude. Constraints are deemed necessary in visual computational processes, because the purpose of computational vision is to describe visual perception, and those constraints are "generally true of the world" (p. 23) — therefore, a reliable indicator of veridical visual perception. The former belief reduces the need to seek a well-defined argument that supports constraint necessity. The underlying reliability framework in Marr's theory of vision covers specific perceptual features as well as constraint discovery.

But although Marr does not seem to invoke a belief in the complete infallibility of visual perception, it seems on the other hand that he considers as a pressing requirement to reconstruct correctly a theory of vision, most importantly at the computational level, in order to be able to reason precisely about vision, and not "in similes" (1982, p. 336). This brings his account close to the criticism of optimality (cf. Kitcher, 1988), the claim that Marr focuses on vision exclusively as a well-defined theoretical problem with a unique solution in the computational level, and thereby neglects less elegant but actual solutions of the same problem formulated by observing only the relevant neurophysiology. On the other hand, Gilman (1996, p. 301ff.) argues against that claim. He states that Marr does not specify beforehand all functional aspects of his proposed solution. Even when he espouses specific constraints, he takes account of the relevant neurophysiological data, and does not rely upon the notion of a unique solution in all of his computational processes. Consequently, Gilman argues that constraint introduction should be seen in terms of a heuristic search, since Marr invokes constraints which may satisfy the process to be described, but do not optimally define it.

The latter position would square with Marr's claim that constraints are necessary and generally true of the world, therefore they should be used in a computational account of vision. What is left unexplained by this interpretation is the sufficiency of constraints thesis, which points against Gilman's claim. Although there might be in practice a construction of the theoretical level according to neurophysiological data, the relevant processes are independently formulated in theory as problems to be solved *inside* the computational level. The fundamental assumption of stereopsis is found valid, because it can be shown theoretically that it leads to a unique solution, before any algorithms are tested. The rigidity assumption rests on the structure-from-motion theorem, and the spatial coincidence assumption entails selection of specific zero-crossings segments according to combination rules formulated in the top level, even though the details are specified in the actual algorithmic part. It might be true that Marr does not provide all the quantitative information that would ideally complete his visual processing proposal, and would totally characterize visual computational processes, constituting his account an optimal explanation of vision. However, at the same time, Marr certainly advocates a specific formalization by invoking constraint sufficiency, and this aspect needs to be explained.

The propositional structure of the assumptions that Marr uses has the form of premise to conclusion, reflecting the process from the image (antecedent) to the world (consequent). The fundamental assumption of stereopsis states that if the correspondence is attained in the computational process according to the relevant constraints, then that correspondence is correct, and similar

is the structure for the spatial coincidence and rigidity assumptions. This is coupled with Marr's description of his vision theory as a theory of inference:

> [T]he true heart of visual perception is the inference from the structure of an image about the structure of the real world outside. The theory of vision is exactly the theory of how to do this, and its central concern is with the physical constraints and assumptions that make this inference possible. (1982, p. 68)

The insistence on a kind of inference from image structures to world information has already been noticed (cf. Segal, 1989, pp. 193–194). It is not meant to be a top–down, general knowledge inference, since Marr's theory of vision is essentially a bottom–up process from elementary image representations to more complex descriptions. External constraints are also purported to be built in, not inferred from cognitive resources in general. On the other hand, the notion of inference as simply information processing (a sequence of particular processes that transform the input and produce the output in the algorithmic part) does not differentiate Marr's theory from other computational accounts, since he insists that environmental facts (which are not computational elements) have an important role in that inference.

If the above considerations are combined with the formulation of the various assumptions that attempt to achieve that solution, it seems that Marr equates the notion of inference with deductive inference. A computational solution according to specific constraints has to lead to the structure of the world, and his various specific assumptions are employed in order to reflect that structure. Although Marr does not talk of deduction, he seems to construct his framework in the form of modus ponens. He independently asserts the existence of constraints (the necessity claim), and then proposes the assumption that if the constraints are satisfied, then there is a correspondence with the visual world. The subsequent problem of the validity of the conditional is confronted by invoking a distinct computational process that will lead to the intended result.

This framework might also interpret the relative explanatory autonomy of the computational level from both the algorithmic and the implementational level. Information processing is not sufficient as an autonomy factor, for Marr proceeds beyond information processing to account for vision. However, if Marr constructs his computational account according to this overall logical construction and there is parallel absence of the latter construction in the subsequent levels, this factor could separate explanation in computational theory from explanation in both the algorithmic level and the neurophysiological implementation. Furthermore, his pressure for uniqueness and exactness in vision would prove to be not quantitative, but qualitative. What

Marr seeks is not a detailed formal description of visual processes, but a valid description of those processes, as he himself views it.

A possible objection against this proposal is that it doesn't seem intuitively clear why Marr substitutes computation with logic in his visual theory. The answer is that Marr does not abandon the information processing part of his theory. However, the quest for validity of this part as being true of the world makes him appeal to a general deductive framework. That might be the meaning of the phrase that constraints are turned into assumptions, being "incorporated into the design of a process" (p. 104).

Another objection would state that Marr does not provide formal proofs for every computational description of a visual process. Although this may hold, the recursive character of the information processing approach binds each computational output which subsequently functions as an input to the previously used constraints according to the formulated assumptions. Consequently, whatever the explanatory input of the other two levels may be, the solution is or depends on the result of a formal proof in the computational level out of the appropriate constraints specified.

This formulation of Marr's account requests a rigorous formal method to individuate computational states according to intentional content and the corresponding sufficiency claim of environmental facts. This claim is located inside a distinct explanatory part (the "why" question of his computational level), which along with the information processing part (the "what" question) define visual computational theory, and Marr expects that the algorithmic and the implementational levels will accord with the prescribed account. If the above considerations are correct, this account of content attribution, with its insistence on the notion of logical deduction, would be more forcefully challenged in the light of embodied theories of cognition and active vision (cf. Brooks, 1991; Clark, 1997). Marr's theory shows that individuation of computational states by means of representational content may be more intricate than it has otherwise been estimated.

References

Bach, K. (1998). Content: Wide and narrow. In E. Craig (Ed.), *The Routledge encyclopedia of philosophy, Volume 2* (pp. 643–646). London: Routledge.

Bermudez, J.L. (1995). Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language, 10*, 333–369.

Biederman, I. (1993). Visual object recognition. In A.I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 1–21). Cambridge, Massachusetts: MIT Press.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 10*, 615–678.

Bontly, T. (1998). Individualism and the nature of syntactic states. *British Journal for the Philosophy of Science, 49*, 557–574.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4*, 73–121.

Burge, T. (1986). Individualism and psychology. *The Philosophical Review, 90*, 3–45.
Butler, K. (1996). Content, computation and individuation in vision theory. *Analysis, 56*, 146–154.
Butler, K. (1998). Content, computation and individuation. *Synthese, 114*, 277–292.
Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.
Churchland, P.S. (1986). *Neurophilosophy: Toward a unified science of the mind–brain*. Cambridge, Massachusetts: MIT Press.
Clark, A. (1997). The dynamical challenge. *Cognitive Science, 21*, 461–481.
Crane, T. (1991). All the difference in the world. *Philosophical Quarterly, 41*, 1–25.
Cummins, R. (1989). *Meaning and mental representation*. Cambridge, Massachusetts: MIT Press.
Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, Massachusetts: MIT Press.
Egan, F. (1992). Individualism, computation and perceptual content. *Mind, 101*, 443–459.
Egan, F. (1996). Intentionality and the theory of vision. In K. Akins (Ed.), *Perception* (pp. 232–247). New York: Oxford University Press.
Field, H. (1977). Logic, meaning and conceptual role. *Journal of Philosophy, 69*, 379–408.
Fodor, J. (1987). *Psychosemantics*. Cambridge, Massachusetts: MIT Press.
Francescotti, R.M. (1991). Externalism and Marr's theory of vision. *British Journal for the Philosophy of Science, 42*, 227–238.
Gilman, D. (1996). Optimization and simplicity: Computational vision and biological explanation. *Synthese, 107*, 293–326.
Kitcher, P. (1988). Marr's computational theory of vision. *Philosophy of Science, 55*, 1–24.
Marr, D. (1982). *Vision*. New York: W.H. Freeman.
Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, Massachusetts: MIT Press.
Morton, P. (1993). Supervenience and computational explanation in vision theory. *Philosophy of Science, 60*, 86–99.
Patterson, S. (1996). Success-orientation and individuation in Marr's theory of vision. In K. Akins (Ed.), *Perception* (pp. 248–267). New York: Oxford University Press.
Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam, *Mind, language and reality* (pp. 215–271). Cambridge: Cambridge University Press.
Segal, G. (1989). Seeing what is not there. *The Philosophical Review, 98*, 189–214.
Shagrir, O. (2001). Content, computation and externalism. *Mind, 110*, 369–400.
Shapiro, L. (1993). Content, kinds and individualism in Marr's theory of vision. *The Philosophical Review, 102*, 489–513.
Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge, Massachusetts: MIT Press.
Treisman, A.M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 194–214.
Treisman, A.M., and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Massachusetts: MIT Press.
Zeki, S.M. (1993). *A vision of the brain*. Oxford: Blackwell.