

# CONSCIOUSNESS MAKES A DIFFERENCE: A RELUCTANT DUALIST'S CONFESSION<sup>1</sup>

*Avshalom C. Elitzur*

To Robert Jahn and Brenda Dunne

## INTRODUCTION (WITH GRUMBLE)

If something odd persists, would its mere persistence make it natural? That would be the case for the layperson, but the scientist and philosopher should know better. Commonness should never mislead us to get used to the miraculous.

Such is the phenomenon known as “consciousness,” underlying the age-old “mind-body problem.” But familiarity breeds contempt; the presence of consciousness at every moment in our waking lives often makes us forget how mysterious it is.

For more than two millennia, the study of this problem has made no real progress. Materialism, dualism, and all other isms keep debating without being able to propose any decisive argument, not to mention an experimental test, which could conclude the debate in favor of one theory or the other.

Is this stalemate inevitable? I submit, with all due modesty, that I have a strong argument in favor of one of the rival sides (Elitzur, 1989, 1996). Alas, this side is dualism, a theory that I dislike. Still, no sound counterargument has been raised against it so far. For most of the time, it has met widespread silence. Hello! Anybody there?

This article reiterates my argument. In sections 1-2 I give an exposition of the mind-body problem through the basic notion of “qualia.” In 3 I point out the “Qualia Inaction Postulate” underlying all non-dualistic theories. Against this postulate I propose in section 4 the “Bafflement Argument,” according to which the fact that people express bafflement about their qualia indicates that qualia play a causal role. Sections 5 and 6 elaborate the argument. Sections 7-10 criticize attempts to dismiss the bafflement argument, mainly the “Bafflement=Misperception Equation.” Finally 11 generalizes the argument to a concise “Asymmetry Proof” against any

---

<sup>1</sup> An earlier version of this article has been published in *Irreducibly Conscious: Alternatives to Reductionist Accounts of Mind*, Eds. A. Batthyany, D. Constant, & A.C. Elitzur. Heidelberg: Universitätsverlag Winter, 2007.

materialistic dismissal of bafflement. Section 12 summarizes by paper, pointing out the inescapability of dualism.

## **1.WHAT'S YOUR MIND-BODY PROBLEM ANYWAY? THE PERCEPTS-QUALIA INEQUALITY**

Often, stating a problem well is half way to the solution. Equally often, the mind-body problem is ill-stated. Chalmers (1996), with typical wit, has shown that, when an author claims to have “solved” the mind-body problem, it is likely that they do not understand problem in the first place. He then introduced his by-now classic distinction between the “hard problem” and the “easy problems” of consciousness. The “hard problem” is the one to be discussed in this article, whereas “easy problems” are

How does the brain process environmental stimulation? How does it integrate information? How do we produce reports on internal states? These are important questions, but to answer them is not to solve the hard problem: Why is all this processing accompanied by an experienced inner life? (pp. *xi-xii*).

Let us, therefore, present the problem first, in a way that will also provide a sufficient basis for the arguments to come. I shall invoke a naïve discussant whose questions will help us focus on the crucial issues.

So what's the mind-body problem with you anyway? Why can't you accept the claim that science gives a satisfactory explanation of consciousness?

When dealing with consciousness, science has miserably failed in what has always been its hallmark of success, namely, *reducing qualities to quantities*. “The qualitative” said Lord Rutherford, “is nothing but poor quantitative.” Consider the following examples: *i*) red differs from blue, *ii*) sweet differs from salty, and *iii*) love differs from hate. These differences seem to be qualitative, but the scientific account neatly converts them to numerical values on the same scales. *i*) Both red and blue light are electromagnetic waves, differing only in their wavelengths: 700 nm for red and 400 nm for blue. Consequently, different cones in our retina react differently to these wavelengths due to different amino-acid sequences of their rhodopsin. *ii*) A sugar molecule,  $C_6H_{12}O_6$ , contains carbon, hydrogen and oxygen atoms, while a salt molecule, NaCl, contains sodium and chlorine atoms. All these atoms contain identical electrons on their shells, differing only in their numbers, which they exchange with the molecules in our tongue receptors. *iii*) Both love and hate involve very similar neurons, differing mainly in their location and spatial arrangement (location, specified by geometry, is also a quantitative measure).

In all these examples, qualitative differences between percepts turn out to be merely quantitative.

Thanks! I'll remember that next time I eat ice cream or hate someone. So why aren't you satisfied with the physical explanations to percepts?

While the explanation does a good job with percepts, rendering them physical events, some intriguing phenomena that accompany these percepts are left out. These are the pure qualities, *qualia*.

What's that?

Qualia ("quale" in singular) are those aspects of our experience that cannot be communicated yet we know they are there. Suppose you and I look at a rose. We assure each other that we both see a red rose. Still, you cannot rule out the possibility that I experience it the way you experience *blue*. True, in all languages each of us would refer to all colors with the same terms that the other would (having verified that our color vision and linguistic abilities are normal). But this only means that we both have correctly learned to associate the appropriate words to the wavelength in question. Nothing of all that can tell you anything about my *quale* of the color. The same holds for all percepts. The percept itself can be accurately communicated, but the accompanying quale remains inaccessible. This is the notorious "problem of inverted qualia." So, it's merely a problem of communication.

Much worse, qualia elude not only words, but observation and experimentation as well. Suppose that, with sufficiently advanced technology, you obtain the fullest real-time description of what goes on in my brain— every neuron, axon, dendrite and synapse, every neurotransmitter molecule – when I perceive a red rose. *You know better than I do what goes on in my brain when I perceive red, and still, that does not bring you any closer to my quale!*

Worse still, it is not only that you cannot be sure that my qualia are similar to yours – you cannot even be sure that I have any qualia *at all*. With today's technology, a machine is perfectly conceivable that will name colors, in any language, with much greater accuracy than all humans. Does such a machine have the qualia of "red" or "blue"?

Returning to humans, the problem of inverted qualia leads to the even more grotesque problem of *absent* qualia. Personally I have no doubt that you, apart from appropriately responding to various colors, sounds, tastes and odors, also have the accompanying qualia. And yet, even this very reasonable belief has no rigorous proof.

Isn't there a law that obliges qualia to be present in the brain?

No, just as there is no known law obliging qualia to exist in any other physical process. Moreover, once you assume that the brain operates in compliance with physical law, qualia have no place in this operation. Here is why.

Consider first the movements of billiard balls. Must you invoke any quale in order to explain their movements? Should you hypothesize that the balls “feel repulsion” upon colliding, or “yearn” to come to rest when slowing down? Their behavior is strictly and solely governed by the laws of mechanics. Next consider a plant that has not been watered for a few days, nearly dying. You water it, and soon its leaves stretch again and regain their vitality. Do you need the qualia of “being thirsty” or “slaking thirst” to explain what happened? The laws of osmosis (different concentrations of salt on the two sides of the plant cell’s semi-permeable membrane) perfectly suffice.

You can guess where I am heading. High up the scale of complexity, above balls and plants, are humans. Their behavior, too, is supposed to be governed by neuronal processes that are, in essence, physical. Likewise their percepts occur with strict accordance with physical law. Now suppose you want to explain a certain behavior, say, picking a red rose. If your explanation invokes not only the *percept* of red but the accompanying *quale* as well, this amounts to asserting that the laws of physics do not sufficiently account for a physical process. It is just as if you assumed that some quale has taken part in the movements of a billiard ball: Something other than mechanical forces seem to govern motion.

Why would that be so bad?

Well, if qualia play a causal role in any process, than some of physics’ most revered laws, such as those of energy and momentum conservation, must be violated. Take again “red.” As long as only the *physical* aspect of color perception affects the person’s picking a red rose, then the color’s *quale* plays no causal role and can be ignored by the scientist. But if the qualia too take part, then the continuous, omnipresent causal network dominating the physical universe must be somewhere broken. Somewhere along the neuronal chain controlling a person’s behavior, an event must occur that is not fully determined by the previous neural events. The very principle of causality is thereby violated.

But surely there is a difference between a few balls and an entire human! Our behavior is so complex...

Do not make the common error of relating qualia to complexity. The UN administration is many times more complex than each of its single officers, yet there seems to be no reason to ascribe qualia to the UN as a whole. Complexity seems to be a necessary but not sufficient condition for

qualia. And it surely does not solve the problem. Conservation laws are supposed to equally apply to simple and complex systems alike. If qualia cannot play a role in the former, the same holds for the latter as well.

The situation brings to mind G. B. Shaw's remark: "If you are not a communist at 20, you have no heart; if you are still a communist at 30, you have no head." A similar choice awaits any one who tries to have a consistent view of qualia: Should you be silly or inhumane? On the one hand, if you grant qualia to, say, mice – believing that they have the quale of fear from cats, you may also ascribe the quale of "fear of light" to a photophobic response of a green alga, supposed to be accounted for by the automatic responses of its flagella<sup>2</sup>. Or you may ascribe the quale of "fear of water" to a hydrophobic detergent molecule, supposed to be governed by electrical forces alone. But on the other hand, if you deny the quale of fear to mice, you may as well deny the quale of "fear of tigers" to a terrified Mogley running for his life...

But you keep talking about "physics" as if present-day physics is the final word. Can't you imagine that future physics will reveal new phenomena, say, some unknown properties of matter or energy, which will eventually account for qualia?

Let me show you why you don't have to be an expert in order to realize that *no physical concept, whether known today or still to be discovered, can account for qualia*. Imagine attending a lecture by a world-renowned neurophysiologist announcing that she has discovered a new neurotransmitter, say, alpha-mindo-enkephaline, that accounts for qualia. A very detailed explanation follows, showing the enormously complex pattern of interactions of that molecule with specific receptors within the neuronal synapses. Naturally, if you have not specialized in neurochemistry and neuroanatomy, no chance you will understand what she is talking about. Or imagine that the lecturer is a distinguished physicist who has discovered a new property of elementary particles. You know that physicists ascribe to particles some properties to which they give fancy names like "beauty," "charm," etc. So this physicist has discovered a new property, "loveliness," and this property, inherent also to the particles within our brains, accounts for our qualia. Here too a frightful mathematics follows, explaining what "loveliness" is and how it gives rise to qualia. And here too, unless you are a competent particle physicist, you will understand nothing. Yet in both cases, if you look around in the audience, you will see senior scientists pensively nodding, perhaps making a few technical objections but saying, "Well, it's interesting. Let's think about it."

---

<sup>2</sup> Facilitated – how inventive is evolution! – by rhodopsin, the same pigments facilitating color vision in our eyes.

Don't bother! Just ask, *a-priori*: Can any such property of matter assure me, in principle, that my quale of red is not my fellow human's quale of blue? Worse, can any such explanation prove that my fellow human has qualia at all? Can loveliness or alpha-whatever rule out the possibility that my fellow human lacks qualia altogether? You see, this is not a question of more knowledge.<sup>3</sup> Qualia lie, *in principle*, out of any possible physical account.

So, if qualia lie outside of any physical account, why not ignore them altogether?

Well, let's see how long you can. Think about sleepwalking. Ridgway (1996) discusses in detail cases where people committed murder while allegedly asleep, raising the question whether they can be accounted guilty for their deeds. Murder is an act requiring fairly advanced cognitive faculties, and yet, the people in these cases are believed to have been totally unaware of their own actions. Wilkes (1984) refers to a less substantiated case of a somnambulist physician who performed a medical examination and even made a correct diagnosis – all while asleep. In principle, there is no reason why this state cannot be extended to all mental functions. One might, in other words, laugh, cry and sing while totally unconscious. So here is a thought experiment for you. How about turning all your qualia off, while leaving all your actions just the same? No one would ever notice any difference, only your qualia would be gone, forever. The guy who wants to conduct this experiment on you offers you \$1,000,000 for as a reward. Would you agree?

But it is doubtful whether it is at all possible to accurately...

Never mind technicalities! In theoretical physics, a *gedankenexperiment* (thought-experiment) is an indispensable tool that enables you to anticipate technology by many years. Recall that physics allows the existence of humans with no qualia at all; in fact, such humans are *more* compatible with physics than we, conscious humans. So, would you agree to go into lifelong sleepwalking, turning off your qualia of red and blue, sweet and salty, love and hate, forever, leaving your observed behavior intact, for a nice sum?

Well, others won't see any difference, but for me it would be nothing short of death.

Welcome to the mind-body problem. Qualia accompany every moment of our waking life; in fact they are the most essential ingredients of it, and yet, they have no place even in the fullest and most detailed scientific explanation of our brain's work and our resulting behavior.

The situation can be summarized as follows.

---

<sup>3</sup> I am indebted to Uzi Awret for the following keen observation: "While the easy problem of consciousness is a result of knowing too little, the hard problem results from knowing too much."

A Percept	A Quale
A state occurring in one's brain upon perceiving a certain stimulus	An experience accompanying the percept
Evolving in strict compliance with physical law	Not entailed by physical law
Can be observed, communicated and quantitatively measured with any desired accuracy	Cannot be observed, communicated or measured

**2.THE CHOICE: DISMISSING QUALIA OR ACCEPTING VIOLATIONS OF PHYSICAL LAWS.**

Several theories have been proposed over the generations to tackle this problem. They can be grouped into two major types, namely, materialism and dualism.<sup>4</sup> Materialism invented a variety of exercises in order to prove that a quale does not really exist, being merely another aspect of the percepts. Dualism, on the other hand, straightforwardly acknowledged that qualia exist alongside percepts.

Materialism never won the full acceptance of the scientific and philosophical communities. If there is something to “red” that I cannot communicate and that even the most detailed description of my brain cannot yield – then something probably *exists* that lies outside of the framework of present-day science. As everyone would confirm upon reflection, have we had no qualia, we would be inwardly dead. Yet physics cannot explain why these qualia exist in the first place.

While dualism cannot be accused of such narrow-mindedness, the cure it offers seems to be worse than the disease. It straightforwardly states that qualia are essentially distinct from percepts. But if these qualia play, in addition to our percepts, any role in our behavior, the clash with physical law is inevitable.

Understandably, some people turned to parapsychology in search of a direct proof for an interference of this kind with physical processes. Most notable among time were the Princeton Engineering Anomalies Research Laboratory (Jahn & Dunne, 1987, 2001). Yet, so far, after many years of admirable labor, the effects they found have not been strong enough to convince the scientific community.

Others tried to invoke quantum mechanics to avoid the breach of conservation laws entailed by dualism. QM, so it seems, has undermined determinism, hence an interference of qualia in the brain's random macroscopic events should not violate physical law. Thus Eccles

<sup>4</sup> A more accurate dichotomy will be the contrast between materialism or physicalism (“the material world is everything; mind is an illusion”) and mentalism (“everything is mind; matter is an illusion”). Conversely, one may distinguish between monism (“there is only one reality, namely, the material/mental”) with dualism (“there are two realities, the material and the mental”). But for our purpose the above dichotomy suffices.

(1994) has invoked QM to allow free will to interfere with the brain neurons' synapses without entailing a violation of the first law of thermodynamics (energy conservation). The trouble is that there is also a *second* law of thermodynamics, dealing with entropy increase. One of this law's derivatives, associated with the famous "Maxwell's demon" paradox (see Elitzur, 1994), says that it is impossible to introduce order into a disordered process without investing energy. That leaves the dualists with two options. Either

- a. *Qualia's effect on behavior is entirely random.* That won't do the trick. To believe that qualia affect our behavior means necessarily that they affect it *nonrandomly, in a certain way*. For example, if the quale of red, in addition to the percept, affects one's picking a red rose, then that person's color preference will turn out to be different every time he or she picks a rose.

Or

- b. *Qualia's effect is systematic.* But then, qualia must be using energy in order to interfere with the brain's random processes in a nonrandom manner, again violating either the first and/or the second law of thermodynamics.

### **3.NEARLY A CONSENSUS: THE QUALIA INACTION POSTULATE**

So, is materialism or dualism offering the lesser evil? Before addressing this question, we must consider a few more variants of these two rival schools:

- a. *Identity/double-aspect theory:* Qualia and percepts are one and the same thing, only perceived as different.
- b. *Parallelism:* Qualia and percepts are different, belonging to different realms. In each realm, events follow one another in strict cause-and-effect relations. However, the two realms run parallel to one another, never interfering with one another. Only by virtue of their perfect correlation they give the illusion that they causally affect one another.
- c. *Epiphenomenalism:* Qualia and percepts are different processes belonging to the mental and the physical realms, respectively. But they maintain asymmetric causal relations: percepts causally affect qualia but never *vice versa*.

Now, we do not need to go into the details of these theories, for they all share with materialism one crucial assumption which we shall shortly put to test: *Qualia play no causal role in themselves.*

Let us examine this assumption with the aid of an ordinary behavior: A woman kisses a man. First, consider the man-in-the-street explanation of this behavior:

1. Alice kisses Bob because she loves him (common sense).

Not good enough. “Love” is both a percept and a quale. Which of them, then, constitutes the kiss’s cause? Ask anyone who has ever been in love whether they would be willing to give up love’s quale, even if the percept and the entire resulting behavior would remain intact, such that even the loved one would never be able to notice the change. To the extent that we are dealing with a normal human, whose reply would be a loud “No!”, you might get the following, more daring account:

2. Alice’s kissing Bob is caused by the quale of her love to him (Interactionist dualism).

No scientifically minded scholar would accept such an account, for reasons explained above. She might propose an alternative description:

3. Alice kisses Bob because the *percept* of love, caused by sensory signals coming from him, triggers in turn the behavior of kissing. (Materialism)

And if you ask where, in all this, is the love’s *quale*, a prudent theorist might clarify her position as follows:

4a. Alice’s quale of loving *is* the percept of loving; she only perceives the two things as distinct. (Identity or Double-aspect theory)

Or

4b. Alice has the quale of loving *alongside* with the percept. However, it is only the latter, never the former, that cause the kissing. (Epiphenomenalism or Parallelism)

The bottom line of all these theories, shared with materialism as well, is this: Any behavior, manifested by any person, would be exactly the same had there been no qualia in the first place (Kim, 1996). This assumption can be succinctly put as:

***The Qualia Inaction Postulate:*** *For any instance where a quale seems to have affected behavior, it can be shown that it was not the quale but its physical parallel, the percept, which has exerted the effect.*

Little wonder that this postulate, in one form or another, has been opted by most modern philosophers.<sup>5</sup> It allows that something very peculiar is going on within us, yet assures us that this thing has no influence whatsoever on our observed behavior. Nothing, therefore, needs to be changed in the scientific worldview. Qualia, whatever they are, would better be left to philosophers, who do not mind wasting their time and energy on things that make no difference.

---

<sup>5</sup> Kim (1996) refer to this postulate as the “exclusion argument,” and Flanagan (1997) as “conscious inessentialism.”

#### 4. THE POINT OF DEPARTURE: THE BAFFLEMENT ARGUMENT

The stakes are now very high. We need only one example in which the Qualia Inaction Postulate fails – where a quale alone, in itself, not its physical counterpart, exerts a causal effect – to falsify all the comfortable alternatives to interactionist dualism. Guess what: This example is occurring to you, dear reader, at this very moment!

For, why do we talk, write, argue and think about the mind-body problem? Why do we feel and say that qualia are not identical with brain processes? I submit that the answer is simply this: We are baffled by the peculiar nature of qualia because we *have* qualia. Hence, as against the Qualia Inaction Postulate, I propose

*The Bafflement Argument: The fact that humans are baffled by the discrepancy between qualia and percepts, and express this bafflement by their observable behavior, is a case where qualia per se – as distinct from percepts – play a causal role in a physical process.*

#### 5. OF KISSES, HEARTBREAKS AND REFLECTIONS

“He who increases knowledge increases pain,” says Ecclesiastes (1, 18), but the opposite is equally correct. Isn’t it significant, now that we come to think of it, that most discussions of qualia take *painful* experiences as a starting point? Happy experiences are taken for granted! So, for knowledge’s sake, let us inflict the following pain on our Alice. She and Bob have broken up, leaving Alice sad over the separation. Now Alice asks herself why there is a quale of sadness.

Study she might as much neuroscience as she can, she will soon encounter the mind-body problem. She may eventually find consolation in a new relationship, but the problem will keep baffling her, hence she will talk, argue and attend conferences on it. She may even present a paper.

Can we subject this bafflement of Alice to the same procedures we have earlier applied to her previous act of kissing? The man-in-the-street account is, again, very simple:

1. Alice says that qualia are baffling because she experiences a quale that is distinct from her percept (Interactionist dualism).

This, of course, is anathema to anyone who believes that the physical world is closed. There seems to be only one way one can explain Alice’s bafflement about qualia without quarreling with conservation laws, namely, to prove her belief in non-physical qualia *wrong*. Consider the less interesting case in which Alice believes that there is evil eye. No one would ascribe this belief to the *existence* of evil eye. Rather, we would prefer to ascribe it to an error in

Alice's belief-system, stemming from inappropriate education, faulty reasoning and some misleading experiences. Why not, then, ascribe the bafflement about qualia to a similar misconception?

Notice, however, that by this explanation the materialist position commits itself, for the first time, to a *falsifiable* hypothesis. By materialism, Alice's presumably false belief that qualia are baffling must be explained in the following way:

2. Alice says that qualia are baffling because there are some processes in her brain, resulting from inadequate education, faulty reasoning, misleading experiences, etc., that make her believe that qualia are distinct from percepts (Materialism).

This, let me stress again, gives a falsifiable prediction. Once future neurophysiology is capable of pointing out the neural correlates of false beliefs such as evil eye, ghosts, etc., it will be equally capable of pointing out the reason why, for more than two millennia, numerous otherwise-intelligent philosophers and scientists keep insisting that there was something inexplicable to qualia.<sup>6</sup>

I don't believe any of this. But the point is that the question is no more a matter of belief alone. The Qualia Inaction Postulate is now committed to a straightforward prediction: *When future neurophysiology becomes advanced enough to point out the neural correlates of false beliefs and superstitions, a specific correlate of this kind would be found to correspond to the bafflement about qualia.* Conversely, the Bafflement Argument makes the opposite prediction: *No neural pattern typical of false belief concerning qualia will be found in a dualist's brain.*

#### **6.A WELCOME CONSEQUENCE: QUALIA MAKE EVOLUTIONARY SENSE**

We have, it seems, an unwelcome argument in favor of interactive dualism, the most unappealing option for anyone brought up in the scientific tradition. Let us appreciate, however, one of its immediate benefits, which for upholders of the Qualia Inaction Postulate is unattainable. If Alice kisses Bob only by virtue of neural mechanisms developed during evolution, whence the quale of loving? If a rabbit escapes a fox only by virtue of neural mechanisms, and if the fox chases it unaffected by the quale of hunger, why are there qualia of fear and hunger in the first place?

Once bafflement indicates the ability of qualia to affect behavior, then, as disturbing the clash with physical law is, the evolutionary question gets a very reasonable answer. Alice's kiss

---

<sup>6</sup> A possible objection to this prediction can argue that a false belief can exist in one's brain while its causes, namely, the previous events that have led to this belief, no longer exist in that person's memory. However, the experience of both psychodynamic and cognitive therapy shows that this is not so. The causation of false belief seems to be accessible to the therapist, albeit with considerable effort, their elucidation leading to the removal of that false belief. Neurophysiology will therefore follow suit.

may take a bit *longer* thanks to the additional effect of the quale of love, and the qualia of hunger and fear may add some *velocity* to the rabbit's and the fox's race. Qualia, in other words, give a clear advantage for survival.

But we still have a long way ahead of us before savoring this insight. Rather than enamored girls, scared rabbits and hungry foxes, we should analyze the output of prudent philosophers.

### **7.THE CRUCIAL QUESTION: IS BAFFLEMENT MERELY DUE TO MISPERCEPTION?**

In the following sections a crucial question will be discussed: *Is the expression of bafflement over qualia obliged by the physical laws governing our brains?* Or put in the language of AI, *Is there a physical/logical principle that obliges an intelligent computer to express bafflement of this kind?* I italicized these questions because they are extremely important, and forgive my didactic tone when I italicize also my answer to them:

*If a proof is ever given that an intelligent system, by virtue of physical laws alone, must state that it has qualia which are distinct from its corresponding percept, then the age-old mind-body problem would finally get a definite solution – a materialist one. The difference between qualia and neural processes would turn out to be nothing but a misperception inherent to all intelligent systems, and the problem would turn out to be a pseudo-problem.* Just as there is no “rabbit-duck problem,” “left-right positioned Necker cube problem” or any other problem entailed by misperception, so would the mind-body problem finally turn out to be merely an unfortunate failure of many wise people to realize that percepts and qualia are just one and the same thing. All dualistic arguments made over the millennia – from Plato to Descartes to Leibniz to Popper and all others – would be dismissed by a simple *ad hominem*, formulated in the precise terms of cognitive science. Philosophy and science will finally move on – for good – to other issues.

Now, the authors quoted in the next sections are proposing just such a proof, evidently without realizing that, if correct, this proof amounts to no less than a final solution to the mind-body problem. The only question is, Is this proof valid?<sup>7</sup>

### **8.CHALMERS' BAFFLEMENT=MISPERCEPTION EQUATION**

In his delightful *The Conscious Mind* (1996) Chalmers' briefly objects to my position:

---

<sup>7</sup> Don't hold your breath.

Indeed, Elitzur (1989) argues directly from the existence of claims about consciousness to the conclusion that the laws of physics cannot be complete, and that consciousness plays an active role in directing physical processes (he suggests that the second law of thermodynamics might be false). But I have already argued that interactionist dualism is of little help in avoiding the problem of explanatory irrelevance (p. 183).

Indeed, Chalmers has struggled against a similar idea in his discussion of “zombies.” Suppose that there are intelligent beings that resemble us in every observable respect but have no qualia.<sup>8</sup> This, as noted above, is *perfectly consistent with physics* – in fact, it accords with physics *more* than the existence of non-zombies like us.

Naturally, when discussing zombies in this context, a crucial question emerges: *Would zombie philosophers be as baffled by qualia as human philosophers are?*

Astonishingly, Chalmers’ answer is in the affirmative. His reasoning is so peculiar that I prefer to use lengthy quotes:

To see the problem in a particularly vivid way, think of my zombie twin in the universe next door. He talks about conscious experience all the time – in fact, he seems obsessed by it. He spends ridiculous amounts of time hunched over a computer, writing chapter after chapter on the mysteries of sensory qualia, professing a particular love of deep greens and purples. He frequently gets into arguments with zombie materialists, arguing that their position cannot do justice to the realities of conscious experience.

And yet he has no conscious experience at all! In his universe, the materialists are right and he is wrong. Most of his claims about conscious experience are utterly false. But there is certainly a physical or functional explanation of why he makes the claims he makes. After all, his universe is fully law-governed, and no events therein are miraculous, so there must be *some* explanation of his claims. But such explanations must ultimately be in terms of physical processes and laws, for these are the *only* processes and laws in his universe (p. 180).

Not even considering the possibility that a zombie will simply *not* be baffled by qualia, Chalmers inevitably reaches the following absurdity:

---

<sup>8</sup> Chalmers distinguishes between a psychological zombie and a phenomenal one. The zombie known from voodoo horror stories (or from the common derogatory term) is a psychological zombie, manifesting a clear behavior of a “living dead” such as apathy and lack of emotions. The zombie with which we deal, in contrast, is a phenomenal zombie, capable of manifesting all emotions manifested by humans, while only lacking the respective qualia.

The explanation of *his* claims obviously does not depend on the existence of consciousness, as there is no consciousness in his world. It follows that the explanation of my claims is also independent of the existence of consciousness (p. 180).

One has to read this passage time and again in order to believe what it says: *A philosopher writes a book about qualia, discussing its enigmatic nature in great detail, and yet, out of adherence to the Qualia Inaction Postulate, states that his own qualia played no role in his writing the book – that he would write just the same book had he lacked qualia!*<sup>9</sup>

But why on Earth should zombies express bafflement about qualia if they don't have any? After all, Chalmers professes physicalism, by which there is a cause for anything zombies say. If that cause is not qualia themselves, what is it? As he himself asks,

To get some feel for the situation, imagine that we have created computational intelligence in the form of an autonomous agent that perceives its environment and has the capacity to reflect rationally on what it perceives. What would such a system be like? Would it have any concept of consciousness, or any related notions?

His answer is “yes,” reasoning, in essence, that the zombie *misperceives* his “direct, unmediated” percept as distinct from what he knows about that percept:

[I]t seems likely that such a system would have the same kind of attitude toward its perceptual contents as we do toward ours, with its knowledge of them being direct and unmediated, at least as far as the system is concerned. When we ask how it knows that it sees the red tricycle, an efficiently designed system would say, “I just *see* it!” When we ask how it knows that the tricycle is red, it would say the same sort of thing that we would do: “It just looks red.” If such a system were reflective, it might start wondering about how is it that things look red, and about why it is that red *just is* a particular way, and blue another. From the system's point of view it is just a brute fact that red looks one way, and blue another. Of course from our vantage point we know that this is just because red throws the system into one state, and blue throws it into another; but from the machine's point of view this does not help (p. 185).

Chalmers' reasoning can be summarized as follows: *i*) People have qualia. *ii*) People express bafflement about qualia. *iii*) Physics allows the existence of zombies that have no qualia;

<sup>9</sup> And to make the irony perfect, it is Chalmers who has added a naughty comment in his book (p. 190) wondering whether the staunch materialist Dennett is a zombie!

yet *iv*) Such zombies should also express bafflement about qualia. Therefore *v*) People express bafflement about qualia for reasons other than their having qualia.

Chalmers is well aware (personal communication, July 2004) that this position is awkward. But the situation is worse. It is obvious that Chalmers' zombie, by Chalmers' own typology (see section 1 above), is talking about one the "easy problems" whereas the real Chalmers addresses the "hard problem." This distinction will shortly enable us to prove the bafflement=misperception equation plainly wrong.

### **9. CHALMERS VS. CHARMLESS: A REVISED TURING TEST**

On the basis of the above debate we can propose an experiment that, though not yet feasible, may one day move the study of qualia from philosophy to empirical science. In essence, it consists of a simple modification of the famous Turing test (Turing 1950), designed for judging whether a sufficiently advanced computer can simulate human intelligence. With the proposed revision, this test can give a clear-cut answer to a much more exciting question.

Turing's test was this: Let a computer and a human dwell in two separate rooms. Let the experimenter, unaware in which room the computer is and in which the human, send whatever questions she has in mind to both rooms via electric cables and get their answers in the same way. If she fails to tell by the answers who is the machine and who is the human, then the machine, for all practical purposes, possesses intelligence like that of the human.

Suppose, now, that such a future computer has passed the test. The time will then be ripe for the greatest question of all: *Does this computer have qualia? Or does it constitute the philosophers' proverbial zombie?*

Before proceeding, let us give our computer a name. How about *Charmless*? The slight difference from its human namesake indicates that, unless we prove that it has qualia, then, even if it is as bright and witty as Chalmers, it is Chalmers' zombie incarnate.

The test is straightforward – a seemingly innocuous question, say, "What is red?"

Assuming that we have followed Turing's recommendation "to provide the machine with the best sense organs that money could buy" (Hodges, 1988) Charmless might give a very accurate answer, such as

1. "Red is the color I perceive whenever electromagnetic radiation of the wavelength 700 nm impinges on the photoelectric device at the back of my obscure chamber, absorbed by photochemical molecules sensitive to this wavelength, and converted into electric pulses that go through optic fibers to the

color-recognition system that in turn activates my memory, language and vocal systems.”

This answer is much more detailed and precise than that of an average human who is oblivious of her own neurophysiology. But it would indicate that this computer is a true zombie, indeed a charmless one.

On the other hand, the above answer might have an addendum, something like

2. “Red is... [see 1 above]. *However*, there is something to my immediate experience of red that is not indicated by the description I just gave you. I know of no way of communicating that additional ingredient. Although I can see that you and I refer to the same color when we use the same word, I can never be sure whether your subjective experience of red is not what I experience as blue. In fact, I am not sure you have *any* subjective experience accompanying your color perception.”

This, as with humans, would indicate that Charmless too has something to his experience that go beyond the physical percepts.

Notice that in this case we can determine with certainty the cause of this bafflement. Since Charmless is man-made, we can rule out the possibility that the bafflement is the result of some pre-installed “bug” such as an explicit command to express bafflement or some deliberate misperception imposed on it. In other words, we can rule out any cause to Charmless’ assertion of having qualia other than his really having them.

### **10. BUT PERHAPS ZOMBIES ARE BAFFLED TOO?**

Following Chalmers’ analysis, a heated debate ensued over the issue of zombies, particularly over the question whether they might be baffled over qualia. Moody (1994) argued that zombies would not be baffled by qualia, a prediction that accords with the Bafflement Argument. Several articles followed Moody’s, most of them expectedly objecting to his conclusion. Flanagan and Polger (1995) argued that zombies might wonder, just as conscious humans, whether qualia are inverted:

Suppose that normal zombies, upon seeing light of a certain wavelength  $x$  go into a state that is the disposition to say, “that object is green”, and then they act on that disposition. [...] All that is necessary for an inverted color judgment problem is that behavioral pathways get crossed twice. In our case (i.e., the usual inverted spectrum problem) one of the pathways is supposed to be the qualitative look of

color, and the other a speech act. Zombies could have an equivalent problem with two non-conscious inversions [...]. First, when seeing an object that reflects a wavelength x, the inverted color judgment zombie enters the state that, in normal zombies, is the disposition to say, "that object is red." However, due to the second crossed wire, the inverted color judgment zombie's "that object is red" state actually causes it to utter, "that object is green." Thus a double inversion can create a problem indistinguishable from the inverted spectrum problem.

This primitive version of the bafflement=misperception equation can be straightforwardly ruled out. Just suppose that these zombies are well informed about their brain physiology! All they need to do is to inspect one another's brain, which is just what for us humans would be useless even with the best possible brain inspection methods.

Most vehement in his criticism on the zombie literature is Dennett (1995):

If, *ex hypothesi*, zombies are behaviorally indistinguishable from us normal folk, then they are really behaviorally indistinguishable! They say just what we say, they understand what they say (or, not to beg any questions, they understand<sup>z</sup> what they say), they believe<sup>z</sup> what we believe, right down to having beliefs<sup>z</sup> that perfectly mirror all our beliefs about inverted spectra, "qualia," and every other possible topic of human reflection and conversation (p. 322, italics original).<sup>10</sup>

While Dennett has detected a linguistic flaw in Moody's formulation of the problem, he misses the point. Moody's question can be easily rephrased such that it will be immune to Dennett's criticism: "Suppose that there are zombies that behave just as we do yet lack qualia. Would their bafflement about qualia be consistent with physical laws?"

Thus phrased, the question has two answers, each of which having far-reaching consequences:

- a. Zombies will be baffled over qualia by virtue of some physical cause. In this case, *our bafflement is not due to the existence of qualia*. But then, another cause for this bafflement must be detected by future neurophysiology, which will indicate that *we somehow misperceive our percepts*.
- b. Zombies will not be baffled over qualia. *Dualism would then be correct*.

---

<sup>10</sup> The superscript z denotes, following Chalmers, zombie mental functions that resemble ours but lack qualia.

## 11. THE ASYMMETRY PROOF: BAFFLEMENT HAS NO PHYSICAL ORIGIN, HENCE...

We are now in a position to show that all the materialistic counter-arguments to the Bafflement Argument run into a contradiction. In general, these authors make two mutually entailing claims:

Bafflement over qualia is due to misperception  $\Leftrightarrow$  Qualia play no physical role

Taking Chalmers' analysis as an example, let us see how such a view leads to an outright contradiction:

1. A presumably conscious human (henceforth Chalmers) states there is a difference between his percept (P) and its corresponding quale (Q).
2. Chalmers further argues that a zombie duplicate of him (henceforth Charmless) is possible, which is similar to him in all aspects, save that he has only P without Q.
3. Chalmers asserts, however, that, by physical law, Charmless must notice a difference between what he knows about the physical process underlying his percept and the unmediated percept itself, which for Charmless plays the role of Q.
4. Chalmers then argues that this difference must produce in Charmless the same bafflement as Chalmers' bafflement about the difference between P and Q.
5. Ask now Chalmers: Can you conceive of a Charmless who will be identical to you but lack Q? His answer, by (2), is "Yes."
6. Then ask Charmless: Can you conceive of a duplicate of you (henceforth Harmless) who will be identical to you but will lack Q? His answer, by (3), must be "No; unmediated percepts, regardless of what is known about them, must occur."
7. As Chalmers can conceive of Charmless but Charmless cannot conceive of Harmless<sup>11</sup>, the two kinds of bafflement, associated with (1) and (3), are essentially different.
8. Hence, the physical explanation for (3) does not hold for (1).

The argument can further be generalized so as to hold for any theory that denies causal role to qualia. Why do we perceive qualia as distinct from our neural firings? If it is not because they *are* different, then we have a *misperception* of them as different. But misperception, just like perception, may or may not have an accompanying quale. Take a simple optical illusion, say, one in which a straight line appears to be curved. *Both the correct and the incorrect percepts may or*

---

<sup>11</sup> Which is why we don't need to worry about Armless and so on.

may not have a quale. Now, the optical illusion is obliged by the laws of perception (which are in essence physical laws), whereas qualia are not! Hence,

*The Asymmetry Proof: If a quale is not distinct from its percept, then its appearance as distinct must be a misperception of a percept. But misperception, being a special kind of perception, is also conceivable without qualia. Hence, beings that only misperceive their percepts will not be baffled as those that misperceive their perception and also possess qualia.*

## **12.SUMMARY: TIME TO FACE THE INESCAPABLE**

At the end of the day, it is astonishing that, in the enormous literature on the mind-body problem, nearly no attention has been given to the very simple question: Why are people baffled over the mind-body problem?

I submit that there are only two consistent answers. Either

1. People are baffled by the mind-body problem because the problem is genuine, i.e., qualia are not identical with the percepts with which they come. Ergo, it is their very existence that gives rise to the physical behavior associated with bafflement. Ergo, something non-physical interferes with physical processes;

or

2. People are baffled by the mind-body problem for erroneous reasons. Qualia are *identical* with the percepts. Ergo, the causes for the misperception of qualia must eventually be found in the some observable failure in brain's rational operation.

The achievement of the Bafflement Argument is that it has made (2), by the premises of materialism itself, a falsifiable prediction. The burden of proof now lies on materialism.

And there is a hybrid of (1) and (2), endorsed by Chalmers (1996) and Flanagan and Polger (1995):

3. People are baffled by qualia, and the problem is genuine, i.e., qualia are not identical with the percepts. However, people are baffled by qualia for reasons other than having qualia.

Allow me to formulate (3) in common language:

I have qualia, but I would have said that I had qualia even if I had not. Still, I have qualia. Believe me, I do!

I must confess that have I heard such a statement from a person in the street I would suspect that he is schizophrenic or at least severely schizoid. That such a position is endorsed by

competent philosophers only attests to the acuity of the cardinal question as to why people are baffled by qualia. Dualism, alas, offers the most reasonable answer.

### Acknowledgments

It is a pleasure to thank the PEARlab members for their stimulating Conscious Academy on July 2003 at Princeton University, where the first draft of this article has been written, and to Harald Atmanspacher for an illuminating dialogue. The imaginary discussant invoked in the article has been, in fact, Moran Cerf. Special thanks are due to Andreas Lindblom, Nili Alon and Vijay Chandrasekaran for their helpful comments. When this article was under its last revision I found out that Uziel Awret has reached similar conclusion to mine in an unpublished manuscript. I thank him for enlightening discussions.

### References

- Block, N., Flanagan, O., & Güzeldere [Eds.] (1997) *The Nature of Consciousness: Philosophical Debates*. Cambridge, Mass.: MIT Press.
- Chalmers, D (1996) *The Conscious Mind*. Oxford: Oxford University Press.
- Dennett, D. C. (1995) The unimagined preposterousness of zombies: Commentary on T. Moody, O. Flanagan and T. Polger/ *Journal of Consciousness Studies*, **2**, 322–6.
- Eccles, J. C. (1994) *How the Self Controls Its Brain*. New York: Springer.
- Elitzur, A. C. (1989) Consciousness and the incompleteness of the physical explanation of behavior. *The Journal of Mind and Behavior*, **10**, 1-19.
- Elitzur, A. C. (1991) Neither idealism nor materialism: A reply to Snyder. *The Journal of Mind and Behavior*, **12**, 303-307
- Elitzur, A. C. (1994) Let there be life: Thermodynamic reflections on biogenesis and evolution. *Journal of Theoretical Biology*, **168**, 429–459.
- Elitzur, A. C. (1996) Consciousness can no more be ignored: Reflections on Moody's Dialogue with Zombies. *Journal of Consciousness Studies*, **2**, 353-358.
- Flanagan, O. (1997) Conscious inessentialism and the epiphenomenalist suspicion. In Block, N., et al. [Eds.] *The Nature of Consciousness: Philosophical Debates*, pp. 357-373.
- Flanagan, O., and Polger, T. W. (1995) Zombies and the Function of Consciousness. *Journal of Consciousness Studies*, **2**, 313-321.
- Hodges, A. (1988) Alan Turing and the Turing Machine. In Herken, R. [Ed.] *The Universal Turing Machine, a Half-Century Survey*, Oxford: Oxford University Press.
- Jahn, R. G., and Dunne, B. J., (1987) *Margins of Reality: The Role of Consciousness in the Physical World*. New York: Harcourt Brace Jovanovich.
- Jahn, R. G., and Dunne, B. J. (2001) A modular model of mind/matter manifestations (M5). *Journal of Scientific Exploration*, **15**, 299-329.

- Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind-Body problem and Mental Causation*. Cambridge (Mass.): MIT Press,.
- Lowe, E.J. (2003) Physical causal closure and the invisibility of mental causation. In Walter, S., & Heckmann, H.-D. [Eds.] *Physicalism and Mental Causation*, pp. 137-154.
- Moody, T. (1994) Conversations with zombies. *Journal of Consciousness Studies*, **1**, 196–200.
- Ridgway, P. (1996) Sleepwalking: Insanity or Automatism?, *E-Law: Murdoch University Electronic Journal of Law*, **3**, No.1.
- Walter, S., & Heckmann, H.-D. (2003) [Eds.] *Physicalism and Mental Causation*. London: Imprint.
- Wilkes, K. (1984) *Is consciousness important?* *British Journal for Philosophy of Science*, **35**, 223-243