

Experimental Methods for Unraveling the Mind–Body Problem: The Phenomenal Judgment Approach

Victor Yu. Argonov

Pacific Oceanological Institute of the Russian Academy of Sciences

A rigorous approach to the study of the mind–body problem is suggested. Since humans are able to talk about consciousness (produce phenomenal judgments), it is argued that the study of neural mechanisms of phenomenal judgments can solve the hard problem of consciousness. Particular methods are suggested for: (1) verification and falsification of materialism; (2) verification and falsification of interactionism; (3) falsification of epiphenomenalism and parallelism (verification is problematic); (4) verification of particular materialistic theories of consciousness; (5) a non-Turing test for machine consciousness. A complex research program is constructed that includes studies of intelligent machines, numerical models of human and artificial creatures, language, neural correlates of consciousness, and quantum mechanisms in brain.

Keywords: mind–body relationship, neural correlates of consciousness, tests for consciousness

In the twentieth century, scientific progress suggested new hypotheses and approaches to the study of consciousness. These hypotheses softened some old problems and created some new ones. However, the most important issue had not changed significantly since Cartesian times. For the mind–body problem, philosophy still had two basic alternatives: (1) consciousness can be studied and controlled as an objective part of matter, or (2) consciousness contains immaterial “degrees of freedom,” which cannot be controlled or observed by objective methods. For the sake of simplicity, I shall call these alternatives “materialism” (including materialistic monism, neutral monism, panpsychism, pantheism, “anomalous” monism, etc.) and “(substance) dualism” (including all theories that regard the human being as a combination of physical body and immaterial soul).

Today, there are four basic views on perspectives of scientific choice between materialism and substance dualism. The first view is that materialism is true, and there are sufficient logical and scientific arguments for it. To support this view, materialist philosophers either consider consciousness as a scientific problem with a materialistic solution (Davidson, 1970; Putnam, 1967) or as a pseudo-problem, and argue that consciousness is a non-scientific, folk term (Churchland and Churchland, 1981). The second view is that science had failed to explain consciousness, and dualism is true (Eccles, 1994; Stapp, 1993). Dualist authors believe that science is obviously incomplete, and modern results (e.g., in quantum mechanics) only support this idea. The third view is radically skeptical: materialism and dualism are the unverifiable doctrines, and mind is incapable of comprehending itself entirely (McGinn, 1989). The fourth view is softly skeptical: in principle, materialism and dualism are verifiable hypotheses, but such verification will be technically possible only in the future (Place, 1956, 1960).

In this paper, I provide arguments for the fourth viewpoint. I argue that today's science has no sufficient arguments for materialism or dualism, but the scientific key to the mind-body problem is our ability to talk about consciousness (produce phenomenal judgments). The study of neural mechanisms of phenomenal judgment production can solve at least some aspects of the mind-body problem and make a scientific choice between various forms of materialism and dualism. Several authors noted the crucial role of phenomenal judgments in the study consciousness (Chalmers, 1997; Elitzur, 1989; Rudd, 2000; Valdman, 1997) and argued that the phenomenal judgment argument can be directly employed for the solution of hard problems (for example, for refutation of epiphenomenalism, see Rudd, 2000; Valdman, 1997). However, their ideas did not become widely accepted. I make an attempt to construct a general phenomenal judgment approach to the mind-body problem and introduce scientific methods for experimental verification of basic theories of consciousness (both materialistic and dualistic). In contrast to Rudd (2000) and Valdman (1997), I do not attempt to disprove any of these theories, but I provide a scientific tool for their study. I also suggest a non-Turing test for machine consciousness (based on phenomenal judgments).

I use a special set of definitions (see next subsection) optimized for my approach. In particular, I view materialism as a theory in which there is a one-to-one correspondence between all subjective facts and some objective facts (constituting a neural correlate of consciousness). I do not discuss whether this means complete metaphysical reducibility of consciousness to matter or only property dualism. The only question I discuss concerns the correspondence between matter and consciousness, not their identity. Some readers might say that this decreases the philosophical value of my study, but I think that the question of correspondence is more scientific than the question of identity. If the state of consciousness is comprehensively determined by the state of matter, then, in practice, consciousness can be studied and controlled as a material object.

Theoretical Basis of the Phenomenal Judgment Approach

Definitions

In this subsection I introduce and discuss the terminology used in this paper. The term “consciousness” (“phenomenal consciousness,” “mind”) will be used in a standard philosophical sense of subjective reality (a totality of person’s subjective [mental] phenomena: sensations, thoughts, volitional acts, etc). The term “matter” will be used in a sense of objective reality (physical particles, energy, space, their properties, and physical laws). In the framework of this paper, information, stored and processed in physical machines (such as deterministic computers) will be also considered as a part of matter.

“Phenomenal judgments” are the words, discussions, and texts about consciousness, subjective phenomena, and the mind–body problem. In this paper, the term “phenomenal judgments” will be used as a synonym of “speech about consciousness” (or other objective phenomenon containing verbal information about consciousness), not “thoughts about consciousness” (subjective phenomena).

The “neural correlate” of a subjective phenomenon (or of a property of consciousness) is a physiological phenomenon containing comprehensive objective (detectable and/or measurable) information about this phenomenon (or a property of consciousness). Here, the words “comprehensive information” mean that there is one-to-one correspondence between any parameter of a given subjective phenomena (for example, a visual image) and some objective parameter in a brain. Any difference between two subjective phenomena must be manifested in their neural correlates. Different subjective phenomena must have objectively distinguishable neural correlates. “Neural correlate of consciousness” is a physiological system containing comprehensive objective (detectable and/or measurable) information on every subjective phenomenon and every property of consciousness of a creature. Note that such correlates must also contain the information that the creature is conscious. If a neural correlate of consciousness exists, then the body of any conscious creature must differ from the body of any unconscious creature. Otherwise, the subjective fact “I am conscious and this body is mine” would not have neural correlates, so the information in a neural correlate of consciousness is not comprehensive.

“Problematic properties of consciousness” are the properties of consciousness having no current satisfactory scientific explanation (no discovered neural correlates). Problematic properties of consciousness are related to so-called hard problems of consciousness (Chalmers, 1997) and are often hypothesized to be immaterial. Examples of such problematic properties: qualia, unity of consciousness, possibility or impossibility of reincarnation, existence of consciousness in a particular creature. Note that some authors deny the existence of “hard” problems and related properties of consciousness (eliminative materialists such as the Churchlands).

Therefore, strictly speaking, all problematic properties of consciousness (discussed in a literature) are hypothetical.

I define “materialism” (physicalism) as a doctrine stating that a neural correlate of consciousness exists, and “dualism” as a doctrine stating that it does not exist. This choice of terms might look strange for some readers. Some materialists do not state that a comprehensive neural correlate of the whole consciousness definitely exists; they state only that matter produces consciousness. However, if some subjective parameters have no neural correlates, then they are hidden from third-person study, and the state of consciousness is not determined unambiguously by the state of matter. Therefore, I define materialism as a doctrine that each subjective parameter has a neural correlate.

For rigorous distinction among the forms of dualism, I introduce the term “immaterial influence on matter”: the change in physical processes meeting two conditions: (1) the change cannot be comprehensively explained (predicted) by physical laws (breaks the causal closure of matter) and (2) it has causal relationship with subjective phenomena that have no neural correlates. I define “interactionism” as a form of dualism stating that at least some forms of conscious behavior are caused by immaterial influence on a creature’s body (Descartes, 1641; Eccles, 1994). In a simplified sense, interactionism is the doctrine that the immaterial soul controls the material body. In contrast, “epiphenomenalism” (in this paper, not distinguished from parallelism) is a form of dualism stating that all basic forms of behavior are possible without immaterial influence on a creature’s body (Hodgson, 1870; Huxley, 1874; Leibniz, 1720). In simplified sense, epiphenomenalism is the doctrine that the immaterial soul exists but does not control the material body. Most authors do not discuss the possibility that an immaterial influence on the human body exists, but that it is not necessary for any important form of behavior. In this paper, I consider this as a special form of epiphenomenalism, because such influence does not play the functional role supposed by classical interactionist authors such as Descartes. However, this is merely a terminological choice.

There are three forms of materialism. “Information-based materialism” (computationalism) is a form of materialism stating that consciousness is a purely informational (functional/computational) object or process (Dennett, 1990; Fodor, 1975; Putnam, 1967). “Substrate-based materialism” is a form of materialism stating that consciousness is a property of a special physical process (chemical, electric, quantum etc.) or of physical substance itself (Anokhin, 1974; Davidson, 1970; Hameroff, 2006; Ivanov, 1998). “Eliminative materialism” is a form of materialism stating that consciousness (or, at least, its problematic properties) is a pseudo-problem, while all cognitive processes have a physical nature (Churchland and Churchland, 1981).

Crucial Examples of the Problematic Properties of Consciousness

Qualia are prominent examples of the problematic property of consciousness. The existence of qualia is often used as an argument against materialism: we are unable to describe qualitative properties of our perception and imagination verbally. For example, we can't explain the essence of red to a human with color blindness (see anti-materialist "knowledge argument" [Jackson, 1982]). And even in the case of normal color vision, we do not know exactly how another person perceives the red color. Maybe the person just calls it "red" but subjectively perceives it as we perceive the blue color (see "inverted spectrum argument" [Shoemaker, 1982]). Discussions on Jackson's, Shoemaker's and other arguments have shown that the existence of qualia does not refute materialism; it rather refutes materialism's most primitive reductive forms. However, it is still unknown whether brain contains comprehensive correlates of qualia or not. The physical world seems to be purely quantitative (space, time, mass, energy, other measurable values). If qualia have neural correlates, then these correlates are supposed to be very special physical phenomena. Materialistic hypotheses on the nature of qualia have been suggested, for example, in Hayek (1952), where qualia are related to functional properties of neuronal analyzers; in Anokhin (1974) and Chuprikova (1985), where qualia are related to neuronal chemistry; and in Hameroff (2006) and Ivanov (1998), where qualia are related to quantum states. However, none of these views is universally accepted today. It should be noted that if neural correlates of qualia exist, then qualia can be comprehensively measured and modified by objective methods. It may be even possible (in principle) to develop equipment, which one person could connect to another person (or to Nagel's [1974] bat) and feel all the other's sensations including unarticulated qualitative content.

The second example of the problematic property of consciousness is the unity of consciousness (binding). Consciousness of a human contains subjective phenomena produced by several sensory organs in a single observer, "self." These phenomena constitute a unified conscious experience of a single person. Self seems to be a fundamentally indivisible thing. Descartes supposed that the unity of self cannot be explained in physical terms, and today this unity still remains unexplained. Brain and its information processes do not demonstrate any fundamental unity that might be interpreted as a neural correlate of the unity of consciousness (Bayne and Chalmers, 2003). Collective quantum effects (Hameroff, 2006), membrane potential oscillation synchrony (Crick, 1995), single-cell (Edwards, 2005; Sevush, 2006) and even single-electron (Argonov, 2012) consciousness concepts have been suggested to explain the unity of consciousness. However, all of these hypotheses remain controversial.

The third example of the problematic property of consciousness is the subjectivity of a particular creature (zombie problem). The existence of consciousness in a particular creature (its subjectivity, sentience) is also often supposed to be an objectively undetectable parameter. In particular, epiphenomenalism supposes that some systems might be twin zombies (unconscious creatures structurally and functionally indistinguishable from a given human). The property of being associated with a particular material system is perhaps the most important problematic property of consciousness. If consciousness has a neural correlate, then it is possible to develop a scientific test for it (applicable to arbitrary systems, including artificial intelligence).

Postulates

Here I declare the postulates of my approach. I provide some argumentation for them but do not pretend that I prove them. They are, rather, based on common sense. All studies suggested in this paper are correct only in the framework of the postulates used. This should not be considered as a drawback, because the explicit appearance of postulates (although controversial) makes the analysis more transparent. It should be also noted that none of the postulates presumes any particular theory of consciousness.

Postulate 1–1. In order to produce detailed phenomenal judgments about problematic properties of consciousness, an intelligent system must have a source of knowledge about the properties of consciousness.

This postulate is based on the hypothesis that problematic properties of consciousness are so complex that occasional production of detailed phenomenal judgments on them is almost impossible (for example, I neglect the possibility that a random algorithm is able to reproduce the books of Descartes and Leibniz). Direct or indirect causation between someone's consciousness and phenomenal judgments is required. I do not state that each phenomenal judgment is caused by the consciousness of a speaking human. Alternatively, phenomenal judgment might be based on knowledge taken from a book written by another person (see "non-eliminative materialism" panel in Figure 1). Moreover, phenomenal judgment might be caused not only by a human consciousness but also by the God who created it (see "dualism" panel in Figure 1). However, at least indirect causation (correlation between problematic properties of consciousness and phenomenal judgments established by a third factor in the past) must exist (see Appendix for additional discussion).

Postulate 1–2. There are only five basic sources of phenomenal judgments on problematic properties of consciousness. Source 1: neural correlates

of problematic properties of consciousness (producing a phenomenal judgment, a creature describes its own brain structure or functions). Source 2: cognitive errors (producing a phenomenal judgment, a creature describes pseudo-problems). Source 3: immaterial influence on a creature's body. Source 4: innate knowledge (causally related to someone's consciousness). Source 5: external material sources such as discussions and books

Note that Sources 1, 2, and 4 are related to brain structure based on genetic information. Therefore, in some sense, all these sources might be called "innate." The difference is their causal relation to consciousness. Source 1 is directly related to the neural correlate of a creature's own consciousness. Source 4 is related to someone's consciousness and may exist even in an unconscious creature. Source 2 is not related to anyone's consciousness.







Postulates 1–1 and 1–2 are very important in the zombie problem. According to these postulates, zombies (unconscious creatures with normal human behavior) can produce correct phenomenal judgments on problematic properties of consciousness only if they have knowledge about these properties. Source 1 cannot provide such information, because a zombie's internal structure is unconscious and cannot produce self-describing phenomenal judgments. Source 2 might provide some phenomenal judgments, but not about all problematic properties of consciousness (otherwise, eliminative materialism is true, and the term "zombie" is incorrect). Therefore, a zombie requires Sources 3–5 to produce correct phenomenal judgments on at least some problematic properties of consciousness.






Postulate 2–1. If Sources 1–2 are able to provide phenomenal judgments on all known/hypothesized problematic properties of consciousness (in the absence of other Sources), then materialism is true.

Postulate 2–2. In particular, if Source 2 is able to provide phenomenal judgments on all known/hypothesized problematic properties of consciousness, then eliminative materialism is true.

These two postulates are based on the fact that all known arguments against materialism are related to problematic properties of consciousness. By the definition, dualism is true, if at least one property of consciousness has no neural correlate. If, however, all known problematic properties of consciousness have neural correlates (i.e., can be comprehensively studied by objective methods), or simply not exist (related to pseudo-problems), then there is no reason to suppose that consciousness has immaterial "degrees of freedom." The materialistic solution is simpler than the dualistic one. Therefore, *caeteris paribus*, it is preferable (see the Appendix for additional discussion).

The theoretical basis for the study of the mind–body problem can be summarized as follows. Production of phenomenal judgments due to description of

-  biological consciousnesses (in materialism); gray nuances symbolize qualia
-  material mechanisms of phenomenal judgment production
-  human soul (immaterial consciousness in dualism)
-  immaterial mechanisms of phenomenal judgment production
-  problematic experience or information (may produce wrong ideas)
-  information about consciousness in material objects (may be wrong)

-  God (immaterial creator of matter and souls in some forms of dualism)
-  Big Bang (may be based on special initial conditions or physical laws leading to the formation of DNA with information about consciousness)
-  philosophical book (may contain information about consciousness)
-  DNA (may contain information about consciousness)
-  speech about consciousness

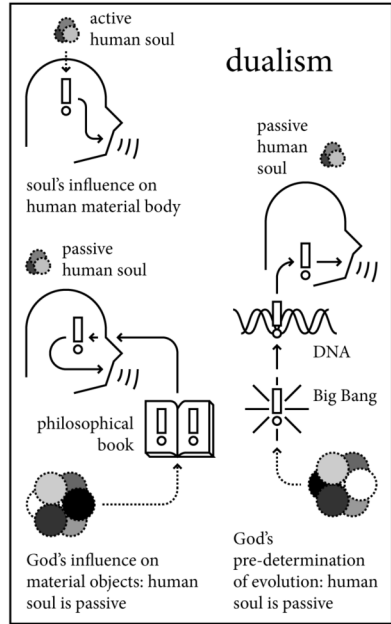
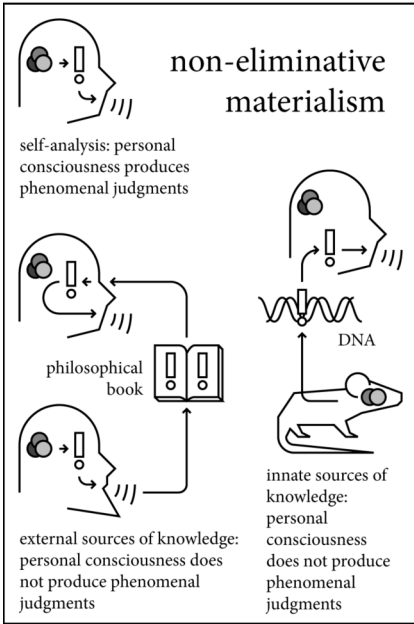


Figure 1: Examples of phenomenal judgment mechanisms in materialism (all sources of knowledge about consciousness are material) and dualism (some sources are immaterial).

the brain's properties (Source 1) is an argument for non-eliminative materialism (see Postulate 2–1). Production of phenomenal judgments due to cognitive errors (Source 2) is an argument for eliminative materialism (see Postulate 2–2). Production of phenomenal judgments due to immaterial influence (Source 3) is an argument for interactionism (see the definition of interactionism). Production of phenomenal judgments due to external or innate sources of knowledge (Sources 4–5) is possible in all theories of consciousness. Complex study of phenomenal judgment mechanisms can give not only “arguments” but also direct scientific verification and falsification of some theories of consciousness (see the next section). In Figure 1, examples of phenomenal judgment mechanisms are shown. Note that external and innate sources of knowledge might be very sophisticated (from philosophical books written by other people to innate knowledge given by God).

Scientific Methods for the Study of the Mind–Body Problem

In this section, I suggest a series of “studies” (some of them may be also called “tests”) focused on the solution of the mind–body problem. Studies 1–3 are the studies of classical computers and computer programs. Classical (deterministic) computations can't be affected by immaterial factors without obvious errors. It is possible to monitor a computer's memory and check that it works according to algorithms (that can't be changed by such factors). Therefore, Source 3 can be easily eliminated in Studies 1–3. However, these studies require very powerful equipment (an intelligent computer). Studies 4–5 are the studies of real humans who definitely have consciousnesses. Real humans have very complex structure. Their physiology involves informational, biophysical, chemical, and, supposedly, even quantum processes. The human brain has no clear algorithm, and there is no simple method to eliminate immaterial influences (that might affect indeterministic quantum processes in cells [Eccles, 1994; Stapp, 1993]).

Study 1. Detection of Phenomenal Judgments Produced by a Deterministic Computer

This subsection describes a simple test that can verify materialism (rigorously¹), detect a machine's consciousness (rigorously), and verify particular information-

¹I do not claim that the suggested studies can prove something in the ideal mathematical sense. Science collects fragmentary experimental facts and interpolates among them. The precision of any experiment is limited and approximations are always present. In this and further subsections the word “rigorous” will mean that some philosophical statement is a logical consequence of some scientific result and postulates (see previous section). This does not mean that a scientific result itself (or postulates) can be rigorously proven.

based materialistic theories of consciousness (non-rigorously). Consider a deterministic (non-quantum) intelligent machine (a computer or a robot) having no innate (preloaded) philosophical knowledge or philosophical discussions while learning. Also, the machine does not contain informational models of other creatures (that may implicitly or explicitly contain knowledge about these creatures' consciousness). If, under these conditions, the machine produces phenomenal judgments on all problematic properties of consciousness, then, according to Postulates 1–1, 1–2, and 2–1, materialism is true and the machine is conscious (if consciousness is not a senseless term). The postulates are applicable here because the machine is deprived of innate and external physical sources of information about problematic properties of consciousness (Sources 4–5), while immaterial factors (Source 3) cannot affect a deterministic machine during the computations.

This test was originally introduced in Argonov (2011). It can be employed not only for verification of materialism, but also for verification of particular materialistic theories of consciousness. However, this application is not rigorous. For example, if somebody thinks that “consciousness is equal to self-learning” then they should build a self-learning machine and test it for consciousness by the described method. Positive results (detection of phenomenal judgments) would be an argument but not a rigorous proof of the tested theory because the machine might have other mechanisms responsible for consciousness. However, complex studies involving several machines of different structures can improve the reliability and validity of the result. The same approach may be employed for verification of any informational (computational) theory of consciousness.

There are limitations to this study. The first limitation is that the study can verify only information-based materialism and eliminative materialism. If, however, consciousness is based on chemical, biophysical, or quantum mechanisms (impossible in the classical computer), then the experiment would demonstrate a negative result. The second limitation is that the study can provide only verification, but not falsification of materialism. A positive result proves materialism but a negative result proves nothing. For example, absence of phenomenal judgments may be caused by lack of the machine's intellect, not by absence of consciousness (at least, in a primitive form).

Beyond these fundamental theoretical limitations, there are several technical problems that may distort the results of the test. The first is human mistakes or unconscious actions during the construction of the machine that may create implicit knowledge about consciousness in software or hardware. In principle, the presence of such knowledge can be discovered in the analysis of memory logs (of phenomenal judgment production). However, a very complex monitoring is required. The second problem is that the machine (having no philosophical education) might have problems with verbal formulation of ideas. The same is true for a human who does not know the philosophical ideas of other people.

It is not easy to produce understandable phenomenal judgments if one is unacquainted with existing discourse and terminology. I suppose that it is merely a technical problem, otherwise philosophy would never emerge in human culture. However, this problem limits language that can be used in the test. Problematic terms such as “consciousness” should be avoided. Maybe it would be better to start the test with the discussion about qualia and religious questions. In human society, even children (e.g., myself in childhood) often have religious hypotheses. An operator may ask the machine: “Can you concede that another computer (identical with you) perceives the red color as you perceive the blue one?” If the computer is not conscious, it most likely will answer “red is red, it can’t be perceived as blue” or “I can’t understand this question.” If the computer is conscious, it can answer “yes” or even “yes, it is a difficult problem. I already thought about it.” Alternatively, an operator may ask: “Did you ever think that your life could be prolonged after the destruction of your body?” A conscious computer might have such ideas. The third (and maybe the most important) problem is that the machine’s design must not be purposefully optimized for the production of phenomenal judgments. If the machine is designed just for passing the test, then such design is an implicit innate knowledge of the problematic properties of consciousness (it is causally related to human knowledge about consciousness). Ideally, the computer should be built according to basic principles of self-learning machines without obvious algorithms of phenomenal judgment production.

In practice, a very intelligent system is needed to produce human-like phenomenal judgments. Real experiments seem to be impossible today due to an absence of intelligent machines with human-like behavior. However, Study 1 can be also performed in thought experiments such as in Argonov (2011), where a self-learning robot seems to be able to produce phenomenal judgments on some problems of consciousness (self, reincarnation etc). However, that robot seems to be unable to understand the problem of qualia.

Study 2. Detection of Phenomenal Judgments Produced by a Deterministic Numerical Model of a Hypothetically Conscious Creature

This subsection describes a more sophisticated test that can verify materialism (rigorously), detect a modeled creature’s consciousness (rigorously), and verify particular materialistic theories of consciousness based on calculable processes (non-rigorously). Consider a deterministic computer containing a numerical model of a hypothetically conscious creature. The model does not contain explicit information about philosophical problems of consciousness, but it may be based on some particular theory of consciousness (this theory will be tested in the study). For example, if one supposes that consciousness has a chemical basis (Anokhin, 1974; Chuprikova, 1985), then the creature’s model must contain

numerical simulation of appropriate chemical processes. If the model correctly describes processes in the creature's material body and produces phenomenal judgments without external or innate sources of knowledge about consciousness, then the modeled creature is also able to produce phenomenal judgments without external or innate sources of knowledge. Therefore, according to Postulates 1–1, 1–2, and 2–1, materialism is true, and the modeled creature is conscious (if consciousness is not a senseless term). The postulates are applicable here for the same reasons as in Study 1. In Study 2, the modeled creature is deprived from external and innate sources of knowledge in the same manner as the computer is deprived of them in Study 1. Note the important differences between Studies 1 and 2. First, in Study 1 the computer must not contain detailed numerical models of other creatures, while in Study 2 it must contain them. Second, in Study 1 a positive result of the experiment (detection of phenomenal judgments) proves that the computer is conscious; while in Study 2 it proves only that the modeled creature is conscious. Study 2 cannot detect the computer's consciousness because the computer contains the numerical model of another creature, and this does not meet the requirements of Study 1.

Most of the fundamental limitations and technical problems for Study 2 are the same as for Study 1 (the word “computer” should be replaced with “creature model”). However, the important new feature is that Study 2 can verify both information- and substrate-based materialistic theories of consciousness, including even some quantum theories (such as Argonov, 2012; Bernroider and Roy, 2004; Ivanov, 1998). Most physicists consider quantum processes as indeterministic (see subsection Study 5), but indeterministic quantum fluctuations can be simulated on a deterministic computer. If a deterministic model of a quantum computer will produce phenomenal judgments, then the difference between computer-generated pseudo-random fluctuations and real stochastic processes is not important in our study. This study, however, is inapplicable to the theory of Penrose (1994), who supposes that some quantum systems may demonstrate incalculable dynamics, and such dynamics are related to consciousness. Such dynamics cannot be simulated in principle.

Study 3. Detection of Phenomenal Judgments in a Deterministic Numerical Simulation of a Human

This study has only a single possible philosophical application: falsification of interactionism (rigorous). Consider a deterministic machine (in particular, a robot) containing a detailed numerical model of a human (including simulation of growth and development) and deprived of external sources of philosophical knowledge (Source 5). However, the robot cannot be completely deprived of innate knowledge (Source 4) because, in principle, such knowledge might exist in a real human. In the beginning of the experiment, the robot contains a model

of a developing embryo. After some time period, the robot begins to move, study language, and communicate with people (avoiding philosophical discussions). If the machine demonstrates normal human behavior (including production of phenomenal judgments on all problematic properties of consciousness), then immaterial influence is not necessary for such behavior, and, by the definition, interactionism is wrong. Even if the robot's phenomenal judgments are based on innate knowledge (Source 4), then a real human also has such knowledge and such phenomenal judgments are a part of "normal behavior."

Study 3 provides only falsification, but not verification of interactionism. A positive result refutes interactionism but a negative result proves nothing. Study 3 is less problematic in its ideology (innate knowledge is not completely prohibited) than Studies 1–2, but it is extremely problematic in pure technological terms. Detailed simulation of human development seems to be much more complex than the construction of basic experimental models of consciousness (in Study 1). Nevertheless, some attempts to create a numerical model of human brain have been already made (Markram, 2006).

Study 4. Study of Human Phenomenal Judgment Mechanisms: The Search for the Neural Correlate of Consciousness and the Search for Cognitive Errors

In contrast with Studies 1–3, this study is not just a "test." It is rather a complex scientific program that, theoretically, can verify materialism (rigorously) and particular materialistic theories of consciousness (non-rigorously).

According to the Postulate 2–1, natural explanation of phenomenal judgment production without Sources 3–5 (see subsection Postulates) can prove materialism in general. The particulars form of materialism can be (less rigorously) determined by the search for neural correlates of problematic properties of consciousness and the study of their role in production of phenomenal judgment. If some brain properties are (1) similar to problematic properties of consciousness and (2) functionally related to the production of phenomenal judgment on them, then these properties, most likely, constitute neural correlates of problematic properties of consciousness. Therefore, non-eliminative materialism is true (phenomenal judgments really describe brain properties, so they are produced by Source 1). If these neural correlates have a purely informational nature, then materialism is information-based. If they are related to physical properties, then materialism is substrate-based. If, however, alternative mechanisms of phenomenal judgment production (not involving brain features similar to problematic properties of consciousness) will be discovered (humans talk about problematic properties of consciousness not because these properties really exist), then there is no reason to believe in the existence of problematic properties of consciousness. These phenomenal judgments contain wrong ideas about consciousness (they are produced by Source 2) and eliminative materialism is true.

Modern science has some arguments for the existence of neural correlate of consciousness (Metzinger, 2000; Wegner, 2003). However, neural correlates of problematic properties of consciousness (such as qualia and unity of consciousness) are not yet found, and science has no comprehensive explanation of phenomenal judgment production. A scheme of a self-educating machine has been demonstrated that seems to reproduce human phenomenal judgment on some philosophical problems (in particular, the idea of soul), but not phenomenal judgment on other problems (in particular, the idea of qualia) [Argonov, 2011].

Study 5. Analysis of Quantum Effects in Phenomenal Judgment Production

The study of quantum effects in the brain has, of course, philosophical implications. In particular, the discovery of a quantum neural correlate of the whole consciousness would prove a substrate-based materialism model (see Study 4). However, I shall discuss only the specific features of research that do not reproduce results of the above-mentioned studies. In particular, I shall show that scientific analysis of quantum effects in the brain may provide falsification (rigorous) and verification (although non-rigorous) of interactionism.

According to the above definitions (see subsection Definitions), interactionism states that: (1) matter is not causally closed (“nonphysical” effects exist), and (2) immaterial degrees of freedom of consciousness have a causal relationship with these effects. In classical mechanics, immaterial influence on matter is impossible: every system evolves according to deterministic equations of motion, and nothing immaterial can change the physical result. Matter contains comprehensive information for the prediction of particle motion with unlimited precision. However, in quantum mechanics, exact predictions are impossible. Most physicists (proponents of indeterministic interpretations such as the “orthodox” Copenhagen interpretation) suppose that random, indeterministic effects are fundamental in quantum mechanics. Some authors (Eccles, 1994; Mensky, 2000; Stapp, 1993) suppose that immaterial consciousness influences random quantum fluctuations in the brain, causing some forms of behavior. These fluctuations might look very similar to random processes but they must contain some additional correlations not predicted by today’s quantum theory (otherwise, intelligent control of behavior is either impossible or produced by physical factors without immaterial influence). Therefore, quantum interactionism must be verifiable or falsifiable in principle.

It must be emphasized that not all proponents of quantum consciousness are interactionists. For example, Argonov (2012), Bernroider and Roy (2004), and Ivanov (1998) suggest quantum but materialistic hypotheses of consciousness. The difference between materialistic and interactionistic quantum theories of consciousness is that interactionism regards quantum systems as “antennae” (receiving control signals from the soul) rather than a real thinking mechanism. Interactionism regards brain quantum systems as windows to other realities rather than normal physical systems.

There are three basic ways to refute interactionism. First, interactionism is wrong if indeterministic interpretations of quantum mechanics are wrong. Today, different interpretations of quantum mechanics are considered as experimentally indistinguishable, but the discussion is not yet finished. For example, Kocsis et al. (2011) made an attempt to measure Bohmian trajectories of photons (predicted by a deterministic Bohmian model). Due to methodological reasons, this experiment is not a rigorous refutation of quantum indeterminism, but it gives at least an aesthetic argument for deterministic interpretation. Second, interactionism is wrong if quantum mechanisms do not take part in information processing in the brain. There are several well-known proponents of quantum consciousness (Bernroider and Roy, 2004; Hameroff, 2006; Penrose, 1994), but most researchers are skeptical regarding these theories, and Tegmark (2000) presented quantitative arguments against the idea of quantum computations in neurons. Third, interactionism is wrong if quantum mechanisms take part in information processing, but the role of these mechanisms can be comprehensively explained by existing theories. Of course, it is problematic to “prove” any result, but it is a completely scientific problem.

Verification of interactionism is more problematic than refutation. Interactionism might be true only if quantum systems in the brain act as black boxes: quantum systems perform complex cognitive operations, but known physical laws can’t explain their functioning. It must be additionally shown that these nonphysical effects are causally related to human subjective phenomena having no neural correlates. For example, a person thinks, “after a minute, I’ll move my hand,” and this thought has no neural correlate, but some quantum fluctuations occur and the hand really moves. Another way to verify interactionism is to show that quantum fluctuations produce phenomenal judgments on problematic properties of consciousness. Such research has obvious methodological problems (lack of knowledge about existence of neural correlates; necessity to take subjective reports into account, etc.), but I suppose that they can be softened by the combined use of several approaches described in this paper.

Summary: Research Program and its Limitations

Studies 1–5 constitute a consolidated research program that, in principle, can give a scientific solution to the mind–body problem. Let me summarize three basic possible results of the research. Two of them help determine the particular nature of consciousness, while the third one leaves some issues unclear.

Result 1. Verification of Materialism and Falsification of Dualism

The most important feature of the research program is its ability to verify materialism. Positive results of Studies 1, 2, or 4 can give scientific support to

materialism (it may be even called “proof” in the present theoretical framework) and, therefore, may be a refutation of dualism.

First, materialism is true if a deterministic computer in Study 1 produces phenomenal judgments on all problematic properties of consciousness. Second, materialism is true if a numerical model of a hypothetically conscious creature in Study 2 does the same. The particular form of materialism (informational, substrate, or eliminative) can be determined in the study of many machines and creature models based on different cognitive and physiological mechanisms (see details in subsections Study 1 and Study 2). Third, materialism is true if Study 4 gives natural explanation to human phenomenal judgment production. In particular, if phenomenal judgments about problematic properties of consciousness really describe some brain properties (neural correlates of problematic properties of consciousness), then materialism is true in non-eliminative form. The particular nature of consciousness (informational or substrate) is determined by the physiological nature of the neural correlate of consciousness. If, however, problematic properties of consciousness have no neural correlates, and all related phenomenal judgments are based on some cognitive errors, then materialism is true in eliminative form.

It should be emphasized that any of these results is sufficient for the materialistic solution of the mind–body problem. For example, if the creature model in Study 2 produces the required phenomenal judgments, then materialism is true, and no other studies are necessary.

Result 2. Verification of Interactionism and Falsification of Materialism

Another important feature of the research program is its ability to verify interactionism and falsify materialism. This can be made within the framework of Study 5. If brain quantum effects take a crucial part in the production of phenomenal judgments on problematic properties of consciousness, but the brain does not contain the neural correlates of these properties (quantum systems “receive” the information about problematic properties rather than produce it), then interactionism is true (see subsection Study 5).

Result 3. Falsification of Interactionism

Studies 3 and 5 can also give a partial solution to the mind–body problem. They can give arguments against interactionism without explicit support of the other two alternatives (materialism and epiphenomenalism).

First, if the robot in Study 3 demonstrates “normal” human behavior and produces phenomenal judgments on all problematic properties of consciousness, then interactionism is wrong. Second, if Study 5 shows that brain quantum effects are not related to problematic properties of consciousness, then interactionism is wrong, too.

However, such results are insufficient for the solution of mind–body problem. Other studies are required to make a scientific choice between materialism and epiphenomenalism, and this may be a problem.

Is it Possible to Verify Epiphenomenalism?

The pro-materialistic results of Studies 1, 2, or 4 refute epiphenomenalism because they refute dualism in general. Therefore, epiphenomenalism is, at least, a falsifiable doctrine. However, none of the studies is focused on the verification of epiphenomenalism. The negative results of Studies 1, 2, and 4 (materialism is not proven) combined with the anti-interactionistic results of Studies 3 or 5 (interactionism is refuted), provide an ambiguous solution: either materialism (Studies 1, 2, and 4 have failed because the natural phenomenal judgment mechanism is extremely sophisticated) or epiphenomenalism (Studies 1, 2, and 4 have failed because materialism is wrong).

Rudd (2000) and Valdman (1997) argued that epiphenomenalism is incompatible with phenomenal judgments. However, I do not share this position. Epiphenomenalism is incompatible only with Sources 1 and 3 (see subsection Postulates), so a creature (even a zombie) can produce phenomenal judgments due to other sources of knowledge about consciousness. The only important restriction is that epiphenomenalism must explain the existence of information about immaterial problematic properties of consciousness in physical reality. The radical epiphenomenalist idea that science (describing physical reality) can completely ignore the existence of consciousness (or, at least, its problematic properties) is wrong. Any theory must explain the existence of the philosophy of consciousness in human culture, and this is a hard question for epiphenomenalism. Two alternative explanations are shown in Figure 1 (fragment “dualism,” schemas regarding the passive human soul).

The first alternative is that the information about problematic properties of consciousness is implicitly “written” in the material world since its creation. For example, according to Leibniz’s idea of “pre-established harmony,” God created consciousnesses and matter and synchronized them. In a deterministic paradigm (classical mechanics and deterministic interpretations of quantum mechanics), the initial conditions of the universe contained comprehensive information on the universe’s future history. Therefore, it may be supposed that humans speak about mind–body problems because God created a detailed “program” of human behavior including phenomenal judgments.

The second alternative is that the information about problematic properties of consciousness is a result of immaterial influence on various physical objects (other than brain). For example, it may be supposed that God controls the evolution of living creatures causing their mutations and changing genetic information. God created implicit philosophical knowledge in DNA, so we have innate knowledge. Also it may be supposed that God created philosophical

books by macroscopic quantum “miracles” (theoretically possible in indeterministic interpretations of quantum mechanics).

Both these alternatives are highly exotic. My personal position is that they are not very realistic, and Studies 1, 2, 4, or 5 will, most likely, support other doctrines. However, I cannot completely exclude the unclear result (interactionism is refuted but materialism is not proven) from consideration. Then, epiphenomenalism may be almost indistinguishable from materialism. Today’s science can’t verify the existence of immaterial influence in a distant past. However, in principle, materialism and epiphenomenalism assume different phenomenal judgment mechanisms, so scientific choice between them seems to be only a technical problem. I hope that in future theoretical works, precise scientific tests for such choices will be developed.

Conclusion

I have built a general phenomenal judgment approach to the mind–body problem. My basic idea is that consciousness is physical if phenomenal judgments about problematic properties of consciousness are produced by purely physical mechanisms. I have proposed a detailed research program for verification and falsification of various forms of materialism and dualism. All suggested tests and methods are focused on the study of phenomenal judgment mechanisms (in humans and machines). Study 1 (originally described in Argonov, 2011) also suggests a novel non-Turing test for machine consciousness.

I understand that some aspects of the suggested approach might seem questionable. I appreciate future discussion on this issue. The main goal of this paper is to demonstrate that the experimental study of “hard” problems is possible in principle. And I hope that it will encourage researchers to further study these “unsolvable” issues.

Appendix: Additional Discussion about the Postulates

Commentary to Postulate 1–1

I expect the following objection: some problematic properties of consciousness seem not very complex. For example, the unity of consciousness might be expressed in four words: “consciousness is something whole.” However, this is a mistake. In practice, complex discussion is needed to explain to another human the essence of each problematic property of consciousness, and Postulate 1–1 states that only conscious and/or a highly educated creature is able to provide such an explanation. Random programs can generate the statement “consciousness is something whole” but not to repeat the books of Descartes or Leibniz.

Note that Postulate 1–1 is weaker than analogous assumptions suggested by other authors. There have been several attempts to use the phenomenal judgment argument for the formulation of some fundamental postulate about consciousness. In particular, Chalmers (1997) said that at least some phenomenal judgments are fully justified because people are acquainted with the phenomenal states that are the objects of such judgments. Valdman (1997) tried to prove the impossibility of zombies using the phenomenal judgment argument. I soften these ideas, and suppose that some unconscious creatures may also produce phenomenal judgments if they have sources of knowledge about consciousness. These sources might be very sophisticated (for example, Leibniz’s “pre-established harmony”). The only thing incompatible with Postulate 1–1 is occasional production of phenomenal judgments on some complex problematic properties of consciousness.

Commentary to Postulate 2–1

Stating that materialism is a preferable theory (in the absence of counter-arguments based on problematic properties of consciousness), I use the positivist principle that any theory should describe phenomena in the simplest manner. This does not mean that, being once “proved,” materialism must be regarded as an eternally true idea. If new anti-materialist arguments will appear in the future, then this position might be changed. The same is true in any scientific branch. Postulate 2–1 might seem a strong claim, but it only follows common scientific practice. For example, the energy conservation law is not verified for all objects in the universe; it is verified mainly on Earth. However, until we do not know experimental facts against it, we consider it as a true law.

References

- Anokhin, P.K. (1974). System analysis of integrative activity of a neural cell. *Uspekhi Fiziolgicheskikh Nauk [Progress in Physiological Sciences]*, 5, 5–92.
- Argonov, V.Yu. (2011). Is a machine able to speak about consciousness? Rigorous approach to mind–body problem and strong AI. In S. Hameroff (Ed.), *Towards a science of consciousness* (pp. 59–59). Tucson: Center of Consciousness Studies.
- Argonov, V.Yu. (2012). Neural correlate of consciousness in a single electron: Radical answer to “quantum theories of consciousness.” *Neuroquantology*, 10, 276–285.
- Ayer, A.J. (1936). *Language, truth and logic*. London: Victor Gollancz.
- Bayne, T., and Chalmers, D. (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The unity of consciousness: Binding, integration and dissociation* (pp. 23–58). Oxford: Oxford University Press.
- Bernroider, G., and Roy S. (2004). Quantum–classical correspondence in the brain: Scaling, action distances and predictability behind neural signal. *Forma*, 19, 55–68.
- Chalmers, D. (1997). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chuprikova, N.I. (1985). *Psyche and consciousness as brain function*. Moscow: Science.
- Churchland, P.S., and Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78(2), 67–90.
- Crick, F. (1995). *The astonishing hypothesis: The scientific search for the soul*. New York: Scribner.

- Davidson, D. (1970). Mental events. In L. Foster and J.W. Swanson (Eds.), *Experience and theory* (pp. 79–101). Amherst: University of Massachusetts Press.
- Dennett, D. (1990). Quining qualia. In W. Lycan (Ed.), *Mind and cognition* (pp. 519–548). Oxford: Blackwells.
- Descartes, R. (1641). *Meditationes de prima philosophia, in qua Dei existentia et animæ immortalitas demonstrator*. Paris: Apud Michaellem Soly.
- Eccles, J.C. (1994). *How the self controls its brain*. Berlin: Springer.
- Edwards, J. (2005). Is consciousness only a property of individual cells? *Journal of consciousness studies*, 12, 429–457.
- Elitzur, A. (1989). Consciousness and the incompleteness of the physical explanation of behavior. *Journal of Mind and Behavior*, 10, 1–20.
- Fodor, J. (1975). *The language of thought*. Cambridge, Massachusetts: Harvard University Press.
- Hameroff, S.R. (2006). Consciousness, neurobiology and quantum mechanics: The case for a connection. In J. Tuszynski (Ed.), *The emerging physics of consciousness* (pp. 193–253). Berlin: Springer.
- Hayek, F.A. (1952). *The sensory order: An inquiry into the foundations of theoretical psychology*. London: Routledge & Kegan Paul.
- Hodgson, S.H. (1870). *The theory of practice. An ethical enquiry in two books*. London: Longmans, Green, Reader, and Dyer.
- Huxley, T.H. (1874). On the hypothesis that animals are automata, and its history. *Fortnightly Review*, 22, 555–580.
- Ivanov, E.M. (1998). *Matter and subjectivity*. Saratov, Russia: Saratov State University Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Kocsis, S., Braverman, B., Ravets, S., Stevens, M.J., Mirin, R.P., Shalm, L.K., and Steinberg, A.M. (2011). Observing the average trajectories of single photons in a two-slit interferometer. *Science*, 332, 1170–1173.
- Leibniz, G.W.F. (1720). *Lehrsätze über die Monadologie*. Frankfurt: Johan Meyer
- Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, 7, 153–160.
- McGinn, C. (1989). Can we solve the mind–body problem? *Mind*, 98, 349–366.
- Mensky, M.B. (2000). Quantum mechanics: New experiments, new applications, and new formulations of old questions. *Physics–Uspekhi*, 43, 585–600.
- Metzinger, T. (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, Massachusetts: MIT Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Nemes, T. (1969). *Cybernetic machines*. Budapest: Iliffe Books.
- Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford: Oxford University Press.
- Place, U.T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 44–50.
- Place, U.T. (1960). Materialism as a scientific hypothesis. *Philosophical Review*, 69, 101–104.
- Putnam, H. (1967). The nature of mental states. In W.H. Capitan and D.D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh: Pittsburgh University Press.
- Rudd, A. (2000). Phenomenal judgment and mental causation. *Journal of Consciousness Studies*, 7, 53–66.
- Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3, 417–424.
- Sevush, S. (2006). Single-neuron theory of consciousness. *Journal of Theoretical Biology*, 238, 704–725.
- Shoemaker, S. (1982). The inverted spectrum. *Journal of Philosophy*, 79, 357–381.
- Stapp, H. (1993). *Mind, matter, and quantum mechanics*. Munich: Springer.
- Tegmark, M. (2000). Importance of quantum coherence in brain processes. *Physical Review E*, 61, 4194–4206.
- Valdman, M. (1997). Will zombies talk about consciousness? The paradox of phenomenal judgment: Its implications for naturalistic dualism and other theories of mind. *Journal of Experimental and Theoretical Artificial Intelligence*, 9, 471–490
- Wegner, D.M. (2003). *The illusion of the conscious will*. Massachusetts: MIT Press.