J. theor. Biol. (1994) **171**, 111–122

Mind as a Force Field: Comments on a New Interactionistic Hypothesis

B. I. B. LINDAHL†§ AND P. ÅRHEM‡

†*Department of Philosophy, University of Stockholm, S-106 91 Stockholm, Sweden and*
‡*The Nobel Institute for Neurophysiology, Karolinska Institute, S-171 77 Stockholm, Sweden*

(*Received on* 8 *December* 1993, *Accepted in revised form on* 5 *May* 1994)

The survival and development of consciousness in biological evolution call for an explanation. An interactionistic mind–brain theory seems to have the greatest explanatory value in this context.

An interpretation of an interactionistic hypothesis, recently proposed by Karl Popper, is discussed both theoretically and based on recent experimental data. In the interpretation, the distinction between the conscious mind and the brain is seen as a division into what is subjective and what is objective, and not as an ontological distinction between something immaterial and something material. The interactionistic hypothesis is based on similarities between minds and physical forces. The conscious mind is understood to interact with randomly spontaneous spatio-temporal patterns of action potentials through an electromagnetic field. Consequences and suggestions for future studies are discussed.

1. Introduction

For a long time the debate about the problem of mind, i.e. the question of how mental life is constituted and how it works, has been dominated by a total renouncement of every form of interactionistic interpretation of the relation between mind and brain. Interestingly, theorists on both sides, opponents as well as proponents of interactionism, agree that the Cartesian theory is wrong—but not *entirely* wrong. The well-known basic problem with the Cartesian theory was the attempt to unite into one theory a dualistic ontology—the distinction between an immaterial mind and a material body—and a materialistic mechanistic notion of causality. Much of the disagreement in the current debate stems from this conflict. To a large extent the two camps for and against interactionism can be said to consist of those, on the one side, who have adopted the dualistic ontology of mind and body and modified the concept of causality accordingly, and of those, on the other side, who have adopted the materialistic notion of causality and adjusted the ontology of mind to this concept.

§ Author to whom correspondence should be addressed.

This is not to say that the distinction between material and immaterial in the interactionistic theories is in every detail Cartesian, or that the notion of causality in the theories of the opposing camp is exactly that of Descartes. The point is that even today the basic issue seems to be whether we should adjust our notion of causality to fit an immaterialistic view on mind, or adjust our view on mind to fit a materialistic notion of causality.

It is in dealing with this issue that the current debate restates the Cartesian dilemma: a materialistic view on mind fails to account for the unique characteristics of consciousness (argue the defenders of interactionism), and a dualistic notion of causality violates certain fundamental principles of physics (argue the opponents). In their respective adherence to an immaterialistic view on mind and a materialistic notion of causality both parties are Cartesian, each in its own way.

Why then do both parties persist in their Cartesian positions? Is there really no way out of this Cartesian dilemma? When trying to answer these questions one must keep in mind the problems that make us need to formulate a theory of the relation between mind and brain. The range and nature of these problems decide

0022–5193/94/210111 + 12 $08.00/0

how well founded and generally applicable the theory needs to be. Differences in these respects are probably at least a part of the explanation why we still seem to be stuck over the Cartesian dilemma. This will be discussed in Section 2.

This paper will examine an interactionistic interpretation of the mind–brain relation. We will concentrate on the relation between consciousness (i.e. not mind in a broader sense, including also the unconscious) and the brain. The conceptual distinction between "the conscious mind" and "the brain" will be seen only as a division into what is *subjective* (the conscious mind) and what is *objective* (the brain). We will argue that the material–immaterial distinction for the most part is played out. Our interest is not primarily in the ontology of consciousness, but in the problem of explaining the existence and development of consciousness in the biological evolution. (This will be further commented upon in Section 2.) The tenability of an interactionistic interpretation of the relation between consciousness and the brain will be discussed by considering some of the major arguments against interactionism. These arguments will be presented in Section 3.

Our discussion of a possible interactionistic interpretation of the mind–brain relation will focus on the theory proposed by Popper (1953, 1955, 1972: chapter 6; 1973; 1978; Popper & Eccles, 1977); the most elaborate of the modern interactionistic theories. Special attention will be paid to the ideas put forward by Popper in a recent article (Popper *et al.*, 1993). The discussion in Section 4 will start with an interpretation of a recent hypothesis about the mind–brain relation proposed by Popper (Section 4.1, with subsections). This interpretation will then be further discussed partly in the light of recent experimental data (Section 4.2, with subsections). The main points and conclusions of our analysis are summed up in Section 5.

2. The Evolutionary Perspective

Even though assumptions about how the brain is related to mental processes underlie and often play a crucial role in physiological and psychological studies, mind–brain issues are seldom at focus in ordinary empirical studies. In detail studies these issues may often be evaded or dealt with in a practical manner.

Relatively few attempts have been made to develop neurophysiological theories about the relation between conscious mental processes and the brain. As Crick and Koch point out, consciousness is seldom even mentioned:

> It is remarkable that most of the work in both cognitive science and the neurosciences makes no reference to consciousness (or "awareness"), especially as many would regard consciousness as the major puzzle confronting the neural view of the mind and indeed at the present time it appears deeply mysterious to many people. (Crick & Koch, 1990)

Most investigations into the problem of consciousness describe the mind–brain relation in rather vague and indirect terms. Roughly, they aim at correlating conscious events with (i) certain parts, or (ii) certain processes of the brain.

Among the studies of the first kind are the investigations of the functional lateralization, pioneered by Sperry *et al.* (1969). Sperry (1977) concludes that surgical disconnection of the hemispheres "yields two conscious minds or selves within the one cranium", whereas MacKay (1966), Penfield (1966) and Eccles (1973) conclude that the unity of the conscious self is preserved. Another example is studies on the visual system. Stoerig & Brandt (1993) call attention to the back-projections to primary areas of visual cortex as one possible explanation of conscious perception.

An example of the second kind of studies is the proposal by Crick & Koch (1990) that the semi-synchronous activity of neurons in the 40–70 Hz range, reported in studies on the visual cortex (Gray & Singer, 1989), is associated with conscious visual experience. To this second kind belong also the studies by Libet (1973; Libet *et al.*, 1979) on the relation between neural activity and perception, indicating that subjectively experienced time need not follow the objectively measured time; findings that Eccles (Popper & Eccles, 1977) and Penrose (1989) suggest support a dualistic view on the mind-brain relation. (For a criticism of this latter interpretation, see Hesslow, 1994.)

Although studies of both kinds contribute to generating and developing plausible mind–brain hypotheses, they provide little or no evidence for or against interactionism.

As Taylor (1903) points out, a physiologist, when being concerned only with certain specific physiological problems, may be justified in treating mental states *as if* they were only epiphenomena, i.e. as if they were caused by physiological events, but not themselves influencing the physiology. Similarly, a psychologist, in dealing with certain specific psychological problems, may treat the physiological processes *as if* they were strictly parallel, i.e. causally unrelated, to the mental events. The "as if" implies, according to these instrumentalistic views, that the theories need not be

taken to be true (only useful), nor do they need to be applicable to other contexts of inquiry than the one at hand. Most, if not all, of the currently dominating theories of mind may be applied in this limited pragmatic way.

In a broader context, however, the significance of the mind–brain issue becomes more obvious and the theories are put to more severe tests. For Descartes, and other scientists of his time, the broader context consisted in the attempt to reconcile the materialistic and mechanistic ideals of science with the conviction that humans possess a free will and an immortal soul. A similarly broad perspective in modern science is the problem of explaining why consciousness has survived and developed in biological evolution. Is it by mere chance? In that case, a parallelistic or epiphenomenalistic theory would do as an explanation. Or has consciousness contributed to the survival of the species possessing this capacity? Then, it seems as if only an interactionistic theory could fully explain this.

Nothing appears to definitely obstruct us from adopting the "mere chance" explanation. But in evolutionary biology this is the last resort, when every effort to find a survival value in a trait has been exhausted. Remarkably little has been written in the theory of biological evolution about consciousness. Richards (1987), however, captures the core of the problem in his summing up of an argument originally formulated by James (1879, 1890):

> Consciousness is a manifest trait of higher organisms, most perspicuously of man; like all such traits it must have evolved; yet it could have evolved only if it were naturally selected; but if naturally selected, it must have a use; and if it has a use, then it cannot be causally inert. Mind therefore must be more than an excretion of brain; it must be (at least in some respects) an independently effective process that is able to control some central nervous system activity. (Richards, 1987: 431)

This issue is also discussed in, for example, Popper & Eccles (1977), Popper (1978) and Eccles (1989).

It is this evolutionary problem that is the reason why we find it worth while to seriously consider an interactionistic theory of the relation between the conscious mind and the brain. We will discuss the tenability of such a theory, not by suggesting possible survival values of consciousness, but by questioning the relevance and validity of the principal objections raised against interactionism.

3. The Criticism

As we noted in the formulation of the Cartesian dilemma, the clash between an ontological mind–body dualism and a materialistic notion of causality is understood by the critics to be crucial in the criticism of interactionism. Four of the major arguments against interactionism (1–4 below) are based on the idea that only material things can influence, and be influenced by, material things.

This view is most explicit in what we can call the mechanistic argument.

1. *The mechanistic argument:* The idea that an immaterial entity (the mind) can influence a material entity (the body) and conversely is not compatible with the notion of causality in natural science, according to which every change in the natural world is produced only by contact of spatially extended bodies. (For a discussion of this concept, see Hesse, 1961: chapter V.)

This argument was raised already by many of Descartes's contemporaries (Williams, 1978: chapter 10). Considering its antiquated conception of matter (as something spatially extended), and its equally outmoded notion of causality (as restricted to action by contact), one would expect this argument to be of only historical interest. It has, however, not entirely lost its influence on the debate. It reappears in partly implicit, partly fragmented ways, and deserves therefore to be paid attention to. For example, P. S. Churchland argues against the existence of "soul-stuff" that is "not spatially extended" (P. S. Churchland, 1986: 318); P. M. Churchland objects to both "Cartesian dualism" and "popular dualism", and points out that the latter is less problematic because, according to that doctrine, mind is "[causally] in contact with" the brain (P. M. Churchland, 1988: 9); Dennett discusses, what he calls, the "standard objection", which "was all too familiar to Descartes" (Dennett, 1991: 33). This objection is basically the mechanistic argument, which Dennett later reformulates in modern terms. However, his illustration to this modern criticism, the incoherent "Casper the Friendly Ghost", who "*both* glide[s] through walls and grab[s] a falling towel" (Dennett, 1991: 35), seems to be more of a problem to the adherents of the mechanistic notions of matter and causality, than to those who are familiar with modern physics. (Already such an early known phenomenon as magnetism seems to restrict the relevance of this illustration.) The mechanistic argument is also discussed by Taylor (1903: book IV, chapter II, § 5) and Popper & Eccles (1977: chapter P5, section 48), who all three argue in favour of interactionism.

A more modern version of the materialistic notion of causality appears in the frequently referred to first law of thermodynamics.

2. *The thermodynamic argument.* The idea of an interaction between an immaterial entity (the mind)

and a material entity (the body) violates the law of conservation of energy (Dennett, 1991: 35–36; see also Taylor, 1903: book IV, chapter II, § 5; Popper & Eccles, 1977: dialogues X, XII).

Given the basic assumption of the criticism that only material things can influence, and be influenced by, material things, it is only logical, the critics argue, to apply Ockham's razor to the postulation of immaterial entities.

3. *The simplicity argument.* To assume the existence of both a material body and an immaterial mind violates Ockham's principle of theoretical parsimony: "... one ought not postulate many items when he can get by with fewer" (Ockham, 1974: 74). (P. M. Churchland, 1988, calls attention to this argument.)

This argument reappears in different forms. Of particular interest to our discussion is the application of this argument to evolutionary theory.

4. *The evolutionary argument.* Modern evolutionary theory does not, and need not, postulate immaterial entities (minds) (P. M. Churchland, 1988: 20–21; cf. P. S. Churchland, 1986: 320–321).

Finally, there is also an influential logical argument.

5. *The logical argument.* The concepts "mind" and "body" are not of the same logical category. Therefore, it makes no sense to talk of mind and body as distinct (or identical) entities, that can (or can not) interact.

This is Ryle's (1949) principal argument. According to Ryle, "mind" refers to an individual's *ability* and *proneness* to do certain sort of things and we can learn about these dispositions by studying the overt behaviour (Ryle, 1949, especially chapter VI, section 4). Similarly, Wittgenstein views behaviour as a criterion (one among several) for someone being in a certain mental state and for the use of psychological language (see Wittgenstein, 1953; Kenny, 1989: chapter 1; Budd, 1991: 17; cf. the view discussed in von Wright, 1994).

Clearly, the relevance of the first four arguments to the interactionistic theory which we will discuss in the next section, depends on which ontological status we ascribe to the mind and the brain—material or immaterial. As we have indicated, our distinction between the conscious mind and the brain is epistemological (subjective–objective) and not ontological (immaterial–material). The ontological status of the conscious mind and the brain is thus still an open question. When discussing this question, we need to consider the fact that "material" is used with many different meanings, for different purposes.

Two conceptions of matter appear in the arguments. The first, in the mechanical argument, is based on spatial extension as a criterion. The action-by-contact notion of causality in this argument seems to imply also impenetrability as another criterion of matter. The other conception of matter, which appears in the thermodynamic argument, uses energy as a criterion.

These two concepts of matter are not always clearly kept apart in the argumentation. (As we noted earlier, it is sometimes difficult to see if the criticism is directed against an interactionistic theory from the distant past, like the original Cartesian theory, or against a theory based on modern physics, or against a popular mixture of both.) Therefore, even though our interest is not historical, we have to take both notions of matter into account in our discussion. (We will also have to discuss other aspects of the definition of matter.)

We will start with the well-known irrelevance of the mechanistic argument to modern physics, and how this opens up the way for a new interactionistic interpretation of the mind–brain relation (Popper's hypothesis). In discussing this hypothesis we will then, in Section 4.2.2, come back to the other ontological arguments and to the logical objection.

4. The Interactionistic Interpretation Further Developed

4.1. THE MIND–FORCE ANALOGY

For Descartes matter was a spatially extended and impenetrable substance. The whole physical universe was understood to be a homogeneous fluid of matter. Individual particles or pieces of matter could be distinguished only through their motion in the fluid. Every physical change was due to action by contact, even magnetism and gravity, and each movement in the universe was taken to be a portion of a constant amount once given by God. Consequently, Descartes repudiated the conception of force on which Kepler and other adherents to the idea of action at a distance based their theories of gravity, and which was later also to become a basis of Newtonian physics. With the continued development of the conception of force, first in terms of action at a distance and subsequently also in terms of fields, ideas were gaining acceptance in physics which were not only incompatible with the mechanistic notion of causality, but which were also to lead to a radically different understanding of the basic elements of physics. The study of electricity and magnetism, especially the works of Faraday and Maxwell, made clear that physics could no longer be built on only the conception of matter in motion, but had to take into account "force fields" as another

fundamental notion. (For a discussion of this whole development, see Jammer, 1957; Čapek, 1961; Hesse, 1961.)

Popper has taken this development as a point of departure for his analysis of the problem of mind.

4.1.1. *Popper's hypothesis*

Popper emphasizes the similarities between minds and forces.

> Minds are (i) located, (ii) unextended, (iii) incorporeal, (iv) capable of acting on bodies, (v) dependent upon body, and (vi) capable of being influenced by bodies.
>
> Most people would say, I think, if one tells them that something with all these properties exists, that it cannot be true. Especially, most materialists would say so, and most physicalists. Now, I say things of this kind *do* exist, and we all know it. So, what are these things? These things are forces. For example, electrical forces. Electrical and magnetic forces have all these properties. (Popper *et al.*, 1993: 168)

Popper adds that minds and forces also have in common that "(vii) they are intensities, not extended but intensities" and that they have "(viii) extension through a span of time" (Popper *et al.*, 1993: 168).

As Popper points out in *The Self and Its Brain* (1977), the analogy between mind and forces is not entirely new. Already Hobbes (1588–1679) and Leibniz (1646–1716) identified a certain part of mind (*conatus*, endeavour, will) with a physical force (see Popper & Eccles, 1977: chapter P5, section 50). Even earlier, Gilbert (1544–1603) in *De Magnete* (1600) "had compared the interaction between magnetic force and a loadstone to that between soul and body" (Watkins, 1974: 410). Of interest in this connection is also the psychological explanation of the origin of the physical concept force, called attention to by Jammer (1957). Both Reid (1710–1796) and Maine de Biran (1766–1824) emphasized our experience of the mind (the will) acting on our body and producing effects in the material world as the source of our universal notion of force (Jammer, 1957: 230–231).

Popper seems to want to go further than to merely an analogy, however. He proposes as a hypothesis

> . . . that the complicated electro-magnetic wave fields which, as we know, are part of the physiology of our brains, represent the unconscious parts of our minds, and that the conscious mind—our conscious mental intensities, our conscious experiences—are capable of interacting with these unconscious physical force fields, especially when problems *need* to be solved that *need* what we call "attention". (Popper *et al.*, 1993: 179)

Here Popper seems to view the "unconscious parts of our minds" as *synonymous* with "physical force fields". He seems to view the electromagnetic field (the unconscious) as an intermediate link between the conscious mind and the neuronal activity. According to this interpretation of Popper's hypothesis (Fig. 1), there are two levels of interaction: the first between a certain spatio-temporal pattern of action potentials and a specific electromagnetic field (the relations 1 and 2); the other between the electromagnetic field/the unconscious, and the conscious mind (the relations 3 and 4). (We will keep to this interpretation. Popper could also be understood to be thinking of an interaction between consciousness and the field, such that the relations 1 and 3 still holds, as in Fig. 1, but consciousness then acts directly on the pattern of action potentials, without passing through the electromagnetic field. Consciousness would then act on the electromagnetic field only through influencing the action-potential pattern. This is perhaps a more far-fetched interpretation; but a direct contact between consciousness and the neuronal activity is what has been discussed by, for example, Beck & Eccles, 1992.)

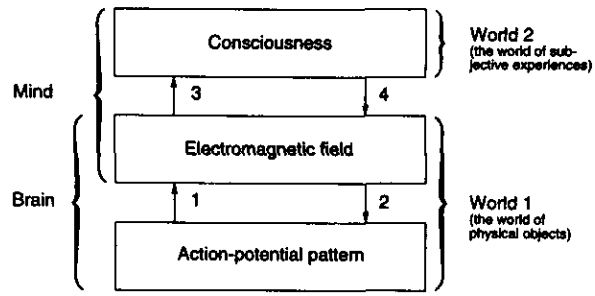

FIG. 1. An interpretation of Popper's hypothesis about the relation between the conscious mind and the brain, based on Popper *et al.* (1993).

In Fig. 1 mind and brain *overlap*. This is a point in Popper's hypothesis. Popper points out that conscious mind may sometimes "sink into physiology" and become unconscious: "a mergent process, a process where [unconscious] mind and brain are no longer really distinguishable" (Popper *et al.*, 1993: 171–172).

> Take a typical case, you learn to play the piano . . . in this there is a stage at which you are very conscious of everything that happens; everything is done consciously. This stage soon disappears . . . The pianist no longer thinks about playing the piano, he thinks about the piece of music which he wishes to present; not at all about how to move his fingers: that is left to the physiology. (Popper *et al.*, 1993: 171)

The decisive distinction therefore seems to be between the conscious mind and the brain.

Popper's hypothesis, in the interpretation of Fig. 1, is both attractive and problematic. The introduction of an intermediate link, the electromagnetic field/the

unconscious, that have properties in common both with conscious mind (the eight properties) and with the spatio-temporal pattern of action potentials (the membership of world 1), makes it somewhat easier to conceive of an interaction between consciousness and the brain, two very different entities in themselves. The most problematic of the two levels of interaction is obviously the 3–4 relation, but also the 1–2 relation needs to be discussed.

Before we enter upon this discussion, we need to put Popper's hypothesis into the broader context of his philosophy.

4.1.2. *The existence of the three worlds*

Popper's view on the role of consciousness in his theory of mind can be best understood in the light of his objectivist theory of knowledge. In his criticism of the theories of knowledge primarily concerned with knowledge as a state of mind, Popper calls attention to knowledge in the objective sense—the logical *contents* of theories, problems, critical arguments, etc, which we find in, for example, journals and books. This is knowledge which exists, Popper points out, without a knowing subject and which is not identical with the words, etc, through which we may acquire this knowledge and make it into our subjective knowledge (Popper, 1972: chapters 3 and 4). In other words, the objective knowledge does not belong to world 2 (the world of subjective experiences) or to world 1 (the world of physical objects), but to a third world, which Popper calls "world 3" (the [world of theoretical, i.e. logical, conceptual, etc] products of the human mind) (the definitions are from Popper and Eccles, 1977: table 1). Popper emphasizes that knowledge in the objective sense has effects on the physical world, and is influenced by the physical world, *through* a knowing subject. This is how Popper most often accounts for the existence and role of consciousness: as an intermediate agent between world 1 and world 3 (Fig. 2).

The point of departure of Popper's argumentation for the existence of world 2 and world 3 is the common-sense acceptance of the reality of world 1: "It is, I think, good common sense to accept the reality or existence of the world 1 of physical bodies ... [and] ... I propose to say that something exists, or is real, if and only if it can *interact* with members of world 1, with hard, physical, bodies" (Popper, 1973; cf. Popper & Eccles, 1977: chapter P1, section 4). Thus, given the reality of world 1, this interaction is, according to Popper, *both necessary and sufficient* for assuming the existence of both world 2 and world 3 (Popper's use of the expression "if and only if" indicates this). In arguing for the existence of world 2, Popper stresses its intermediate role:

> My main argument for the existence of the world 2 of subjective experiences is that we must normally grasp or understand a world 3 theory before we can use it to act upon world 1; but grasping or understanding a theory is a mental affair, a world 2 process: my view is that world 3 usually interacts with world 1 *via* the mental world 2. (Popper, 1973; cf. Popper & Eccles, 1977: chapter P2, and dialogue X)

We shall not go further into discussing the issues of objective knowledge, world 3 and its relations to world 2 and world 1. These issues are important, but lie outside the scope of the present paper. We shall continue to focus on world 2 and world 1. Two things should be noted, however. The first is that Popper uses the interaction with world 1, as a (both necessary and sufficient) criterion for the existence of world 2 and world 3. This is the reversed order of reasoning from the evolutionary approach, called attention to at the end of Section 2 above. According to this view it is the existence of subjective experiences that calls for a causal explanation.

The other thing to be noted is that Popper in arguing for the existence and role of consciousness allows—even emphasizes—that so widely different entities as physical objects, our subjective experiences, and the logical content of theories can interact; world 1, world 2 and world 3 are what Popper calls *causally open* to each other. This is what the opponents of mind–body interactionism find so hard to accept. As Watkins (1974) points out in his discussion of Popper's interactionism, the Cartesian materialistic-mechanistic notion of causality, which, as we have seen, has inspired much of the criticism, "is a special case of the old idea that only like acts on like".

The influence of forces on bodies and the influence of bodies on forces seem to refute this idea of only like acting on like. As Watkins puts it:

> If body–body interaction involves forces, then there is certainly some kind of interaction between *physical* intensities and extended bodies. But if there is interaction between (physical) *intensities* and bodies, if something can act on a body without itself being extended, then there is no longer any compelling reason to deny mind–body interaction. (Watkins, 1974: 395)

By calling attention to the similarities between mind and forces, the eight properties they have in

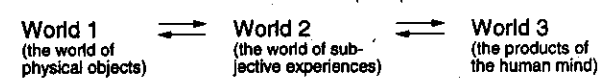

FIG. 2. The interactive relationships between world 1, world 2 and world 3.

common, Popper can be said to have highlighted how the development of physics solved the Cartesian dilemma in its *original* form, i.e. when matter is defined in terms of spatial extension and impenetrability. To what extent the mind–force analogy also sheds light upon other aspects of the Cartesian dilemma—in a more modern interpretation—will be discussed in Section 4.2.2.

4.2. THE MIND–BRAIN INTERACTION

In the following we will more closely examine our suggested interpretation of Popper's hypothesis (Fig. 1). We will, as far as possible, analyse each of the two levels of interaction separately. This should not be understood to imply that one level ought to, or even can, be analysed in complete isolation from the other. What this division of the analysis will amount to, however, is that the more scientific, partly experimental, aspects will be discussed in Section 4.2.1, whereas the more philosophical (the conceptual, epistemological, and ontological issues), will be addressed in Section 4.2.2.

4.2.1. *The action-potential pattern–electromagnetic field interaction*

Relation 1 appears to be the least problematic of the four in Fig. 1. This relation is in principle covered by the Faraday–Henry and Ampère–Maxwell laws, and may therefore in principle be studied within classical electrodynamics. It may, however, in practice be exceedingly difficult to calculate fields caused by the activity of millions or even billions of connected neurons.

Relation 2 (Fig. 1) may at first seem more difficult to accept, especially if one thinks of this as an excitation of *inactive* resting neurons. In order to excite a totally inactive neuron, it is necessary to change the potential over the membrane by 20 mV (Hodgkin, 1964; Hille, 1992). Simple calculations show that under the most favourable conditions an electric field of at least 0·5 V/cm would be necessary. (A field of this strength may excite myelinated axons oriented parallel with the field lines.) However, the electric field around a nerve cell, induced by a normal impulse activity, is many times weaker owing to the low resistance of the extracellular part of the local circuit. Thus, under these circumstances, an electromagnetic field effect on the brain seems highly unlikely.

However, judging from what Popper says in Popper & Eccles (1977: dialogue X), the electromagnetic field in Popper's hypothesis (Fig. 1) does not seem to be expected to trigger inactive neurons, but to sculpture *ongoing* neuronal activity; to affect neurons in constant spontaneous activity. Popper seems to propose that the electromagnetic field affects membranes already fluctuating around a threshold potential. In describing this, Popper draws a parallel with the selection of mutations in biological evolution.

> Thus, what I am here suggesting is that we might conceive of the openness of World 1 to World 2 somewhat on the lines of the impact of selection pressures on mutations. The mutations themselves can be considered as quantum effects; as fluctuations. Such fluctuations may occur, for example, in the brain. In the brain there may at first arise purely probabilistic or chaotic changes, and some of these fluctuations may be purposefully selected in the light of World 3 in a way similar to that in which natural selection quasi-purposefully selects mutations. [. . .] (The all-or-nothing principle of the firing of nerves may indeed be interpreted as a mechanism which would allow arbitrarily small fluctuations to have macroscopic effects.) The action of the mind on the brain may consist in *allowing certain fluctuations to lead to the firing of neurones* while others would merely lead to a slight rise in the temperature of the brain. (Popper & Eccles, 1977: 540–541; our emphasis)

In referring to "quantum effects" Popper seems to assume the existence of a truly random (i.e. not only in practice difficult to predict) neuronal activity. But does such activity really exist? There seems to be some evidence in support of Popper's proposal.

While the mechanism of regular, pacemaker activity is well studied and known in some detail (see Jack *et al.*, 1975; Berridge & Rapp, 1979), the mechanism of random spontaneous activity is not so well understood. Several possibilities are conceivable, however.

The simplest hypothesis is that nerve impulses are triggered by the opening of single ion channels and that the ion-channel gating is random. Let us call this *the ion-channel hypothesis*. (Ion channels are membrane proteins through which the current causing the nerve impulse passes.) That the ion-channel gating is random is indeed the most widely held position (for an opposite view, however, see Liebovitch & Toth, 1991). Some of the hypotheses take the basic randomness to be on a quantum-mechanical level (Donald, 1990). For a single ion channel to trigger a nerve impulse, the cell needs to have rather unique properties. Normally thousands of channels are necessary for exciting a neuron at rest. For a single channel current to induce a sufficient potential change, either the current or the membrane resistance must be unusually large (due to Ohm's law). In some secretory cell types and in olfactory sensory neurons the membrane resistance seems sufficiently large and single

channel induced activity has been recorded (see e.g. Fenwick *et al.*, 1982; Lynch & Barry, 1989). Recently such activity has also been demonstrated in certain brain cells (Johansson & Århem, 1994). These studies thus suggest that single ion channels may cause spontaneous activity in areas of the brain associated with consciousness.

In two respects the ion-channel hypothesis is in line with Popper's proposal above, that the all-or-nothing principle of the firing of nerves allows arbitrarily small fluctuations to have macroscopic effects. The first is that the ion-channel hypothesis suggests that the fluctuations may be caused by quantum events. The other is that the hypothesis assumes that the fluctuations are amplified to macroevents (nerve impulses). It should also be pointed out that it has been observed that *single* nerve impulses may cause conscious experience (Ochoa & Torebjörk, 1983).

The detailed mechanism of channel gating is still not known. It seems likely, however, that the crucial trigger structure is only a fraction of an ion-channel protein. Lee (1992) has proposed as the trigger event an electron transfer between two amino acids, critically located in the ion-channel protein.

The ion-channel hypothesis seems more promising than the alternative hypotheses, based on the probabilistic nature of synaptic activity, which are receiving more discussion. ("Probabilistic" here means that a nerve impulse releases a packet of transmitter molecules at a probability between 0·5 and 0·01 depending on the type of synapse; see Jack *et al.*, 1981; Cunnane & Stjärne, 1982; Korn & Faber, 1987.) One of the alternative hypotheses is *the microsite hypothesis*, proposed by Eccles (1986), based on observations of the specific microstructure of the synaptic apparatus in certain central neurons, and assuming quantum mechanical fluctuations. A weakness of this hypothesis, as an explanation of spontaneous activity, is that the critical structure (part of the presynaptic vesicle) seems several orders of magnitude larger than the critical structure according to the ion-channel hypothesis. Another weakness is that it requires the joint action of a large number of synapses on the same neuron; random release of a single packet of transmitter molecules is as a rule not sufficient to trigger action potentials in central neurons.

These hypotheses, both the ion-channel and the microsite hypothesis, as explanations of spontaneous activity, are still far from being experimentally corroborated. They are bold conjectures. What has been observed are single channel induced action potentials. What is also known is that action potentials induce release of packets of transmitter molecules at a probability lower than 1·0. From an evolutionary perspective, however, these observations call for a functional explanation. This justifies the formulation of the hypotheses.

4.2.2. *The electromagnetic field–consciousness interaction*

The relations 3 and 4 are clearly the most difficult to grasp in Fig. 1. We have in part already dealt with this level of interaction in the discussion of Popper's analogy between conscious mind and forces (Section 4.1.1) and in the comments on his idea of a causal openness of world 1 to world 2 (Sections 4.1.2. and 4.2.1) To sum up, the conclusions of our analysis so far are:

(i) that Popper's basic idea, that "[t]he action of the mind on the brain may consist in allowing certain [purely probabilistic or chaotic] fluctuations to lead to the firing of neurones ..." (Popper & Eccles 1977: 541), seems to require that the potential across the cell membranes already fluctuate around a threshold potential;

(ii) that Popper's suggestion of an electromagnetic field, as an intermediate link between the conscious mind and the neuronal activity, not only makes it easier to conceive of an interaction between two such disparate entities, the conscious mind and the brain (an effect of the family resemblances between the conscious mind, the electromagnetic field, and the action-potential pattern), but also enables us to conjecture a possible physical mechanism between two of the components in the mind–brain relation (the electromagnetic field and the action-potential pattern);

(iii) that the purely probabilistic or chaotic nature of the potential-fluctuations may possibly be explained by the ion-channel hypothesis: that nerve impulses are triggered by the opening of single ion channels and that the ion-channel gating is random.

In this section we shall concentrate on the fundamental conceptual, epistemological, and ontological aspects of our interpretation of Popper's hypothesis, which manifest themselves in the analysis of the relations 3 and 4. In connection with this we will also discuss the relevance and tenability of the major arguments against interactionism, presented in Section 3. We will focus on three-issues: (i) the status of the distinction between world 1 and world 2; (ii) the problem of examining conjectures about the relation between conscious mind and the brain; and (iii) the nature of consciousness.

(i) The first thing one notices about the 3–4 relation is that consciousness and the electromagnetic field belong to two different "worlds", the one physical (world 1), the other mental (world 2). We have already noted some problematic consequences of an ontological interpretation of this distinction. "Physical" is then understood as synonymous with "material" and, consequently, what is not physical (in Fig. 1, consciousness) is then immaterial. This interpretation—brain as material and conscious mind as immaterial—was, as we have seen, the basic problem with an interactionistic interpretation of the mind–brain relation, according to the critics. In the epistemological interpretation, which we indicated at the outset of this paper was the point of departure of our analysis, "physical" is understood as that which is objective.

Also this interpretation has its problems, however. This is clearly illustrated by the physicalistic doctrine of the early logical positivists. According to this doctrine, all scientific statements should be possible to express in a "physicalistic language", i.e. in terms referring only to observable physical properties. Statements about subjective experiences needed therefore to be "translated".

Despite this, it may still be less problematic to analyse the division between world 1 and world 2 as an epistemological distinction, because the distinction does not then depend on our ideas about of what "stuff" consciousness and the brain, respectively, consists. That consciousness is subjective and the brain objective is less of a speculation. But this is not the issue here. What decides the choice of interpretation is the problem we have set out to solve. And since this is about explaining the existence and development of something *subjective*, consciousness, the problem of objectivity cannot be avoided. "Translations" in physicalist or behaviourist terms will be of no use. Theories in the instrumentalist tradition of Ryle (1949), for example the theory proposed by Dennett (1991), and other theories treating attributions of beliefs, desires, intentions and other mental states as if they do not refer to anything involved in causing behaviour, or theories, such as the one put forward by P. S. Churchland (1986), treating such ascriptions as if they do not refer to anything at all, intersubjectively unobservable but yet existing, are not fruitful for our purpose. They all show how we can *do without* the subjective experiences in explaining behaviour and in the use of psychological language. But since our interest is not in explaining behaviour or the use of language, but in explaining consciousness, these theories are of little relevance. These differences in the focus of interest also make, as we see it, the logical argument in Section 3 irrelevant.

Thus, our interest in explaining the existence and development of consciousness makes it practical to interpret the division between world 1 and world 2 as an epistemological distinction. With this as a basis, we may then discuss the ontological requirements for a causal interpretation of the relation between world 1 and world 2; for example whether consciousness needs to be material, and, if so, in what sense.

(ii) But, if we include the first-person view like this, will our analysis then not be restricted to introspection? Not necessarily. Introspection is our point of departure, but does not take us very far as a means of analysis. (We are not primarily interested in *what it is like* to have subjective experiences; what is known as the "qualia" problem. To try to include this in the theory would cause special problems: Svensson, 1994.) It is through introspection that we know that (at least one) consciousness exists. (We can, of course, question the validity of this as well, but there is not much point in that.) And it is this knowledge that (at least one) consciousness exists that gives us reason to wonder how such a phenomenon has survived and developed in our species. Our conjectures about this are not likely to be conclusively demonstrable or refutable, but they have at least to be rationally arguable. (For a discussion of this problem, see Popper, 1972.)

One way to argue, as we have seen, is by analogy. In the mind–force analogy the conjecture that consciousness interacts with the brain was examined, as a first step, by comparing the properties of mind with the properties of forces. Through this Popper called attention to the fact that spatial extension no longer is a requirement of causation. Another, more modern requirement, we have noted, is the exchange of energy (the thermodynamic argument). In this respect there is no evidence about the relation between the world 2 phenomena and world 1. Instead, Popper questions the power of this argument by referring to Schrödinger (1952) and the theory that the law of conservation of energy may be only statistically valid (Popper & Eccles, 1977: dialogues X and XII). Science may, or may not, decide this, and there may be other requirements of causation to be taken into account. What is important here, however, is that the 3–4 relation in Fig. 1 can be analysed, and our conjectures about this relation can be criticized, without "consciousness" being "translated" into terms referring only to something intersubjectively observable.

The remaining two arguments (the simplicity argument and the evolutionary argument in Section 3) are quite tricky. They are either trivially correct or

historically somewhat misleading. How many items one can get by with depends, of course, on how the field of research, or the particular problem one is dealing with, has been defined. If the field or problem has been so defined that it comprises only material processes (meaning, let us say, only processes to which we know the law of conservation of energy applies), then it is self-evident that the stipulation of items beyond this frame would be redundant. This is a truism and does not amount to much. If the arguments are meant to say more, their tenability will obviously be time-bound; what is irrelevant at one point of time, may be relevant at another.

(iii) However, this historical fact also affects any attempt to attribute a special ontological status to a natural trait, like consciousness. Although it may have been justified to classify consciousness as "immaterial", when "material" was reserved for processes that could be analysed in mechanical terms, it seems less motivated in natural science of today. Not only does the division material–immaterial no longer distinguish the domain of science from other areas, and not only has the understanding of matter changed dramatically only during the twentieth century, but there is also no compelling reason, to begin with, why consciousness *as a phenomenon in nature* should not be studied within the natural sciences. (This does not preclude, of course, that conscious mental processes may be studied also from other points of view, within other fields or research.) Methods and theories may have to be adjusted; but is this not what science is all about?

Already the limits of the mind–force analogy need to be further examined. For example, to what extent, if any, can a concept of physical force account for the apparent *autonomy* of mind? Popper raises this issue. He points out that it is not well known how forces are related to *processes* (i.e. biochemical processes) and that we tend to think of forces as something *attached* to bodies, and not as something that can obtain autonomy (Popper *et al.*, 1993). The fundamental question is, as Popper formulates it: "How can these forces, which are set up in the brain, continue themselves, so to speak, and continue to have a kind of identity which is even able to initiate in its turn biochemical processes in the brain?" (Popper *et al.*, 1993: 169).

How well Popper's hypothesis about an intermediate electromagnetic field will stand up to a more thorough scientific and philosophical scrutiny remains to be seen. An interesting consequence that seems to come out of a possible corroboration, however, is that it would seriously challenge the idea of some computer scientists, that the running of the right algorithms on a computer would give rise to consciousness. According to Popper's hypothesis, consciousness would depend on a certain field, and consequently also on the spatial structure of the brain, which so far has been virtually ignored in computer science.

5. Concluding Remarks

What we hope to have shown by this analysis is that, even if most of the currently influential mind–brain theories manage to serve certain limited pragmatic purposes in a satisfactory way, an interactionistic theory has greater explanatory value in a broader context; for example as an explanation of the problem we focused on: how consciousness has survived and developed in biological evolution. We have suggested an interpretation of a central hypothesis within such a theory. This interactionistic interpretation was based on the idea that the conscious mind interacts with randomly spontaneous spatiotemporal patterns of action potentials, through an electromagnetic field. How such a randomly spontaneous neural activity could be possible, we explained by the ion-channel hypothesis. Our examination of the tenability of the interactionistic interpretation is by no means conclusive, but we believe we have demonstrated the insufficient relevance of the major arguments against interactionism to this interpretation.

Much remains to be discussed. We have rejected most of the main arguments against interactionism as irrelevant to our interpretation of the interactionistic hypothesis, because they are based on an alleged ontological difference between mind as something immaterial and brain as material; and we have rejected some of the most influential alternative theories as irrelevant to our problem, because they fail to deal with the subjectivity of consciousness. However, while our interpretation of Popper's distinction between world 1 and world 2, as a division into what is objective and what is subjective, escapes this criticism, it also confronts us with the problem of explaining how objective processes can interact with subjective or, to put it in Popper's terminology, how it is possible for world 1 to be *causally open* to world 2 (which we commented upon in Section 4.1.2 above). Some would say that this problem is impossible to solve, not only for empirical reasons, owing to the subjectivity of consciousness, but also for conceptual and logical reasons, because a theory about this relation cannot be formulated in a purely physicalistic language nor in a purely mentalistic language (cf. the logical argument in Section 3). Nevertheless, we often

use arguments from analogy in science, and it is not unusual that theories are constructed in terms taken from entirely different disciplines. (A clear example of the latter is psychosomatic medicine.) We see no compelling reason why the development of a theory of the role of consciousness in the biological evolution needs to be more restrictive.

Our main conclusion is that Popper's hypothesis of consciousness interacting with neural activity through an electromagnetic field is a thought-provoking suggestion worth closer examination; and that his theory of mind as a whole is possibly the most promising proposal yet made for a future explanation of the survival and development of consciousness.

We owe a particular debt of gratitude to Sir Karl Popper for contributing to the improvement of this article, both through discussions in the course of writing and by constructive criticism of the final drafts. We would also like to thank Lars-Göran Johansson, Dugald Murdoch, and Gunnar Svensson for helpful comments on various versions of this paper. This paper was written within a project financially supported by the Swedish Council for Research in the Humanities and Social Sciences (F724/93). The experimental parts were supported by the Swedish Medical Research Council (B94-04X-06552-12B).

REFERENCES

BECK, F. & ECCLES, J. C. (1992). Quantum aspects of brain activity and the role of consciousness. *Proc. natn. Acad. Sci. U.S.A.* **89,** 11357–11361.

BERRIDGE, M. J. & RAPP, P. E. (1979). A comparative survey of the function, mechanism and control of cellular oscillators. *J. expl. Biol.* **81,** 217–279.

BUDD, M. (1991). *Wittgenstein's Philosophy of Psychology.* London: Routledge.

ČAPEK, M. (1961). *The Philosophical Impact of Contemporary Physics.* Princeton, NJ: D. van Nostrand.

CHURCHLAND, P. M. (1988). *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind.* Revised edition. Cambridge, MA: MIT Press.

CHURCHLAND, P. S. (1986). *Neurophilosophy: Toward A Unified Science in the Mind–Brain.* Cambridge, MA: MIT Press.

CRICK, F. & KOCH, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* **2,** 263–275.

CUNNANE, T. C. & STJÄRNE, L. (1982). Secretion of transmitter from individual varicosities of guinea-pig and mouse vas deferens: all-or-none and extremely intermittent. *Neuroscience* **7,** 2565–2576.

DENNETT, D. C. (1991). *Consciousness Explained.* Boston, MA: Little, Brown and Company.

DONALD, M. J. (1990). Quantum theory and the brain. *Proc. R. Soc. Lond.* **A427,** 43–93.

ECCLES, J. C. (1973). *The Understanding of the Brain.* New York: McGraw-Hill.

ECCLES, J. C. (1986). Do mental events cause neural events analogously to the probability fields of quantum mechanics? *Proc. R. Soc. Lond.* **B227,** 411–428.

ECCLES, J. C. (1989). *Evolution of the Brain: Creation of the Self.* London: Routledge.

FENWICK, E. M, MARTY, A. & NEHER, E. (1982). A patch-clamp study of bovine chromaffin cells and of their sensitivity to acetylcholine. *J. Physiol., Lond.* **331,** 577–597.

GILBERT, W. (1600). *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure; Physiologia Nova, Plurimis & Argumentis, & Experimentis Demonstrata.* Londini: Petrus Short.

GRAY, C. M. & SINGER, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. natn. Acad. Sci. U.S.A.* **86,** 1698–1702.

HESSE, M. B. (1961). *Forces and Fields: The Concept of Action at a Distance in the History of Physics.* London: Thomas Nelson and Sons Ltd.

HESSLOW, G. (1994). Will neuroscience explain consciousness? *J. theor. Biol.* **171,** 29–39.

HILLE, B. (1992). *Ionic Channels of Excitable Membranes,* 2nd edn. Sunderland, MA: Sinauer Associates, Inc.

HODGKIN, A. L. (1964). *The Conduction of the Nervous Impulse.* Liverpool: Liverpool University Press.

JACK, J. J. B., NOBLE, D. & TSIEN, R. W. (1975). *Electric Current Flow in Excitable Cells.* Oxford: Clarendon.

JACK, J. J. B., REDMAN, S. J. & WONG, K. (1981). The components of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia afferents. *J. Physiol., Lond.* **321,** 65–96.

JAMES, W. (1879). Are we automata? *Mind* **4,** 1–22.

JAMES, W. (1890). *The Principles of Psychology,* Vol. 1. London: Macmillan.

JAMMER, M. (1957). *Concepts of Force: A Study in the Foundations of Dynamics.* Cambridge, MA: Harvard Univesity Press.

JOHANSSON, S. & ÅRHEM, P. (1994). Single-channel currents trigger action potentials in small cultured hippocampal neurons. *Proc. natn. Acad. Sci. U.S.A.* **91,** 1761–1765.

KENNY, A. (1989). *The Metaphysics of Mind.* Oxford: Clarendon.

KORN, H. & FABER, D. S. (1987). Regulation and significance of probabilistic release mechanisms at central synapses. In: *Synaptic Function* (Edelman, G. M., Gall, W. E. & Cowan, W. M., eds) pp. 57–108. New York: John Wiley & Sons.

LEE, C.-Y. (1992). A possible biological role of the electron transfer between tyrosine and tryptophan. *FEBS Letters* **299,** 119–123.

LIBET, B. (1973). Electrical stimulation of cortex in human subjects, and conscious sensory aspects. In: *Handbook of Sensory Physiology,* Vol. 2 (Iggo, A., ed.) pp. 743–790. Berlin: Springer-Verlag.

LIBET, B., WRIGHT, E. W. JR., FEINSTEIN, B. & PEARL, D. K. (1979). Subjective referral of the timing for a conscious sensory experience. A functional role for the somatosensory specific projection system in man. *Brain* **102,** 193–224.

LIEBOVITCH, L. S. & TOTH, T. I. (1991). A model of ion channel kinetics using deterministic chaotic rather than stochastic processes. *J. theor. Biol.* **148,** 243–267.

LYNCH, J. W. & BARRY, P. H. (1989). Action potentials initiated by single channels opening in a small neuron (rat olfactory receptor). *Biophys. J.* **55,** 755–768.

MACKAY, D. M. (1966). [Discussion.] In: *Brain and Conscious Experience* (Eccles, J. C. ed.) pp. 312–313. Berlin: Springer-Verlag.

OCKHAM, W. (1974). *Ockham's Theory of Terms. Part I of The Summa Logicae* (Loux, M. J., transl.). Notre Dame, IN: University of Notre Dame Press.

OCHOA, J. & TOREBJÖRK, E. (1983). Sensations evoked by intraneural microstimulation of single mechanoreceptor units innervating the human hand. *J. Physiol., Lond.* **342,** 633–654.

PENFIELD, W. (1966). Speech, perception and the uncommitted cortex. In: *Brain and Conscious Experience* (Eccles, J. C., ed.) pp. 217–237. Berlin: Springer-Verlag.

PENROSE, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford: Oxford University Press.

POPPER, K. R. (1953). Language and the body–mind problem: a restatement of interactionism. In: *Proceedings of the XIth International Congress of Philosophy,* Vol. 7, pp. 101–107. Amsterdam: North-Holland.

POPPER, K. R. (1955). A note on the body–mind problem. *Analysis* **15,** 131–135.

POPPER, K. R. (1972). *Objective Knowledge: An Evolutionary Approach.* Oxford: Clarendon.

POPPER, K. R. (1973). Indeterminism is not enough. *Encounter* **40,** 20–26.

POPPER, K. (1978). Natural selection and the emergence of mind. *Dialectica* **32,** 339–355.

POPPER, K. R. & ECCLES, J. C. (1977). *The Self and Its Brain: An Argument for Interactionism.* Berlin: Springer-Verlag.

POPPER, K. R., LINDAHL, B. I. B. & ÅRHEM, P. (1993). A discussion of the mind–brain problem. *Theor. Med.* **14,** 167–180.

RICHARDS, R. J. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior.* Chicago, IL: The University of Chicago Press.

RYLE, G. (1949). *The Concept of Mind.* London: Hutchinson.

SCHRÖDINGER, E. (1952). Are there quantum jumps? Part I–II. *Brit. J. Phil. Sci.* **3,** 109–123, 233–242.

SPERRY, R. W. (1977). Forebrain commissurotomy and conscious awareness. *J. Med. Phil.* **2,** 101–126.

SPERRY, R. W., GAZZANIGA, M. S. & BOGEN, J. E. (1969). Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection. In: *Handbook of Clinical Neurology,* Vol. 4 (Vinken, P. J. & Bruyn, G. W., eds) pp. 273–290. Amsterdam: North-Holland.

STOERIG, P. & BRANDT, S. (1993). The visual system and levels of perception: properties of neuromental organization. *Theor. Med.* **14,** 117–135.

SVENSSON, G. (1994). Reflections on the problem of identifying mind and brain. *J. theor. Biol.* **171,** 93–100.

TAYLOR, A. E. (1903). *Elements of Metaphysics.* London: Methuen.

VON WRIGHT, G. H. (1994). On mind and matter. *J. theor. Biol.* **171,** 101–110.

WATKINS, J. W. N. (1974). The unity of Popper's thought. In: *The Philosophy of Karl Popper* (Schilpp, P. A., ed.) pp. 371–412. La Salle, IL: Open Court.

WILLIAMS, B. (1978). *Descartes: The Project of Pure Enquiry.* Harmondsworth: Penguin Books.

WITTGENSTEIN, L. (1953). *Philosophical Investigations.* Oxford: Basil Blackwell.