

# Interactionism, Energy Conservation, and the Violation of Physical Laws

Ulrich Mohrhoff

## Abstract

The law of energy conservation is either analytic and not threatened by psychophysical interactionism or contingent upon the causal closure of the physical world which interactionism denies. In either case, interactionism implies departures from the laws of physics, despite attempts to demonstrate the contrary by exploiting the loophole of quantum mechanical indeterminism. These departures are best formulated in terms of modifications, by the conscious self, of the electromagnetic interactions between particles. The electromagnetic vector potential is essentially a summary representation of effects on the motion of particles; these effects can be caused by mental events just as well as by physical events. The interactionist hypothesis implies that the modifications caused by the self cannot be described by mathematical laws. Therefore, the interactions between particles that are modified by a self cannot take place in strict accordance with any physical law. Interactionism, finally, is not irrevocably committed to dualism: it can be incorporated into a panpsychism which permits the self and the particles constituting the body to be thought of as different aspects of the same causal agent.

**Key words:** conservation laws, energy, free will, indeterminism, indistinguishable particles, mental causation, momentum, panpsychism, psychophysical interactionism, vector potential

## 1. INTRODUCTION

A recurrent theme in discussions of the mind-body problem is the allegation that interactionism is incompatible with the conservation of energy and therefore not viable. (Interactionism is the doctrine that mind and matter are two mutually irreducible, interacting domains.) Interactionists have defended their position claiming that by exploiting the loophole of quantum mechanical indeterminism, the nonmaterial mind is capable of influencing matter without violating conservation laws. The present essay argues that both the charge and the defense are misconceived. While the law of energy conservation is either analytic (true by virtue of the meaning of "energy") and therefore not threatened by interactionism or contingent upon the causal closure of the physical world that interactionism denies, the loophole hypothesis violates basic physical laws other than the law of energy conservation. This violation nevertheless cannot be held against interactionism, as transpires from a subsequent inquiry into the epistemological status of the violated laws. Should the material world turn out to be open to nonmaterial influences — and there are no logical or *a priori* reasons why this cannot be the case — we *must* expect departures from the physical laws. Nor should interactionism be blamed for the shortcomings of metaphysical dualism. While dualism is a legitimate and sufficient framework for formulating the interactionist hypothesis in an empirically testable manner, interactionism is not irrevocably committed to dualism. This will

be made obvious by sketching out a panpsychist framework that encompasses all essential tenets of interactionism.

Section 2 reviews the debate about energy conservation. Most of Sec. 3 is devoted to the view that the conservation laws of energy and momentum are analytic: whenever it is possible to define "energy" and "momentum," the quantities denoted by these concepts are necessarily conserved. If it were true, interactionism could not entail a violation of their conservation, but like the general theory of relativity it might undermine their definitions and, consequently, restrict their employment. Section 4 explains why it would nevertheless be chimerical to hope that by exploiting the loophole of quantum mechanical indeterminism, interactionism could steer clear of any violations of physical laws.

But if interactionism renders modifications of physical laws unavoidable, there is no longer any advantage in accounting for the causal efficacy of the nonmaterial mind in terms of modified quantum mechanical probabilities. Any physical effect that can be accounted for in terms of a direct modification of such probabilities can then just as well be accounted for in terms of a suitable modification of the physical entity that determines the probabilities during the deterministic phase of their evolution, the electromagnetic vector potential.<sup>1</sup> Quantum mechanical probabilities cease to be essential for formulating the interactionist hypothesis. The causal efficacy of the nonmaterial mind becomes a "classical" effect.<sup>2</sup>

There are two reasons for preferring the latter account, one of which is an unseemly feature of the former account. This consists in a disparateness between the respective treatment of effects caused by matter and effects caused by mind, or in having probabilities determined twice over, once during their deterministic evolution by the physically determined vector potential and once at the end through direct modification by the mind. By ranking nonmaterial mental events alongside the physical contributors to the vector potential, this disparateness is removed. The second reason ensues from the epistemological status of the vector potential. As a summary representation of possible effects on moving particles that makes no reference whatever to the nature of the corresponding causes, the vector potential ought to represent the effects of both material and nonmaterial causes alike. All of this is the subject of Sec. 5.

Metrical inhomogeneities are the only means of mathematically representing influences on the motion of particles. If there exists an experimental realm in which the inhomogeneities due to mind are absent or negligible, then it makes sense to attribute energy and momentum to matter even where matter is open to nonmaterial mental influences and energy and momentum are not conserved as a matter of course. This conclusion is reached in Sec. 6. It disposes of the view that the conservation of energy and momentum is an analytic truth. The sketch of a panpsychist extension of interactionism is given in Sec. 7.

A semiclassical presentation has been adopted in this essay. The quantum mechanical sum-over-histories approach is used to introduce the classical vector potential and to explicate the meaning of this theoretical construct. Second quantization does not significantly alter its meaning, not least because the indeterminacy of the vector potential to which second quantization gives rise, is irrelevant to discussions of the causal efficacy of the nonmaterial mind. Thinking of the vector potential as an externally prescribed classical field is therefore adequate for the purpose of the present study.

## 2. ENERGY CONSERVATION AND THE INTERACTIONIST HYPOTHESIS

Attempts to address the mind-body problem along interactionist lines have traditionally been faulted for taking liberties with physical conservation laws, notably the principle of the conservation of energy (also known as the first law of thermodynamics). Bunge and Dennett will speak for the prosecution:

If immaterial mind could move matter, then it would create energy; and if matter were to act on immaterial mind, then energy would disappear. In either case energy...would fail to be conserved. And so physics, chemistry, biology, and economics would collapse.<sup>(1)</sup>

Let us concentrate on the returned signals, the directives from mind to brain. These, *ex hypothesi*, are not physical; they are not light waves or sound waves or cosmic rays or streams of subatomic particles. No physical energy or mass is associated with them. How, then, do they get to make a difference to what happens in the brain cells they must

affect, if the mind is to have any influence on the body? A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy, and where is this energy to come from? It is this principle of the conservation of energy that accounts for the physical impossibility of "perpetual motion machines," and the same principle is apparently violated by dualism. This confrontation between quite standard physics and dualism has been endlessly discussed since Descartes' own day, and is widely regarded as the inescapable and fatal flaw of dualism.<sup>(2)</sup>

Dualists on the whole have taken these strictures to heart. Even Popper, by proclaiming himself not to be "in the least impressed by the danger of falling foul of the first law of thermodynamics,"<sup>(3)</sup> implicitly acknowledges the danger. From the early days of quantum mechanics, the strategy of the defense has consisted in claiming that quantum mechanical indeterminism allows nonmaterial mental events to act on matter (specifically the brain) without infringing conservation laws. Eddington<sup>(4)</sup> was probably the first to publicly speculate that the mind may influence the body by affecting the configuration of quantum events within the brain through a causal influence on the probability of their occurrence.

More recently, Margenau<sup>(5)</sup> has suggested that the mind may be "regarded as a field in the accepted physical sense of the term," yet not be "required to contain energy in order to account for all known phenomena in which mind interacts with brain": "In very complicated physical systems such as the brain, the neurons and the sense organs, whose constituents are small enough to be governed by probabilistic quantum laws, the physical organ is always poised for a multitude of possible changes, each with a definite probability."

Standard axiomatizations of quantum mechanics recognize two kinds of change: the probabilistic collapse of a quantum mechanical superposition that occurs during a measurement and the deterministic evolution of the quantum state that takes place between measurements.<sup>(6)</sup> What Margenau proposes is that the causal efficacy of the mind rests on the following sequence of steps: (1) the relevant physical system develops, in accordance with the deterministic ("type II") evolution of states, into a superposition of alternative states, each associated with a probability; (2) the mind "loads the dice" by altering the physically determined probabilities or by superimposing its own probability field on the physically determined probability field; (3) the resulting superposition collapses to one of its elements in accordance with the probabilistic ("type I") change of states.<sup>3</sup> In this way, Margenau maintains, the mind can act on the brain without disturbing the balance of energy.<sup>(7)</sup>

Seizing on Margenau's proposal, Eccles, in collaboration with Beck,<sup>(8,9)</sup> has put forward the most elaborate and specific hypothesis of mind-brain interaction to date. It capitalizes on the basic unitary activity of the cerebral cortex, exocytosis. Exocytosis is the emission of a synaptic transmitter into the synaptic cleft by a vesicle of the presynaptic vesicular grid, a paracrystalline structure situated inside the terminal expansion

(bouton) of a nerve fiber. It is an all-or-nothing event, which has been found to occur with a probability of about one-fourth to one-third when a bouton is activated by a nerve impulse. Eccles and Beck assume this probability to be of quantum mechanical origin. They cite increasing evidence of a trigger mechanism that may involve quantum transitions between metastable molecular states and propose a model for the trigger mechanism based on the tunneling of a quasiparticle through a potential barrier.<sup>4</sup> During a period of the order of femtoseconds, according to their model, the quasiparticle is distributed over both sides of the barrier. One side corresponds to the activated state of the trigger, the other side to the nonactivated state. At the end of this period exocytosis has been triggered with the aforesaid probability. The Eccles-Beck hypothesis of mind-brain interaction (Part I) is that mental intentions act through a quantal probability field altering the probability of exocytosis during this brief period.

While this change in probability of exocytosis by a single vesicle is many orders of magnitude too small for modifying the patterns of neuronal activity even in small areas of the brain, there are many thousands of similar boutons on a pyramidal cell (the principal type of neuron of the cerebral cortex), and there are about 200 neurons in the region of a dendron (the basic anatomical unit of the cerebral cortex). The hypothesis of mind-brain interaction according to Eccles and Beck (Part II) is that mental intention becomes neurally effective by momentarily increasing the probabilities for exocytosis in the hundreds of thousands of boutons in a whole dendron.

In summary it can be stated that it is sufficient for the dualist-interactionist hypothesis to be able to account for the ability of a nonmaterial mental event to effect a changed probability of the vesicular emission from a single bouton on a cortical pyramidal cell. If that can occur for one, it could occur for a multitude of the boutons on that neuron, and all else follows in accord with the neuroscience of motor control.<sup>(10)</sup>

It is reassuring that all of the richness and enjoyment of our experiences can now be accepted without any qualms of conscience that we may be infringing conservation laws!<sup>(11)</sup>

### 3. ENERGY CONSERVATION: A CLOSER LOOK

Originally momentum had been defined as "mass times velocity." It soon became apparent that (within Newtonian physics) this was a conserved quantity. Then the special theory of relativity superseded Newtonian physics, and mass times velocity was no longer conserved. By this time, however, the property of being conserved was accorded much greater importance than the original definition in terms of mass and velocity. Momentum accordingly was redefined to match its original definition in the low-speed limit, where the two theories make identical predictions, as well as to retain its status of a conserved quantity.<sup>5</sup>

But a redefinition that consists in the substitution of one theory-dependent *definiens* for another can only be a halfway

stop. It must be possible to define the *definiendum* at a more basic level, independently of the specific principles of either theory and hence in a way that is valid for both. It indeed soon transpired that the different mathematical embodiments of momentum in the respective theories of Newton and Einstein were specific instances of a quantity that could be invariantly defined for a large class of theories. According to Noether's 1918 theorem, there exists in all theories that can be based on a minimum principle (i.e., all theories permitting the formulation of a Lagrangian) a deep connection between symmetry principles and conservation laws. In all such theories, including all experimentally well-confirmed theories to date, the homogeneity of space implies the existence of a locally conserved quantity. And this locally conserved quantity, the existence of which is implied by the homogeneity of space, is what is meant by "momentum." Thereafter it was possible to claim that this has always been the true definition, even when the concept was insufficiently differentiated from its then sole instantiation, "mass times velocity."

What has just been said about momentum and the homogeneity of space holds equally true, *mutatis mutandis*, of energy and the homogeneity of time. Both energy and momentum are defined as *conserved quantities*. They are conserved *by definition*. Either they make sense and are conserved, or they don't make sense, and this should put an end to talk about possible violations of the conservation of energy or momentum. The only way for these quantities not to be conserved is to be meaningless.

So when are they meaningless? The answer is: whenever it is not the case that space and/or time are, or can be considered as being (approximately), homogeneous. Homogeneity of space consists in the mechanical equivalence of all locations, homogeneity of time in the mechanical equivalence of all moments. Differently put, these homogeneities express the fact that any closed mechanical system behaves in the same way anywhere and/or anytime. This has the nature of a postulate: differences in the outcomes of identical experiments performed at distinct locations and/or times are to be ascribed to the different physical conditions (known or unknown) prevailing at these distinct locations and/or times, not to these locations and/or times *per se*. An instantiation of the synthetic *a priori* judgment that everything that happens has a cause, it has to do with what we (investigating humans) make of our experiences rather than with any particular experience of ours. If we did not assume the existence of a cause, we would not look for one; and if we did not assume the existence of physical causes to explain the apparent spatial or temporal inhomogeneities we observe, we would not look for such causes but rest content with attributing the inhomogeneities to space or time *per se*.

And so it would seem that the homogeneity of space and the homogeneity of time are *a priori* certain; that momentum and energy are always well-defined; and that they are always conserved. However, there are riders to this series of conclusions. Whatever is *a priori* certain is so only with regard to our mental constructs. Whether or not these can consistently be thought of as descriptions of objective reality is a matter beyond our control. Also, before anything can be derived from the said

homogeneities, they must be given formal expression within the framework of a physical theory. And there is no *a priori* guarantee that this is possible. In fact, there are reasons to surmise the opposite, as will become apparent in what follows.

There is nothing controversial about the way in which space and time are rendered manifestly homogeneous or the way in which their respective homogeneities find mathematical expression in a physical theory. One employs either an inertial coordinate system or, if a coordinate-free formalism is preferred, a metric with vanishing connection coefficients (flat space plus affine time in Newtonian physics, flat space-time in relativistic physics).<sup>6</sup> But that which is capable of manifesting homogeneity also lends itself to the manifestation of inhomogeneities. The metric needed to manifest flatness could instead serve to manifest (Riemannian) curvature. In other words, the metric texture of space or space-time offers a handle for the formulation of an interaction law. Matter could act on matter via the intermediate representation of the metric connection (curvature, for short) in much the same way as electric charges act on electric charges via the intermediate representation of the electromagnetic field. The curvature at any given space-time point  $p$  could depend on the distribution and motion of matter on any spacelike hypersurface  $h$  intersecting with the past light cone of  $p$ . It could thus represent an influence on the motion of matter at  $p$  due to the distribution and motion of matter on the intersection of  $h$  with the past light cone of  $p$ .

It is no secret that this *a priori* conceivable possibility is an actual feature of the objective world. The interaction in question is gravity; the theory just outlined is the general theory of relativity. Now, gravity appears to be quite indispensable to the creation of what Squires<sup>(12)</sup> has called an "interesting world." Without gravity there would exist no stars, no planets, nor (for all we can imagine) any sites hospitable to anything as interesting as life. In view of this it might be asserted that curvature is implied by our own existence, although to be able to make this assertion appears to presuppose the wisdom of hindsight.

Be that as it may. The metric connection lends itself to the manifestation either of spatiotemporal homogeneity or of gravity. As far as the description of objective reality is concerned, the choice is not ours but nature's. And nature has opted for gravity. The metric that could have offered a handle for the incorporation, in our mental picture of reality, of a homogeneous space and a homogeneous time is already used up. From this and what has been said earlier one might draw the conclusion that in situations in which gravity plays a significant role, energy and momentum are undefined. But in doing so one ignores that curved space-time is locally flat and that, as a consequence, the combined energy-momentum of all *nongravitational* fields is locally conserved. This is sufficient for it to be well-defined even though it is not globally conserved. What is ill-defined in any generic space-time is the gravitational and hence the total energy-momentum. The energy-momentum associated with a curved region of space-time is, strictly speaking, definable only in model space-times that are asymptotically flat.<sup>(13)</sup> (This does not mean that gravitational energy-momentum cannot be *approximately* defined wherever and whenever space-time can be con-

sidered flat to a good approximation. There is an important theory in its own right, the so-called linearized theory of gravity,<sup>(14)</sup> which treats departures from the global Minkowskian metric as a gravitational field in flat space-time. This deals adequately with such areas as terrestrial ballistics and interplanetary travel.)

At certain junctures in the history of physics the law of energy conservation has been called in question. Bohr at one time felt that he had to renounce it, and not a few particle physicists despaired of it before the neutrino was proposed and, in due course, discovered. It is not to be supposed that these physicists were unaware of the deep connection between the conservation laws for energy and momentum and the homogeneity of time and of space. Rather they were driven to contemplate the possibility that these homogeneities were not, after all, respected by nature. Bohr considered the problems facing atomic theory to be "of such a nature that they hardly allow us to hope that we shall be able, within the world of the atom, to carry through a description in space and time that corresponds to our ordinary sensory perceptions."<sup>(15)</sup> If the feasibility of such a description cannot be taken for granted, the homogeneity of space and of time cannot be taken for granted either.

More recently, in connection with the so-called reduction problem in quantum mechanics, the stochastic generation (and hence nonconservation) of energy has emerged as a theoretical possibility.<sup>(16,17)</sup> This amounts to introducing stochastic inhomogeneities in the "flow" of time and to redefining energy as the quantity whose conservation *would* be implied if those inhomogeneities (as well as other possible violations of time-translation invariance) *were* absent. Needless to say, if this definition is adopted, the view that the conservation of energy is analytic (energy being either conserved or undefined) can no longer be entertained.

The situation, then, is this: as long as the said view is tenable, the interactionist hypothesis cannot imply any violation of energy conservation. And if physicists can invoke inhomogeneities in the "flow" of time and define energy in such a way that it is conserved only when and where those inhomogeneities are absent, so can interactionists. The causal efficacy of nonmaterial mind could rest on its generating similar (but not stochastic) inhomogeneities.

#### 4. INTERACTIONISM VIOLATES PHYSICAL LAWS

While the argument that interactionism violates conservation laws is either groundless or carries no weight, the notion that mental events could influence physical events through the loophole of quantum mechanical indeterminism, without in any manner whatsoever infringing on the deterministic regime of physical laws, is chimerical, as the following will show.

Consider an ensemble of identical physical systems that are capable of being influenced by mental intention (such as the intention to flex the right index finger) and that invariably produce the intended action if so influenced. If these systems are initially in identical states and are influenced by identical mental intentions, then there is no randomness in the causal concatenation between intention and intended action. Throughout the

ensemble, the same mental event brings about the same physical event. If this regular causal connection is postulated to involve modifications of quantum mechanical probabilities associated with "collapsible" wave functions, these modifications must be statistically significant. In the simplest case in which the modifications amount to the selection of a definite outcome in a single collapse, the same definite outcome must be selected throughout the ensemble. In the Eccles-Beck model, in which weak modifications accumulate over a large number of collapses to produce a definite effect, the fact that equal definite effects must be produced throughout the ensemble entails that the individual modifications likewise exhibit statistically significant trends. If  $S$  is a set of individual collapses, each occurring in a different member of the ensemble but all playing identical roles in the production of identical definite effects, the modifications of the probabilities associated with the collapses must be statistically significant across  $S$ .

In other words, what gets modified is not merely the individual quantum event but the statistics of entire ensembles of such events. And these statistics, unlike the individual events, are rigorously determined by physical laws. Changing them means changing the physical laws. Altering the probabilities associated with individual measurement-like events without changing the laws of physics is possible only if the relative frequencies associated with every ensemble of identical such events remain unaltered. But this is possible only if the individual modifications of probability are themselves probabilistic in nature; and if this were the case there could be no talk about causation, mental or otherwise. An event that causes nothing but random modifications of quantum mechanical probabilities cannot be causally effective in the realm of definite occurrences.<sup>7</sup>

The basic tenet of the dualist-interactionist position — causal openness of the material to the nonmaterial mental — thus entails a violation (i.e., an occurrence of modifications) of physical laws.<sup>8</sup> Probability distributions, determined jointly by initial conditions and some quantum mechanical equation of motion such as the Schrödinger equation, are altered. One might leave it at that. But also one might wonder if any such alteration cannot just as well be formulated in terms of the well-known physical quantities that determine probability distributions during the deterministic phase of their evolution. The answer is that it can.

## 5. INTERACTIONISM: THE "CLASSICAL" VERSION

As an illustration of how the altered probability distributions entailed by the interactionist hypothesis could arise within the formalism that physicists use to calculate probability distributions, rather than as *ad hoc* modifications of the results of the calculations, an open one-particle system will be considered. It seems reasonable to assume that mental events do not cause the creation or annihilation of particle pairs. It likewise stands to reason that the weak and strong forces need not be considered as possible vehicles of mental causation. Both are short-range forces. The strong force is confined to the interior of color-neutral mesons and baryons, with a residual force, the nuclear force, which in brains is very much confined to the interior of atomic nuclei, as is the weak force. (If the weak force could

reach from the flavored quarks inside the atomic nucleus to the electrons surrounding the nucleus, it would sooner rather than later turn these into neutrinos, with disastrous consequences for the stability of matter.) But the goings-on inside atomic nuclei do not seem to have any influence on when neurons fire, or how likely they are to fire, which is where mental causation must make itself felt. For the purpose of studying the possible effects of nonmaterial mental events on the motion of a particle, it therefore suffices to consider an isoscalar particle, for only the weak and strong forces cause transformations in internal symmetry spaces — only they require the particles on which they act to be isospinors. And since it turns out that the most general formulation of effects on the motion of a spinless particle already includes the possible effects on a spinning particle, we may further confine our discussion to a spinless particle.

The entire physics of such an open one-particle system is formally contained in the probability amplitude  $\langle \psi_f | \psi_i \rangle$  for the transition from the particle's initial state  $|\psi_i\rangle$  at time  $t_i$  to its final state  $|\psi_f\rangle$  at time  $t_f$ . This in turn is known, in principle, if the probability amplitude  $\langle z|x \rangle$  for the particle to travel from point  $x$  to point  $z$  in the time between  $t_i$  and  $t_f$  is known.  $\langle z|x \rangle$  may be calculated<sup>(18)</sup> by summing over all timelike curves connecting  $x$  and  $z$ . Each directed curve  $c$  from  $x$  to  $z$  contributes a complex number of magnitude 1 and phase  $\phi[z,x;c]$ . It is readily shown (using axioms of standard quantum mechanics) that  $\phi$  is additive in the following sense: if  $y$  is a point on  $c$  between  $x$  and  $z$ , then

$$\phi[z,x;c] = \phi[z,y;c] + \phi[y,x;c]. \quad (1)$$

This additivity makes it possible to interpret  $\phi$  as a length (in the sense of a Finsler geometry) of curves in space-time. For a freely propagating particle this *mechanical length* is proportional to the geometric length of  $c$  in a locally Minkowskian space-time,

$$d\phi = -mds, \quad (2)$$

where  $d\phi$  is the mechanical line element,  $ds$  is the geometric line element that can (at least locally) be given the standard Minkowskian form, and  $m$  is the mass of the particle. (A system of units in which  $\hbar$  and the velocity of light are equal to unity has been adopted.)

Clearly, the only way of influencing the particle's motion is to modify  $d\phi$ , or the mechanical lengths of curves in space-time.<sup>9</sup> This can be done either in the manner of gravity, by warping space-time itself, or by effectively bending<sup>10</sup> the curves of extremal mechanical length — the classical trajectories — relative to the (locally) flat metric background. When it is weak enough to permit a human brain to function normally, gravity plays no significant role in a region of space the size of a brain; so only the latter option needs to be considered.<sup>11</sup> Because the inclusion in  $d\phi$  of any nongravitational effects on the particle's motion must leave the free term (2) unaffected, it can only consist in the *addition* of a term to (2). If we further assume that inertial behavior is determined (as it is known to be) by gravity

alone, the additional term must be linear in the coordinate differentials,<sup>12</sup> that is, it must be of the form

$$-qA_k dx^k. \quad (3)$$

Relativistic invariance rules out the dependence of  $A$  on more than one position. It also requires the whole term to be a scalar, so that  $A$  itself must be a one-form.

So the one and only way of influencing the motion of a particle under the above restrictions is to contribute to a vector field  $A$ . This field is the summary representation of all possible nongravitational effects on the motion of a spinless isoscalar particle, when pair creation or annihilation can be ignored. But it is well known that

$$d\phi = -mds - qA_k dx^k \quad (4)$$

is the mechanical line element for a particle of (electric) charge  $q$  moving in a given electromagnetic field represented by the vector potential  $A$ . The punch line is that the vector field, in contributing to which the one and only way of influencing the motion of a particle (under the above restrictions) consists, is none other than the electromagnetic vector potential.

This isn't exactly how the vector potential is introduced in the textbooks. We tend to associate it not only with the way in which the motion of charged particles is influenced by it, but also with a *particular* way (given by Maxwell's laws and the well-known formulas expressing the electric and magnetic fields  $E$  and  $B$  in terms of  $A$ ) in which the motion and distribution of charges contribute to it. And we have let ourselves be bamboozled into thinking that this is the *only* way of contributing to it. But, in fact, *anything* — be it physical, mental, or whatnot — that (under the above restrictions) has an effect on the motion of a particle will necessarily contribute to the electromagnetic vector potential. If a mental event is to influence the behavior of the quasiparticle in Eccles' model of a trigger mechanism for exocytosis, it must modify the barrier — a *potential* barrier — penetrated by the quasiparticle.<sup>13</sup>

When the electromagnetic field (which we now think of as the exterior derivative of the one-form  $A$ ) was introduced by Maxwell, it was thought of as the property of a mechanical substrate pervading space. When Einstein disposed of this substrate, the erstwhile property became a physical entity in its own right. The symbol took on a life of its own; the mathematical description took the place of the thing described. Today many physicists opine that reality is mathematical. While the present investigation ought not to be biased in favor of any such metaphysical claim, it is safe to say that the empirical reality investigated by science is, first of all, a complex of mental constructs. (I am not saying that it is “nothing but” mental constructs.) What these constructs have in common, and what distinguishes them from mere fantasies, is that they are objectifiable, that is, they are capable of being consistently thought of as features of an objective world. The vector potential is such a construct (after quantization, at any rate), and from the role it plays in our account of particle motion it is clear that it

cannot be partial to any particular type of causal agent. It serves to represent the effects of mental causes just as well as those of physical causes.<sup>14</sup>

If nonphysical causes do indeed contribute to (the physically relevant aspect of) the vector potential, the well-known dynamical laws for the vector potential (i.e., Maxwell's laws or their quantum mechanical counterparts) would be violated, in the sense that they would describe *some* but not *all* contributions to (the physically relevant aspect of) the vector potential.<sup>15</sup> It is worth emphasizing that there are neither theoretical nor experimental reasons to rule out such a violation. While empirical evidence of nonphysical contributions to the vector potential may as yet be lacking, absence of evidence is not evidence of absence, as they say. Systems in which such contributions would occur are notoriously complex, difficult to analyze, and no less difficult to experiment with. It also stands to reason that the nonphysical contributions to  $A$  *cannot* represent more than minuscule modifications of the physically determined component of  $A$ ; if they did, the causal efficacy of nonmaterial mind would not require an amplification device as complex as the human central nervous and locomotor systems with their various support systems.

As for theoretical derivations of the dynamical laws for the vector potential, they tell us no more than what was initially assumed. Let  $\langle A_f | A_i \rangle$  be the amplitude for the transition from an initial field configuration  $A_i$  at time  $t_i$  to a final field configuration  $A_f$  at time  $t_f$ . The entire dynamics of  $A$  is expressible in terms of amplitudes of this sort. As there were contributions to the transition amplitude  $\langle y | x \rangle$  from all timelike curves connecting  $x$  and  $y$ , so there are contributions to  $\langle A_f | A_i \rangle$  from all “histories” of the field  $A$ , that is, from all parametrized curves in the infinite-dimensional space of field configurations. However, to ensure the existence of a classical limit (or, what comes to the same, of an extremal field history), a new term needs to be added to the mechanical line element  $d\phi$  when calculating field transition amplitudes. As it turns out, this term is fully specified by the requirements of local Lorentz invariance, additivity of  $\phi$ , and gauge invariance of  $A$ , if in addition  $A$  satisfies the principle of superposition. None of these assumptions give reason to pause.

But there is a crucial difference. To calculate the amplitudes  $\langle y | x \rangle$  for our open one-particle system, we had but to assume that the (externally prescribed) vector potential was capable of representing all possible nongravitational effects on an isoscalar particle, and this was an analytic truth. On the other hand, in order to calculate the amplitudes  $\langle A_f | A_i \rangle$  using the formalism just outlined, one must make sure (1) that all sources contributing to  $A$  are explicitly included in  $d\phi$ , and (2) that all effects represented by  $A$  are linked to their causes according to immutable mathematical laws; only then can these causal links follow from a Lagrangian. In order to be able to derive Maxwell's equations (as well as their quantum mechanical counterparts), these two conditions have to be met.

That the first condition is or can be satisfied will become clear in Sec. 7. The second condition, however, cannot be satisfied by the action of the nonmaterial mind on matter, as the following

section will show. The argument boils down to this: if, as interactionists believe, our sense of free will is no mere illusion, then the link between mental events and their physical effects cannot be described by mathematical laws.

## 6. ENERGY CONSERVATION AND THE INTERACTIONIST HYPOTHESIS REVISITED

Not only do metric inhomogeneities in space and/or time lend themselves to formulating effects on the motion of particles, they are also our only means of mathematically representing such effects. (We decided to disregard pair events and the forces requiring internal symmetries.) There are the inhomogeneities equivalent to Riemannian curvature, which are "experienced" by all particles; and there are the inhomogeneities represented by the electromagnetic vector potential, to which electrically charged particles and particles with a magnetic moment are susceptible. Both types of inhomogeneity can be treated as modifications of the mechanical lengths associated with paths in space-time, that is, modifications of the homogeneity of space-time as "experienced" by particles.

Noninteracting particles "experience" the unmodified homogeneity of space-time, that is, their behavior reflects or manifests this homogeneity. When electromagnetic interactions are added, particles both create and "experience" a type of inhomogeneity that can alternatively be thought of as a physical field. So conceived, it is a physical system in its own right that carries energy and momentum, mediates interactions between particles, and together with the particles exists in, or "experiences," a metrically homogeneous space-time. This is why in closed systems of electromagnetically interacting particles energy and momentum are conserved. When gravitational inhomogeneities enter the picture, the possibility of reinterpreting them as a physical field in a metrically homogeneous space-time exists only as an approximation, in weak-field situations with trivial topologies. Generic space-time manifolds are not metrically homogeneous, and this is why in generic situations the gravitational (and hence also the total) energy and momentum are ill-defined, as has been shown.

Open systems of electromagnetically interacting particles "experience" inhomogeneities, which is why in open such systems energy and/or momentum is not conserved. What we now want to determine is whether a causally efficacious nonmaterial mind can be part of a closed system, and if not, whether or not it makes sense to attribute (nonconserved) energy to the open system with which this mind interacts.

The following could be advanced in favor of treating the nonmaterial mind as if it were a component of a closed physical system. Think of the vector potential as a physical field in a metrically homogeneous space-time. Combine the field and all its sources into a closed system. Include mind among the sources, and the behavior of the system will remain invariant under space-time translations on condition that mind (or at least its physical effects) co-translates with the rest of the system. Since minds are associated with bodies, it is difficult to see how this condition could not be fulfilled. Nothing therefore seems to stand in the way of the conception of a closed system containing

not only the mental modifications of the vector potential but also the mind (or minds) causing them. In such a system energy and momentum would be exchanged between matter (particles), the vector potential or its derivative, the electromagnetic field, and the mind.

The notion that matter and mind exchange energy is vacuous all the same. For one thing, a principal objective of interactionism is to be able to treat volition as an independent causal agent, independent at least in the sense of not being determined by any combination of physiological microstructure and physical law. For such a causal agent one cannot write down a Lagrangian. (If one could, this causal agent would be just another kind of matter subject to just another kind of physical law, something whose existence neither dualists nor materialists would endorse.) But where there is no Lagrangian, there is no mathematical expression for energy, and in this case the necessary condition for attributing energy to mind is not satisfied. It is only if the inhomogeneities caused by mind and "experienced" by matter are quantifiable, that is, if they can be given mathematical expression, that a formula for the energy exchanged between matter and mind can be derived, and that the energy associated with mind can be defined (necessarily in such a way that the total energy is conserved).

For another thing, if there is anything that justifies the appellation "physical system," it is existence in space and time (at least in the sense of subjection to the experiential conditions of spatiality and temporality). But the nonmaterial mind, though subject to temporal succession, does not exist in space. It has no position, not even an indefinite one. Only the physical effects it is, *ex hypothesi*, capable of producing, are localizable in space.<sup>16</sup> Such an entity cannot legitimately be conceived as part of a *physical* system. (If the mind were viewed as a nonmaterial property of physical substance rather than as a substance in its own right, it could not even be thought of as a *part* of a physical system.)

Furthermore, energy can be attributed only to entities existing in time, while momentum can be attributed only to entities existing in space. In a relativistic world, energy cannot be attributed without at the same time attributing momentum, that is, it too can be attributed only to entities existing in space. Hence neither energy nor momentum can be attributed to mind.

Again, in a world in which simultaneity is relative, nongravitational energy and momentum must be *locally* conserved; they cannot disappear in one place and immediately reappear in another, far-away place. They can propagate from one place to another only by passing through whatever boundary lies between the two places. Hence they are in space, and so must be the objects to which they are attributed (considering that objects are where their localizable properties are). This, too, disqualifies the mind as a carrier of energy or momentum. It is matter that carries energy and momentum, and if that is open to the nonmaterial mind, its energy and/or its momentum are not conserved.

But even when matter is causally open to the nonmaterial mind and its energy and momentum are not conserved, these remain as well-defined as they are in the absence of the mind.

This is because there exists an experimental realm in which the inhomogeneities due to mind are absent or negligible.<sup>17</sup> The necessary conservation of energy and momentum in this realm makes it possible to verify the correctness of the mathematical expressions for the energy and the momentum of matter and of the Lagrangian by which these are implied. The same expressions may then be used for the energy and the momentum associated with matter where matter is open to the action of the nonmaterial mind.<sup>18</sup>

## 7. INTERACTIONISM AND PANPSYCHISM

The modifications of the electromagnetic field "experienced" by certain constituents of the body could be effected in either of two ways: the self could contribute to the electromagnetic field as a separate nonmaterial source or it could modify the way in which the field is built up by material sources. However, to act as a separate source of the field, the nonmaterial self would have to exist in space, a notion that has just been rejected. This leads to the conclusion that material particles are the only sources of the electromagnetic field, but that the self, under certain circumstances, influences the action by particles on particles that the electromagnetic field represents.

In other words, the causal efficacy of the self rests on the causal efficacy of the particles, that is, on the ability of the particles to modify their individual contributions to the electromagnetic field. The self can contribute to this summary representation of effects on the motion of charged particles only by modifying contributions from charged particles. The causal behavior of particles (meaning the way particles influence each other's motion, as distinct from the way particles move) thus comes in two modes: a physical mode which obeys the laws of physics and a nonphysical mode in which modifications of the physical mode are effected. Being the determinant of this nonphysical mode, the self relates to the individual particle in much the same way as it relates to the entire conscious person: just as it causes voluntary modifications of the latter's autonomous behavior, so it causes modifications of the causal behavior of certain particles.

But it is the self's causal efficacy with respect to the causal behavior of some of the body's constituent particles that explains its causal efficacy with respect to the body as a whole. Its ability to initiate actions such as raising a hand rests on its ability to modify the way in which those particles influence each other's motion. The self therefore is more directly associated with certain particles than it is with the body as a whole.

It now takes but a couple of more or less straightforward steps to go beyond dualism. The implications of interactionism elaborated in previous sections have an obvious extension in the direction of panpsychism, as the following will show.

First extend to all of the body's constituent particles the self's ability to modify the causal behavior of certain particles. Each particle of the body then is capable in principle of being affected in its causal behavior by the self, even though actual modifications of a particle's causal behavior may occur only in specific detector organs that assure causal efficacy on a macroscopic scale. (It could be an important function of the brain to amplify

the weak modifications of the causal behavior of particles which the self is able to generate and to so enable the self to cause macroscopic effects such as flexing an index finger. Modifications of the causal behavior of particles outside the brain are not ruled out, but they are not positioned to contribute to the self's causal efficacy with respect to the body's macroscopic behavior.)

Next extend to all particles in the universe the capacity for being affected in their causal behavior by a conscious self. Each particle in the universe possesses this capacity, but only such aggregates of particles as are equipped with the appropriate detector organs (brains) are capable of manifesting, or evincing the presence of, a causally efficacious self.

If consciousness is as causally efficacious as it is because of its ability to affect the causal behavior of particles, then the logical next step is to take the view that consciousness is rooted in the particles. It is because consciousness is already present in the particles that it can arise in suitably equipped aggregates. The existence of special organic structures (such as brains) may be necessary for the causal efficacy of consciousness (and presumably also for its awareness of the physical world), but it is not sufficient<sup>19</sup> for the existence of conscious states and the subjective point of view whence these are experienced. The emergence of consciousness with its irreducible first-person ontology is made possible by the fact that every particle already possesses some sort of self or consciousness, though presumably one without structured content.

How shall we conceive of the relation between the selves  $\{S_p\}_b$  associated with the constituent particles of a particular body and the self  $S_b$  that is both one (barring split-personality pathologies) and causally efficacious with respect to this body as a whole? Clearly,  $S_b$  cannot be simply the set  $\{S_p\}_b$ . It is not the *combined* consciousness of the particles. Nor can  $S_b$  be merely of the same *kind* as an  $S_p$ . If the primordial, content-free consciousness of the particles is to account for the existence of the unified, content-filled consciousness of the body, it must *constitute* what is essential to the consciousness of the body, that is, its subjective, first-person quality. So we must find a manner of constitution that does not work by simple combination.

To this end it might help to recall the common sense-defying behavior of particles of the same type. Consider an elastic collision between two identical bosons, one coming from the north, the other coming from the south. Suppose that after the collision one boson travels to the east and another travels to the west. Let's call these respective particles  $N$ ,  $S$ ,  $E$ ,  $W$ . Then it is not the case that either  $N$  is the same as  $W$  and  $S$  is the same as  $E$  or  $N$  is the same as  $E$  and  $S$  is the same as  $W$ . If this were the case, the possibility that we presently consider (namely, that one of the outgoing particles is eastbound and the other westbound) would be realized with probability

$$P_c(EW) = |\langle EW | NS \rangle|^2 + |\langle WE | NS \rangle|^2, \quad (5)$$

whereas in actual, quantum mechanical fact it is realized with probability

$$P(EW) = |\langle EW | NS \rangle + \langle WE | NS \rangle|^2. \quad (6)$$

$\langle EW|NS \rangle$  and  $\langle WE|NS \rangle$  are the respective probability amplitudes associated with the alternatives ( $N \rightarrow E, S \rightarrow W$ ) and ( $N \rightarrow W, S \rightarrow E$ ), in obvious notation. For bosons  $\langle EW|NS \rangle = +\langle WE|NS \rangle$ , whence it follows that  $P(EW)$  is twice as large as it would be if the incoming or outgoing particles had permanent identities:

$$P(EW) = |2\langle EW|NS \rangle|^2 = 4|\langle EW|NS \rangle|^2 = 2P_c(EW). \quad (7)$$

[In the case of fermions  $\langle EW|NS \rangle = -\langle WE|NS \rangle$  and  $P(EW) = 0$ .] The punch line is that a system consisting of two particles of the same type must not be thought of as a pair of intrinsically distinct entities. No conception that implies this kind of distinctness can be adequate. When the two particles are in different places, we can distinguish them by their positions, but when they are in the same place, as during a collision, all distinctness is lost. Only “twoness” persists; for instance, the system still has (approximately) twice the mass of each particle.

Here is how one can conceive of this “twoness” without implying a persistent distinctness: think of the two particles as two *manifestations* or two *instantiations* of identically the same entity. The manifestations are conceptually identical but numerically distinct. What is manifested twice is not only conceptually but also numerically identical. Distinguish between manifestations *when* they can be distinguished by their accidental properties, but do not think of that which manifests or instantiates itself twice as two things, and especially do not think that there are two things when there are no distinguishing accidental properties. When there are no such properties, there is “twoness” in the sense of a capacity for twofold manifestation or instantiation, but in absolutely no sense are there two things; whoever thinks so is immediately led to the wrong conclusion, namely, to expression (5) instead of (6).

In all particles of the same type there is something that is *numerically* identical. Particles of the same type are so many times one and the same “Particle.” And this “Particle” is not merely the same Platonic Form which is present or instantiates itself in materially distinct individuals, for there are no *materially* distinct individuals<sup>20</sup>; there are only *accidentally* distinct individuals. Abolish the accidental distinctness (as happens in many actual physical situations), and all there is is one entity that has “*n*-ness,” or the capacity for *n*-fold manifestation.

Needless to say, the manner in which this is “in” all particles of the same type is not that of spatial inclusion. It becomes necessary to conceive of another “dimension,” another extension or differentiation of reality besides the temporal and spatial ones, and this may be thought of as a plurality of ontic levels. At present we have two such levels: one at which particles of the same type are distinct individuals and another at which they are one and the same entity. (Without the additional distinction between ontic levels it would be self-contradictory to say that what is many is one. What is many at one ontic level can be one only at a different ontic level.)

Particles have other accidental properties besides their (more or less definite) positions. All changeable properties are accidental. A particle’s being of a certain type (for instance, its being

an electron or a charmed quark) is therefore necessarily an accidental property if an interaction exists that is capable of converting the particle into one of a different type, as the weak interaction is capable of converting electrons into electron-neutrinos and charmed quarks into strange ones. Whether or not such an interaction exists depends on the theory. In the Standard Model, hadrons and leptons cannot be converted into each other, nor can bosons and fermions, while in grand unified theories hadrons and leptons are mutually convertible, and in supersymmetric theories “once a fermion, always a fermion” is no longer true either. In these theories, being a hadron, lepton, boson, or fermion has become an accidental property of something that by itself is neither hadron nor lepton nor boson nor fermion.

But regardless of whether or not the ultimate theory will permit every conceivable type conversion, being of a certain type can be thought of as an accidental. Nothing therefore stands in the way of the notion that there exists but one fundamental species of particle, and that what has just been affirmed of particles of the same type is in fact true of all particles.

We are now armed with a conceptual tool for dealing with the emergence of a unified, experiencing, and causally efficacious self  $S_b$  out of a multitude of not manifestly conscious particles. “Internally” all fundamental particles are identical in the radical sense of strict, numerical identity. Something “in there” is the same self-identical being, and the familiar mental self  $S_b$  is one of the forms under which this manifests itself. There is a basic ontic level  $L_1$  at which the particles in a material aggregate  $b$  are numerically different entities; and there is an emergent ontic level  $L_2$  at which they are identically (in the sense of numerical identity) the same entity.  $S_b$  inheres in this entity.

At  $L_2$  the self-identical being in the constituent particles of one aggregate remains distinct from the self-identical being in the constituent particles of another aggregate (provided that the two aggregates have no common constituents). This is why there are many unified, experiencing and causally efficient selves  $S_b$  instead of just one. But then we may also conceive of something that transcends the multiplicity of all particles in the universe, something that is identically the same in all particles and, *a fortiori*, in all level-2 entities (which, individually and by definition, transcend the multiplicity of a particular aggregate but not the multiplicity of mentally equipped aggregates). In other words, we may conceive of a third ontic level  $L_3$  at which there is but one entity which is constitutively related to the level-2 entities as each of these is to a particular set of level-1 entities. (The locution “constitutively related” is noncommittal where the direction of the relation is concerned. But once we conceive of a single, undivided substratum for the multiplicity of the universe, it may be more natural to consider the higher-level entity or entities as constituting the lower-level ones.)

The facts that “internally” (i.e., at  $L_2$ ) the particles that make up a mentally endowed aggregate  $b$  are a single entity  $E_b$  and that  $S_b$  inheres in  $E_b$  may help explain how it is possible for  $S_b$  to modify the interactions between those particles. But several further questions arise. How is it that a subject  $S_b$  inheres in  $E_b$ ? What brings about the multiple instantiation of  $E_b$  to which the

particles that "make up"  $b$  owe their existence? What accounts for the causal behavior of  $b$ 's constituent particles (i.e., what enables them to affect each other's motion)? And what is it that has put and keeps in force (to the extent that it does) the laws according to which they affect each other's motion? Finally, supposing that there is an answer to the previous question, what kind of relation must exist between that causal agent and  $E_b$  so that  $E_b$  or its inherent self is able to modify those laws inside  $b$ ?

It is clear that questions of this kind will always lead to further questions. The name of the game is to find a patently irreducible set of postulates that can account for everything but themselves. Let's try to put together such a set.

**Postulate I:** At bottom being and consciousness (object and subject, matter and mind) are the same. Hence for every true fundamental statement about being there is a corresponding true fundamental statement about consciousness.<sup>21</sup>

**Postulate II:** At bottom there is nothing but one intrinsically unstructured or indeterminate being with the power to structure or determine itself. In other words, at bottom there is nothing but one intrinsically unstructured or content-free consciousness with the power to give itself structure or content.

In a fundamental way, the content of consciousness is just that. There is no consciousness *of being*. There is no being *over and above* consciousness. When we use the language of consciousness, we may talk about consciousness, the content of consciousness, and the power to create it. We may instead use the language of being, and then we may talk about being, the determinations of being, and the power to create them. But we must not mix languages. There are no relations between being and consciousness that would warrant statements referring to both being and consciousness. The only warranted statement referring to both of them is the affirmation of their strict identity. Being identical, they cannot play different roles in the same ontology, as this would suffice to render them distinct; they can only play corresponding roles in different formulations of the same ontology. We may say either that being has determinations or that consciousness has content because the two statements mean the same. But we cannot pick and choose: we cannot appeal to the identity of the determinations of being with the content of consciousness and, disregarding the identity of being with consciousness, conclude that consciousness is of (the determinations of) being, for to say so is to ascribe to them the distinct roles of subject and object. Consciousness is not *of* being; it *is* being.

By the same measure, there is no consciousness *over and above* being. Being is not something that *has* consciousness; being *is* consciousness. Being does not need to have consciousness in order to be aware of its determinations if saying that being has such and such properties is exactly the same as saying that consciousness has such and such content. Consciousness is no more a property of being than being is an object of consciousness.

Next we will postulate the obvious.

**Postulate III:** The determinations of being (the contents of consciousness) are spatial (i.e., they are shapes and/or spatial relations).<sup>22</sup>

Now for a crucial postulate:

**Postulate IV:** The power inherent in being (consciousness) not only enables being (consciousness) to give itself determinations (content) but also supports a multiple instantiation of being (consciousness) within the same ontology.

(It will be convenient to henceforth distinguish between the being/consciousness that is instantiating itself and the instantiations of being/consciousness by capitalizing the former.) If the many beings (consciousnesses) are members of the same ontology, their spatial determinations (contents) are spatial with respect to one and the same space; that is, their determinations (contents) must exist in the same extension as the determinations of Being or the content of Consciousness. As a consequence, the isomorphism that exists between the determinations of Being and the content of Consciousness does not exist between the determinations of an individual being and the content of an individual consciousness. Since each consciousness contains the spatial determinations not of one but of many beings, the spatial structure that the individual being possesses as a property cannot be the same as the spatial structure that the individual consciousness has for its content.

This casts the individual instantiation of Being = Consciousness into a double role. Intrinsically this is still inseparably and indistinguishably both a being and a consciousness. It is not a being that has something else, consciousness, as a property. What we call a being is the individual instantiation of Being = Consciousness in the role of the substrate of its determinations, the object "behind" the individual form. What we call a consciousness is the same instantiation in the role of the consciousness which has spatial structures for its content, the subject "in front of" the forms.<sup>23</sup>

Two final postulates will take us to the phenomenology of inanimate matter.

**Postulate V:** The individual instantiations of Being = Consciousness lack spatial determinations (i.e., they are formless and without internal spatial relations or internal structure).<sup>24</sup>

This makes it possible to think of the fundamental particles of matter as the products of a multiple instantiation of the original, *indeterminate* Being = Consciousness. Fundamental particles are generally thought of as having a pointlike form rather than no form at all. It therefore needs to be emphasized that the possible empirical evidence for the fundamental nature of a particle species (namely, the lack of evidence of internal structure) is consistent with either view. But the more parsimonious view that the spatial forms of things resolve themselves into sets of (more or less indefinite) relative positions has an obvious edge over the view that, in addition to the forms that can be accounted for in this manner, there exists another (empirically unwarranted) type

of spatial form: the pointlike form of a fundamental particle.

If ultimately all spatial determinations resolve themselves into relative positions, the creative power inherent in Being is reduced to a force that acts between particles and that statically determines (as in an atom in a stationary state) as well as dynamically changes their relative positions; in other words, a force (or forces) by which particles affect each other's motion.

**Postulate VI:** The force (or forces) by which fundamental particles affect each other's motion is (are) subject to immutable mathematical laws.

By giving the causal efficacy of Being = Consciousness the aspect of a mechanical force, consciousness has been successfully concealed. But consciousness does not stay concealed. In every suitably organized aggregate of particles, an intermediate (level-2) instantiation *E* of Being = Consciousness evinces (1) the power to affect its spatial form, (2) consciousness, and (3) the causal efficacy of consciousness (volition).

Given our basic postulates, it is not hard to understand how anything material can have freedom and consciousness and how consciousness can be causally efficacious. We don't have to worry about how there can be such things as consciousness and freedom because they are part of an irreducible set of ontological categories (postulated, respectively, in I and II). Nor need it puzzle us how anything material can be in possession of these things because the particles are instantiations of Being = Consciousness, while their interactions are the power inherent in this at work in a world of particles.

The strictly material is characterized by the adherence of this power to seemingly inflexible laws. Freedom and consciousness can manifest themselves by a partial reversal of the multiple instantiation of Being = Consciousness to which the particles owe their existence.<sup>25</sup> Unlike the individual fundamental particle, the level-2 individual *E* has a form (i.e., it is related to the form that consists of the aggregate's internal relative positions as an intrinsically indeterminate substrate is to its extrinsic determinations). So it can have a genuine if limited freedom to modify this form (necessarily, via modifications of the aggregate's physically determined internal interactions). Nor need *E*'s relation to form be limited to the relation of a substrate and/or determinant to its determinations. Given the essential identity of being and consciousness, it can just as well be related to forms as consciousness is to its content.<sup>26</sup>

## 8. SUMMARY AND CONCLUSION

The following results have been obtained:

- (1) Conservation of energy and momentum is a consequence of the metric homogeneity of time and space, and this is warranted for systems that are causally closed. That the energy and/or momentum of a physical system that is open to causal influences from the nonmaterial mind is/are not conserved is a matter of course.
- (2) Assuming that part but not all of matter is causally open to the nonmaterial mind, it remains justifiable to attribute energy and momentum to physical systems interacting with

the nonmaterial mind.

- (3) The interactionist hypothesis implies departures from the statistical laws of quantum physics, independently of whether or not the conservation of energy and momentum is an analytic truth.
- (4) These departures are capable of being reformulated in terms of modifications, by the self, of the electromagnetic interactions between particles; and they are more consistently formulated in this manner.
- (5) The electromagnetic vector potential being a summary representation of effects on the motion of particles, nothing stands in the way of the assumption that the effects represented by it are caused in part by mental events.
- (6) If freedom of will is no illusion, the causal behavior of the self cannot be subject to mathematical laws. Consequently, the interactions between particles that are modified by a self cannot take place in strict accordance with Maxwell's laws, or with any other physical law for that matter.
- (7) The difference between the material and the mental has been reduced to a distinction between two modes (out of three) of causal behavior of a single causal agent. As the self-identical being in all particles this originates the physically determined forces and sustains both these and their mentally determined modifications. As a plurality of causally efficacious conscious selves it is responsible for the actual form taken by these modifications. And as the multitude of particles it is what exerts those forces and thereby constitutes matter.<sup>27</sup>

Although there are no compelling theoretical or experimental reasons why mental events should not be capable of causing departures from physical laws, it may remain difficult for the interactionist, at least for some time to come, to disabuse the contemporary physicist, biologist, or philosopher of science of the doctrine of physicalism, which has been a reigning orthodoxy for well over a century. So much was this doctrine taken for granted that until recently it has been considered as almost indecorous to waste much thought over the dismissal of its antithesis. Thus, after stating that "very few people"<sup>19</sup> any longer suppose that living things violate any laws of physics (as some thinkers supposed as late as the 19th century), Putnam<sup>20</sup> makes known why this should be so: "Physics can, in principle, predict the probability with which a human body will follow any given trajectory." Are we to suppose that the mountaineer who fell to death would have been able to choose a less ruinous trajectory if only Eccles' hypothesis of mind-brain interaction had been true?

What the interactionist claims, in effect, is that the nonmaterial self becomes materially effective by modifying the electromagnetic interactions between certain constituents of the body. Not only is this consistent with the assumption that the trajectory of the body's center of mass is fully determined by physical laws, but it also agrees with our sense of free will which the interactionist wishes to take seriously. I decide to raise my hand and up it goes, but nothing in my experience leads me to expect that I could alter my trajectory once I have jumped off a cliff.<sup>28</sup>

Yet there is cause for optimism. As far as philosophers are concerned, "consciousness" has, over the last decade, made the transition from a virtual taboo word to one of the hottest items on their agenda. Many people in cognitive studies now admit the irreducibility of consciousness.<sup>(21)</sup> And most of the philosophers who speculate about the shape of a fundamental theory of consciousness invoke some form of panpsychism.<sup>(22)</sup>

However, most of these philosophers still find it necessary to reduce conscious events to "causal danglers": they affirm that pain is not reducible to its physical correlate yet deny that it causes us to pull our hands out of fires. Such a position is inher-

ently unstable, as Lowe<sup>(23)</sup> has pointed out; it is under intense pressure either to lapse back into materialism (which restores the causal efficacy of conscious feelings by identifying them with their physical correlates) or to take the further step of admitting the causal efficacy of consciousness. The present essay has shown that this further step can be taken by combining the best of interactionism and panpsychism (while avoiding their shortcomings<sup>(24)</sup>) into an integrated theory of consciousness and view of nature.

Received 12 December 1996.

### Résumé

*Soit la loi de conservation de l'énergie est impliquée dans le concept même d'énergie et par conséquent n'est pas remise en question par l'interactionisme psychophysique, soit elle est la conséquence de la fermeture causale du monde physique que l'interactionisme nie. Dans les deux cas l'interactionisme implique des dérogations aux lois de la physique en dépit des tentatives pour démontrer le contraire en s'appuyant sur l'indéterminisme de la mécanique quantique. Ces déviations sont le mieux formulées en termes de modifications, par le moi conscient, des interactions électromagnétiques entre particules. Le potentiel vectoriel électromagnétique est essentiellement une représentation sommaire des tous effets, quelles qu'en soient les causes, sur le mouvement des particules; ces effets peuvent être produits par des événements mentaux aussi bien que physiques. L'hypothèse de l'interactionisme implique que les modifications produites par le moi ne peuvent être décrites par des lois mathématiques. Par suite, les interactions entre particules qui sont modifiées par le moi ne peuvent se produire en conformité absolue avec aucune loi de la physique. Finalement, l'interactionisme ne se limite pas au dualisme; il peut s'intégrer à un panpsychisme selon lequel le moi et les particules constituant le corps sont des aspects différents d'un même agent causal.*

### Endnotes

<sup>1</sup> What is meant by this throughout this essay is the 4-potential or 4-vector potential.

<sup>2</sup> To be taken with a grain of salt: while it is a classical potential rather than a quantum mechanical probability that gets modified, quantum phenomena like tunneling are likely to remain essential to the causal efficacy of the nonmaterial mind.

<sup>3</sup> In the present study, the legitimacy of attributing type I changes to processes that are not in any obvious way measurements will be left unchallenged.

<sup>4</sup> The microtubule hypothesis adopted by R. Penrose, *J. Consciousness Stud.* 1, 241 (1994), is a membrane-physiological proposal for this trigger mechanism, as F. Beck, *J. Consciousness Stud.* 1, 253 (1994), has pointed out. It realizes the motion of the quasiparticle as the motion of one, or a few, hydrogen atoms in the membrane.

<sup>5</sup> In present-day physics the requirements of local interactions and local invariance carry even more weight than the local conservation laws derived from them.

<sup>6</sup> The precise connection between homogeneity and flatness will be pointed out in Sec. 5.

<sup>7</sup> It may be borne in mind that quantum mechanics in its

standard interpretation is not consistent with the instrumentalist's identification of probabilities with relative frequencies. Quantum probabilities reflect objective indeterminacies associated with individual events and not merely our ignorance about them. It is therefore legitimate to discuss single-case probabilities. See also K.R. Popper, *The Open Universe* (Hutchinson, London, 1982), p. 95, for his insistence that likelihood is more fundamental than frequency. *A* happens more often because it is more likely, and not *A* is more likely because it happens more often.

<sup>8</sup> The same violation is entailed when the nonmaterial self is replaced by a "superintelligence" who, as F. Hoyle, *The Intelligent Universe* (Michael Joseph, London, 1983), surmises, guides the evolution of the cosmos by altering the probabilities associated with quantum processes.

<sup>9</sup> This explains why the *mechanical* equivalence of locations in space(-time) referred to as homogeneity is tantamount to the *metric* equivalence of locations in space(-time) referred to as flatness.

<sup>10</sup> "Effectively" because one must not think of one curve being transformed into another curve. What happens (in general) when the dynamical lengths of space-time curves are changed in a manner consistent with flatness is that previ-

ously extremal curves cease to be extremal while other curves become extremal.

<sup>11</sup> Further reasons for dismissing gravity as a possible vehicle of mental causation are given in Endnote 14.

<sup>12</sup> If the additional term were nonlinear in the differentials, the Lagrange equation would contain another term proportional to the acceleration besides the term containing  $m$ .

<sup>13</sup> For completeness' sake, another way in which the mind could be causally efficacious ought to be mentioned. According to E.P. Wigner, in I.J. Good, *The Scientist Speculates* (Heinemann, London, 1961; NY, 1962), p. 284, J. von Neumann,<sup>(6)</sup> and H.P. Stapp, *Mind, Matter and Quantum Mechanics* (Springer, Berlin, 1993); J. Consciousness Stud. 3, 194 (1996), the collapse of a quantum mechanical superposition to one of its components is precipitated by consciousness. Even if this were true, it would be irrelevant to the causal efficacy that interactionists have in mind. Consciousness would play the same role as the measurement apparatus in the orthodox Copenhagen interpretation; it would select the set of alternatives with respect to which collapse takes place (a basis in Hilbert space) without in any way affecting the physically determined probabilities associated with alternatives. The particular outcome of a type I change would be "determined" by pure chance, while on the interactionist hypothesis the physical correlate of free will is anything but pure chance.

<sup>14</sup> Now that we know that the second manner of modifying the mechanical lengths of space-time curves is, in actual fact, the way of the electromagnetic force, we have another reason for dismissing gravity as irrelevant to mental causation. Since exocytosis is controlled by the influx of  $\text{Ca}^{2+}$  ions into a synaptic vesicle (Ref. 9, p. 149), mental causation is likely to be effected through a modification of the physically determined force exerted on such ions or, at any rate, on charges, especially those involved in the propagation of nerve impulses. But the electromagnetic interaction between, say, two protons is about  $10^{36}$  times stronger than their gravitational interaction. Hence if the mentally generated modification of the force exerted on a charged particle were of gravitational nature, the mind would have to generate an implausibly strong gravitational field (about that many times stronger than the physically generated one), while it would only need to generate an electromagnetic field that is weak in comparison with the physically generated one.

Yet another reason why the electromagnetic interaction is the more likely vehicle of mental causation is the selectivity of the electromagnetic force. While this acts on charged particles only, gravity affects everything. If one wants to accelerate a charged particle in a neutral medium, one had better not also accelerate the medium, since this would simply produce a congestion.

All said, nothing stands in the way of assuming that the mind contributes to any or all of the four fundamental interactions, inasmuch as the gauge fields of the weak and strong forces no less than the metric connection and the

electromagnetic vector potential are simply ways of formulating possible effects on the behavior of particles, whether their origin be physical or mental or whatnot. But since we have reasons to assume that the most effective vehicle of mental causation is the electromagnetic field, the other possibilities are not considered here.

<sup>15</sup> So far, I have made no mention of the physical irrelevance of the chosen gauge. It goes without saying that to be physically efficacious, mental causes must contribute to the physically relevant aspect of the vector potential and hence to the electromagnetic field. This is why I do not consider it necessary to always strictly distinguish between the vector potential and the electromagnetic field.

<sup>16</sup> M. Lockwood, *Mind, Brain and the Quantum* (Basil Blackwell, Oxford, 1989), p. 101, takes special relativity to imply that mental states must be in space given that they are in time. This conclusion appears to rest on a too naive identification of two distinct concepts of time. What "time" means in the context of psychological experience is not the same as what it means in the context of special relativity. Without an in-depth study of their relation (not offered by Lockwood), only the physical effects of mental states can be said to necessarily exist in space-time. See C.J.S. Clarke, J. Consciousness Stud. 2, 231 (1995), for a refutation of Lockwood's arguments for the spatiotemporal localization of mental events.

<sup>17</sup> Given present experimental limitations, they may well be negligible everywhere.

<sup>18</sup> If there existed no experimental realm in which the inhomogeneities caused by mind are absent or negligible, attributing energy or momentum to matter would be gratuitous and otiose, because in this case any mathematical expression would do. None could be falsified on the basis of experimental data, for the corresponding conservation law needed for this purpose could nowhere be tested. For any expression, violations of its conservation could be imputed to mind.

<sup>19</sup> Some would deny even the necessity. According to A.A. Cochran, Found. Phys. 1, 235 (1971), the known facts of quantum physics and biology "strongly suggest" the related hypotheses that atoms and fundamental particles have a rudimentary degree of consciousness, volition, or self-activity and that the so-called indeterminism of quantum mechanics is a result of this fact. For different reasons the proponents of panexperientialism [C. de Quincey, J. Consciousness Stud. 1, 217 (1994)] hold that experience goes all the way down.

<sup>20</sup> In the Platonic view matter is what makes possible multiple instances of the same type and what is different in conceptually identical individuals.

<sup>21</sup> What exactly is meant by a fundamental statement about being/consciousness will be spelled out in Endnote 23.

<sup>22</sup> What about *temporal* determinations? If one thinks of the power postulated in Postulate II as the power by which being gives itself spatial and temporal determinations, one places the agent, being, not only "outside space" but also

"outside time." As the creator of spatial (temporal) determinations and thus of space (time), it cannot itself be spatially (temporally) differentiated. The temporally undifferentiated act of creation then has to be thought of not as situated in or at the beginning of time but as coextensive with it. The power inherent in being may well have this "atemporal" aspect, and this may well be responsible for the unchanging spatiotemporal regularities to which dynamical laws give expression. Mental causes, however, are themselves subject to time. Hence, if the power inherent in being is to account for their causal efficacy, it must also have this temporal aspect, and this can create spatial but not temporal determinations. (It of course also can *change* the spatial determinations.) Postulate III implicitly assumes an already temporally differentiated causal agent.

<sup>23</sup> The fundamental statements about being/consciousness are those that refer to Being/Consciousness.

<sup>24</sup> Lacking changeable determinations, they cannot be temporally differentiated, and this is the same as saying that they also lack temporal determinations.

<sup>25</sup> The *multiplicity* of constituent particles in a mentally equipped aggregate is not abolished by this reversal, although in a sense the multiplicity of *constituent* particles is: the manifestation of a level-2 individual *E* makes it possible to conceive of *E* as the aggregate's constituent substance and of the discrete multiplicity entailed by a form consisting of relative positions, as an extrinsic property of *E*.

<sup>26</sup> The essential identity of having a form and being conscious of it "survives" at *L*<sub>2</sub> as far as *E* and its own form are concerned. Through it the level-2 individual can be directly conscious of certain goings-on in its brain, and if these contain objective images of external objects (patterns of neuronal activity linked to visual input, for instance), it can construct from these an indirect, representational consciousness of forms other than its own.

<sup>27</sup> The fact that the multitude of externally interacting particles are, at another ontic level, a single self-interacting entity may also be needed to understand how "one body may act upon another at a distance through a vacuum, without the mediation of any thing else" which to Newton was "so great an absurdity" that he believed that "no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it." (Quoted in Ref. 13, p. 41.) Ultimately all action is, or at least is made possible by, Being's power of self-determination (or the power of Consciousness to give itself content). To think of action-at-a-distance as mediated by the transmission of, say, photons does not solve Newton's problem, for one would be hard-pressed to explain how the transmission of a photon is different from an elementary, irreducible action-at-a-distance.

<sup>28</sup> I (the consciousness inherent in what is identically the same in the constituents of my body) can affect the position of my hand relative to the rest of my body, because this only involves modifications of interactions that are internal to my body. I cannot alter my trajectory because this would involve modifications of interactions between my body and

other material objects. Such modifications could only be caused by what is identically the same in my body *and* those other objects.

<sup>29</sup> Such as (for interactionism, given its acceptance as a species of dualism) the difficulty, if not logical impossibility, to account for the causal efficacy of the mental without explicitly or implicitly uniting mind and matter in a single ontology (and thus going beyond dualism). Such as (for panpsychism) the propensity to grant sentience to room thermostats.

## References

1. M. Bunge, *The Mind-Body Problem* (Pergamon, Oxford, 1980), p. 17.
2. D.C. Dennett, *Consciousness Explained* (Allen Lane/Penguin, London, 1991), p. 35.
3. K.R. Popper and J.C. Eccles, *The Self and Its Brain* (Routledge and Kegan Paul, London, 1983), p. 564.
4. A. Eddington, *The Nature of the Physical World* (Dent, London, 1935).
5. H. Margenau, *The Miracle of Existence* (OxBow, Woodbridge, CT, 1984), p. 96.
6. J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955).
7. D. Hodgson, J. Consciousness Stud. 3, 69 (1996), likewise invokes the mind's ability to load the quantum dice.
8. F. Beck and J.C. Eccles, Proc. Natl. Acad. Sci. U.S.A. 89, 11357 (1992).
9. J.C. Eccles, *How the Self Controls Its Brain* (Springer, Berlin, 1994).
10. *Ibid.*, p. 78.
11. *Ibid.*, p. 170.
12. E.J. Squires, Eur. J. Phys. 2, 55 (1981).
13. See, for instance, C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (W.H. Freeman, San Francisco, CA, 1973), p. 451.
14. *Ibid.*, p. 435.
15. N. Bohr, Letter to Høffding, 22.9.1923, Bohr Scientific Correspondence, microfilm No. 3, p. 5; quoted in J. Honner, Stud. Hist. Phil. Sci. 13, 1 (1982).
16. G.C. Ghirardi, A. Rimini, and T. Weber, Phys. Rev. D 34, 470 (1986).
17. P. Pearle, Phys. Rev. A 39, 2277 (1989).
18. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, NY, 1965).
19. Among them are the biologists A. Szent-Györgyi, in *Light and Life*, edited by W.D. McElroy and B. Glass (Johns Hopkins, Baltimore, MD, 1961), W. Elsasser, *Atoms and Organism* (Princeton University Press, 1966), M. Delbrück, *Mind from Matter? An Essay on Evolutionary Epistemology* (Blackwell Scientific, Palo Alto, CA, 1986), and Mae-Wan Ho, *The Rainbow and the Worm: the Physics of Organisms* (World Scientific, Singapore, 1993). Their work supports the view that particles behave differently in living systems.
20. H. Putnam, *Renewing Philosophy* (Harvard University Press, Cambridge, MA, 1992), p. 83.

21. See the issues of *Journal of Consciousness Studies* devoted to the "hard question" about consciousness: 2 (3) (1995); 3 (1,3,4,5/6) (1996); 4 (1) (1997).
22. For explicit panpsychist proposals see, for example, P. Hut and R. Shepard, *J. Consciousness Stud.* 3, 313 (1996), G.H. Rosenberg, *J. Consciousness Stud.* 3, 76 (1996), and W. Seager, *J. Consciousness Stud.* 2, 272 (1995).
23. E.J. Lowe, *J. Consciousness Stud.* 2, 266 (1995).

**Ulrich Mohrhoff**  
Sri Aurobindo Ashram  
Pondicherry 605002, India