Quantum Leaps in Philosophy of Mind

David Bourget (www.dbourget.com)

Appeared in *Journal of Consciousness Studies* 11 (12): 17-42. 2004.

Abstract: The quantum mechanical theory of consciousness and freewill offered by Stapp (1993, 1995, 2000, 2004) is exposed and clarified. Decoherence based arguments against this view are undermined in an effort to draw attention to the real problems it faces: Stapp's separate accounts of consciousness and freewill are incompatible, the interpretations of QM they are tied to are questionable, the Zeno effect could not enable freewill as he suggests because weakness of will would then be ubiquitous, and the holism of measurement in QM is not a good explanation of the unity of consciousness for essentially the same reason that local interactions may seem incapable to account for it.

I: Introduction

Quantum theories of mind are routinely derided as having the explanatory power of 'pixie dust in the synapses.'[1] A few among the dozens which have appeared since the beginnings of quantum mechanics nevertheless gained considerable visibility in the last decade: those of Penrose and Hameroff (1996), Eccles (1994), Lockwood (1989), and Stapp (1993, 2001, 2004).[2] Of these, Stapp's theory is on the face of it the most interesting. It is supposed to (a) be immune to the standard criticism based on quantum decoherence which threatens the Penrose-Hameroff hypothesis; (b) avoid the dualism of Eccles's account; and (c) shed considerably more light than Lockwood's proposal on the central questions in philosophy of mind. The theory has two complementary parts: a quantum mechanical explanation of the unity of consciousness and an account of conscious will which frees it from the constraints of physical laws. It is still unclear whether freewill and consciousness can be accounted for in the world of classical physics (see for instance the arguments put forward in Kane [2002] and Shear [1997]). It is thus interesting to see what progress could be made by taking the quantum leap with Stapp. I offer a brief survey of the relevant aspects of quantum mechanics in §II.

[1] I take the expression from Rosenblaum and Kuttner (1999).

[2] References to the first edition of Stapp's *Mind, Brain and Quantum Mechanics* (1993) are kept separate from those made to the second edition of this work (2004) as this essay was written on the basis of the former and only subsequently updated to incorporate the new material which appears in the latter. The continuity (at least with respect to the points discussed here) between this new material and Stapp's previous writings is highlighted in several footnotes of the present essay.

I then explicate and clarify his approach in §III-V. In §VI-X I shed new light on the relevance of quantum decoherence, raise doubts concerning the viability of the kind of interpretation of quantum mechanics Stapp is committed to, and put forward serious technical challenges for his two main proposals.

II: The Measurement Problem

Stapp's views stem from considerations concerning the measurement problem in quantum mechanics (hereafter, 'QM'). I thus start by briefly rehearsing the relevant aspects of this problem.

At the heart of the postulates of quantum mechanics is a distinction between two kinds of dynamical processes that can govern the evolution of a system. These were clearly identified by von Neumann (1955) in what has become the orthodox formulation of the theory. The first process (hereafter, 'process 1') governs the evolution of a system subject to measurement. It is described by the so-called *projection postulate.* It is not deterministic in that the state a system evolves into under it (the outcome of the measurement) is not normally uniquely determined by the system's current state (also called its wave function), although probabilities are assigned to all possible outcomes of all possible measurements. Following Stapp (1996), I will sometimes construe measurements as 'empirical questions' one can put to nature. Basic empirical questions always take a yes-or-no form and must be formulated in terms of observable quantities, e.g. 'is the particle at position p?' The second process (hereafter, 'process 2') is defined by Schrödinger's equations, is deterministic, and specifies in what state a system will be at any point in time as a function of its initial state given that no measurement is effected on it. This quantum state determines the probabilities associated with all possible outcomes of process 1. When the state of a system gives non-null probabilities for more than one outcome of process 1 for an empirical question, we say that the system is in a superposition of states for the corresponding observable.[3] Such superpositions are often said to be 'reduced' or 'collapsed' by process 1, whereas they are preserved by process 2.

[3] A *definite* state that gives non-null probabilities for more than one outcome is not the same as that of a system which is in an unknown state. See §VI for an explanation of the relevant distinction between superpositions and unknown states ('mixtures').

The measurement problem arises from the fact that the postulates of quantum mechanics do not specify which process occurs when. Since process 1 and process 2 are not equivalent, its silence on the question makes the theory incomplete. There are three aspects to this incompleteness; they correspond to three specific questions the theory leaves unanswered:

(1) When does measurement occur? The theory does not say when a measurement occurs, that is, when it is appropriate, from the point of view of an experimenter, to apply the projection postulate instead of process 2 to make predictions.

(2) What measurement occurs? Even if the theory did tell us when process 1 applies, it would also have to tell us *what* measurement takes place, i.e. what empirical question is being asked. As is illustrated by the Zeno effect (see §IV below), the choice of an empirical question can make an important difference to the dynamics of a system.

(3) *Where does a given measurement occur?* Even if we knew what measurement occurs when, there could still remain a certain amount of indeterminacy concerning the exact physical events that must be modelled with process 1. For instance, a common assumption is that we know when a measurement occurs in the time of our conscious experience, as well as what we are measuring at that time. But, as von Neumann remarked, this does not tell us where process 1 instead of process 2 applies in the chain of physical events that ranges from the collision of a photon on a photoelectric plate to the triggering of the relevant neural events in the observer's brain.

Typically, only the third kind of indeterminacy is singled out as 'the measurement problem', but all three deserve serious attention.[4] Many solutions have been proposed to the measurement problem; far too many, in fact, to do justice to the relevant debates here. In order to properly inter-

[4] It is important to note that von Neumann's argument to the effect that the third kind of indeterminacy has merely 'metaphysical' implications is not conclusive. He is often said to have demonstrated that wherever we choose to locate process 1 along that chain does not matter to the resulting measurement. But his demonstration works only when observer and observed are left alone. As Hughes (1989) explains, if a second observer is brought in to make observations along the causal chain of measurement, the exact localization of process 1 makes an empirical difference, because superpositions along the chain must stop after process 1, and they can be detected because of the interference effects they can generate. Given the ubiquity of multi-observer systems, the third kind of indeterminacy is a problem even for a theory with no greater aim than that of being a complete set of rules for making empirical predictions.

pret Stapp, however, we must keep in mind a central point of divergence concerning the reality of process 1. Several interpretations or revisions of QM purport to yield the same empirical results as von Neumann's theory *without* the projection postulate. These include Everett-type interpretations (many-worlds, many-minds, and consistent histories), Bohm's deterministic version of QM, and several 'apparent collapse' accounts of measurement, which are supposed to reconstruct the empirical predictions of the projection postulate by taking into account factors hitherto neglected (decoherence, quantum-computing style interference in measurement apparatuses, thermodynamical effects, etc.; see Stamatescu, 1996). On the other side of this divide are interpretations which take process 1 to be irreducible to process 2. In addition to von Neumann's orthodox formulation, these include Wigner's earlier interpretation involving 'consciousness', Heisenberg's version of the Copenhagen interpretation, and several attempts to single out special conditions of collapse in physical terms (e.g. Pearle, 1990; Penrose and Hameroff 1996; Ghirardi *et al,* 1990). Finally, some interpretations are rather agnostic and gloss pragmatically over this issue in the fashion of Bohr and van Fraassen (1991).

Stapp (1993) claims to be following Heisenberg in (1) recognizing a minimal reality to superpositions of states, (2) taking the (actual) collapse of such states to constitute the real, concrete events of this world (which he calls 'actual events'), and (3) drawing a line based on scale between the quantum realm, where everything is mere potential, and the macroscopic realm, where objects appear in classical states.[5] He also rejects most other interpretations of QM on the basis of the arguments typically levelled against them. I must gloss over these debates for now, but I return in the critical part of this essay to the questions of how much of Stapp's theory hinges on a particular interpretation, of the nature of this interpretation, and of its plausibility (§VII).

III: Stapp's First Account of Freewill and the Neurobiological Basis for his Second

To my knowledge, Stapp never clearly situates his views with respect to the main positions on freewill. It is thus worth pausing to characterize his aims before dwelling into the details. He distin-

[5] By 'classical' I mean a state which has the minimal amount of superposition allowed by the uncertainty principle for typical observables (position, momentum, energy).

guishes two forms of determinism: internal and external. Any factor not described in phenomenal terms is external to one's mind and will (the kind of will that interests Stapp is always conscious; see his 1993, pp. 92-3). One's brain is thus on this account an external determinant of one's actions. Stapp is dissatisfied with theories of freewill formulated within the classical framework. He argues that on all such views external factors leave no room for one's conscious will to determine one's actions, making it 'superfluous,' in the sense that 'the evolution of the physical universe would be exactly the same' whether it existed or not (1993, p. 37; see also his 1996b). That is to say that he charges ordinary theories of consciousness and freewill of epiphenomenalism. What this amounts to is better explained in his 1999 account, in which he states that he is after a theory which frees the conscious will from the 'complete dominion of myopic (*i.e., microlocal*) causation and random chance,' a goal that will be achieved only if it is found not to be 'reducible to' or 'replaceable by' the microphysical (p. 160). Characterizing as a 'complete failure' the 'three-century-old effort to reconcile the properties of mind [one of which is freedom] with the concepts of classical physics', he proposes to switch to the quantum mechanical framework (1993, p. 38). Since the problem arises on Stapp's view from the premise that conscious states should be reducible[6] to physical states in the classical framework, his project should be to show that these internal factors could in the quantum framework play a role that eludes any such reduction.

Stapp appears to have developed two different accounts of the relevance of quantum mechanics to the question of freewill. Even if he silently switches from the first to the second in his most recent papers, a brief examination of the former is in order, since it provides the background for several aspects of his more recent work. As he admits, this first account is a 'more detailed form of Eccles's general idea' (1993, p. 105). I thus start with a summary of Eccles's proposal before pointing out where Stapp's account departs from it.

Eccles (1994) suggests that irreducible, non-physical mental events can cause brain events of macroscopic scale without violating conservation laws. The plausibility of his hypothesis rests on

[6] I use the notion of reduction in the epistemological sense that A is reducible to B iff A can with the help of bridge principles be replaced by B for the purpose of making predictions. The reducibility of the mental to the physical is, even in this sense, enough to compromise genuine freewill on Stapp's view (see his 1996b).

the claim that quantum-level physics are relevant to macroscopic brain dynamics, which he supports with findings concerning the process leading to the release of neurotransmitters from the extremity of an axon (exocytosis).

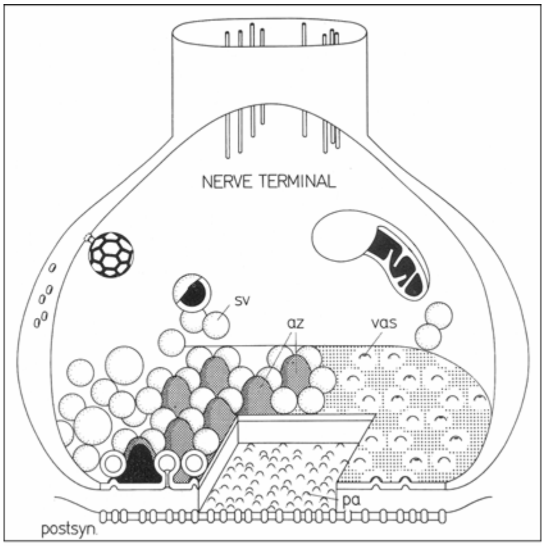

Fig. 1: The extremity of an axon. Reproduced with permission from Akert et al (1975).

Figure 1 illustrates his account of the process in question. A bouton (the extremity of an axon) contains vesicles filled with neurotransmitters (SV). At any time, a small number of these vesicles (between 30 and 50) are in contact with the membrane of the synaptic cleft, kept in place by a sort of grid (AZ, VAS). The process begins when a nerve impulse causes an influx of a large number of calcium ions into the bouton. Some of these ions then bind to a protein called calmodulin, after which they can trigger the release into the synaptic cleft of the neurotransmitters held in the vesicles. More recent work has put in question several aspects of this picture, but Eccles's general account of the role played by small calcium ions has not to my knowledge been disputed (see Wilson [1999] for a survey of the relevant literature).

An interesting finding with which Eccles supports his hypothesis is that at most one vesicle normally releases its neurotransmitters when a bouton receives an impulse. According to him, these observed statistics are surprising given the high concentration of calcium ions and calmodulin that can be expected in a bouton after an impulse reaches it. He thus conjectures that the mind exerts a fine control at the quantum level to tune the above probabilities by affecting the state of the pre-synaptic grid, a scenario which makes possible a strong form of freewill. Eccles also notes that the will must simultaneously influence large groups of boutons to significantly affect the dynamics of the whole brain. These large groups ('dendrons') are supposed to be associated with psychic counterparts ('psychons').

Eccles's account has the major flaw that it is incompatible with the principles of quantum mechanics: it requires either a violation of the projection postulate in the long run or a modification of the brain's wave function by some non-physical factor before collapse occurs (he at times seems to waver between the two). These violations of the postulates of quantum mechanics do not seem more acceptable than the blunt violation of the principles of conservation he wants to avoid (Mohrhoff 2002; Vaas, 2001). Also, the recent refutation of his hypothesis concerning the functioning of the presynaptic grid implies that the interventions of the will could not be hidden under the quantum fluctuations characterized by Heisenberg's uncertainty principle; they would require the violation of conservation laws (Wilson, 1999).

The least controversial aspect of Eccles's picture –the role of calcium ions in exocytosis— provides part of the empirical basis for Stapp's two accounts of freewill. Stapp (1993, 1997, 2000) maintains that the small scale of calcium ions entails by Heisenberg's uncertainty principle that they should be found in significant superpositions of position by the time they can trigger exocytosis. He concludes that 'the brain must evolve into an amorphous superposition of states corresponding to a continuum of different possible macroscopic behaviours.' (1997, p. 83)

More specifically, he (1993, p. 152-6) argues that the brain should evolve into such superpositions in a mere 200μs. According to Fogelson and Zucker (1985; cited by Stapp), it takes 200μs for a calcium ion to travel a distance of 50nm before reaching the release site. Stapp concludes that a bouton should be found in a superposition of firing and non-firing states within this timeframe. So far, however, there is no reason to think that the brain should be found in superpositions of macrostates in 200μs: the indeterminacies pertaining to the firing of all the boutons in the brain could average out (this is something Eccles takes for granted). But Stapp holds that this is not what happens. He makes the further hypothesis that the brain's non-linear dynamics can amplify some such microscopic fluctuations into large patterns. He postulates that the relevant neural mechanisms (e.g. those for decision-making) involve mutually exclusive, self-sustaining neural groups. In giving his ac-

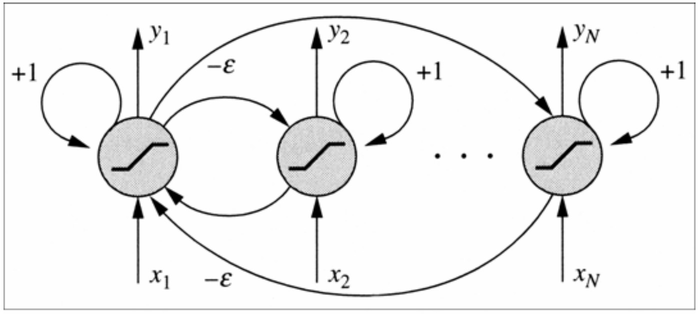

Fig. 2: A simple winner-take-all network. Excitatory connections are marked '+1' and inhibitory connections '-ε'. Xs are inputs and Ys outputs. Reproduced with permission from Principe et al (1999).

count he describes without mentioning it explicitly the winner-take-all network topology illustrated in Figure 2. Networks of this type can be used to extract the strongest of a range of signals: the group which gets the strongest initial activation rapidly remains the only one activated (see Principe et al*,* 1999, p. 335). Since a very small difference in the initial activation levels of two groups can result in a large discrepancy, Stapp estimates that even if a brain starts in a classical macrostate, it will evolve into a superposition of large activation patterns within 200μs, the time it takes for sufficiently many boutons to be in superpositions of firing and non-firing states.

Stapp first designated the will as the factor which determines in what state superpositions of macroscopic brain states generated by process 2 collapse into:

> The basic idea of the present psychophysical theory is to identify the *selection* of one of these mutually exclusive self-sustaining patterns of neural excitations as the image in the physical world, represented by quantum theory, of a creative act from the realm of human consciousness. (1993, p. 102; original emphasis)

This approach to freewill has the same flaw as Eccles's. Stapp does not suggest that the will has the power to alter process 2 in order to change the actualization probability of macroscopic brain patterns, but only that it can choose which one is actualized within the probabilistic constraints set by the wave function evolved through this deterministic process. If, as he insists (1997, p. 187), the postulates of QM should not be violated, this scenario has the absurd consequence that after having decided often to do X in a state which is a superposition of, say, X and Y, one would have to somehow be *forced* to do Y for no meaningful reason.[7] This probably explains why Stapp has silently switched to a different account in his 1999 and later papers.

IV: Stapp's Second Account

In Stapp (1999, 2000, 2001), freewill is introduced not through the stochastic character of process 1, but in the gap left open by the indeterminacies surrounding quantum measurement (see §II

[7] To be more precise, this would happen as the number of repetitions tends to infinity.

above).[8] As explained above, there are three questions left unanswered by current physical theory: what measurement occurs? when does it occur? and where? According to Stapp, it is the indeterminacy pertaining to the timing of measurement which leaves room for a causally efficacious freewill. The two other questions are answered as follows. He first argues that collapse is triggered in the brain of human observers, which settles the third question.[9] As for the first question, in his 1999 account he hints that the will can determine what measurement occurs, but he retracts this suggestion in his more recent writings. Instead, he holds that the empirical question is always determined by the wave function of the brain at any time. In principle, there is a discrete quantum mechanical observable corresponding to any volume V of a brain's state space. This observable has two values: yes or no. A measurement of this observable for some V yields a positive outcome if the brain is in V and negative otherwise. There should thus be a measurement that can tell us if a brain is in one of the many possible states in which any decision has been made. For instance, an empirical question which can in principle be asked in process 1 is, 'am I moving a finger?' Such a question verifies the implementation of a plan of action. According to Stapp (2000; 2004, ch. 12), the will's freedom at any point in time is limited to the choice of asking or not the question of this kind which is most likely to get a positive answer. At one point in time the question most likely to yield a positive outcome might be 'am I moving a finger?', at another time it might be 'am I stepping up?', and so on. All the will can do at any time is decide to ask or not the question which is independently determined according to this rule.[10]

It is thus only by choosing the rate at which process 1 takes place that the will can influence the brain. Stapp points out that the choice of the frequency at which process 1 occurs in a system can alter its dynamics. This peculiar aspect of QM has been dubbed 'Zeno Effect' by Misra and Sudarshan (1977), and it is reasonably well understood and verified (Joos, 1996; Itano *et al,* 1990). Take a measurement M that can have 'yes' and 'no' outcomes. In QM the likelihood that the outcome of M at $t_1 = t_0 + \Delta t$ be the same as the outcome of M at t_0 increases as M is performed with increased

[8] This approach is neglected by Wilson in his survey of the possible entry points for freewill in the physical world (published in this journal, 1999).

[9] Stapp has been inconsistent on this point. See §VII.1.

[10] Although the empirical questions considered here (and by Stapp) are always of the binary kind, it is clear that more complex choices could be reconstructed from these. What is said below about such binary choices applies to more complex ones.

frequency between t_0 and t_1. In the limit case of continuous measurement, the Zeno effect can prevent a system from evolving with respect to the degrees of freedom that are relevant to the outcome of M. Stapp wants to exploit the Zeno effect to give some control to the will over its associated brain. His approach is encapsulated in the following passage:

> If a person can, by wilful effort, acting through his power to consent, increase the rapidity of the [actual events] in his stream of consciousness, then he could control the activity of his brain by keeping [its] activity … confined to the subspace it is already in. The brain state would be prevented from 'wandering' in the way that it would if there were no [measurement process]. Thus wilful effort would alter the behaviour of the quantum brain, at the statistical level, from what it would be if there were no macroscopic quantum effects. (Stapp, 2000)

Stapp does not provide details concerning the 'wandering' in question in his latest writings, but he repeatedly alludes to his earlier account (§III).[11] On this picture, the brain should often be found in superpositions of macrostates, and these could presumably be superpositions of different decisions in a decision-making context. This is not yet to say, however, that the brain would 'wander' without the intervention of the will through the Zeno effect, if by this we mean that no course of action is chosen for good. This is because the different courses of actions would be taken in different superposed macrostates if collapse never occurred, and there is no sense in which the agent would wander in any of these 'parallel universes'. But process 1 must also take place in absence of conscious will, as Stapp's account of the unity of consciousness makes clear (see §IV below). Presumably, then, the idea is that the brain superposes significantly between each unwilled collapse, yielding an unstable behaviour which can be stabilized by increasing the rate of process 1 through a 'wilful effort'.

It is worth noting that the proposed model implies that a truly free will may have to wait for variable amounts of time for the brain to collapse into the desired state before increasing the rate of process 1 to prevent it from wandering away from the desired plan of action. Suppose, for instance,

[11] In particular, he recapitulates this account in his 2004 book (ch.12).

that I have to decide whether I will move a finger or not. The relevant empirical question would then be whether I am moving my finger or not, and my brain's potential oscillation between the two possible answers to this question would be the result of quantum indeterminacy plus sporadic measurements. There is no guarantee that the result of any of a finite series of such sporadic occurrences of process 1 will be the desired state. For instance, my brain may collapse onto the 'move finger' region of its state space any number of times in a row. If that is not the desired state, freewill has to wait. A truly free action may thus take any finite amount of time to implement.

Stapp does not make explicit this consequence of his model, but he seems committed to it by his criterion of freedom and his account of the macroscopic implications of quantum-level indeterminacy in exocytosis. The latter account not only suggests that the foregoing interpretation of his view is correct, but also commits him to it inasmuch as it forbids a deterministic 'wandering' by the brain through the available plans of action (on the assumption that collapse occurs sporadically). Since the outcomes of sporadic reductions in the brain are the products of chance and physical determinism, the will is not free at all by Stapp's criterion (§III) if it is forced to stabilize one specific such outcome (say, the first). It can be free only insofar as it can wait for the desired outcome to enforce it, and ultimately free only if it can wait for it any finite amount of time.

It might be hoped that we could obtain a better picture by giving the will not only the option to stabilize any outcome, but also that not to stabilize any outcome. This, however, does not advance us much as far as freeing the will from temporal constraints goes. The relevant empirical questions each divide a brain's quantum mechanical state space in two partitions corresponding to their two possible answers. Let us see what the proposed third option of letting the brain free to oscillate between two such regions amounts to. It may or may not in fact oscillate, as it may repeatedly collapse onto the same state even without the intervention of the Zeno effect. If it does not oscillate in a significant manner, one of the two plans of action will be implemented (e.g. the plan to move a finger). If it does, then presumably some novel (and strange) behaviour will result, such as a series of beginnings to move a finger. A complete spectrum of such unstable behaviours is possible, as the brain may alternate in many different ways between the two relevant states. We can nevertheless say without loss of generality that three plans of action may be implemented. In the case at hand, they would be the plans to move a finger, not to move a finger, and to alternate between these two actions. None of these three courses of action can be said to be more likely than the others in general. However, the two possible outcomes of the relevant empirical question must be considerably likely; otherwise the brain would not even be wandering. If that is the case, unstable behaviour is much more likely than anything else in absence of the Zeno effect (how likely it is depends on how much the brain is supposed to wander). We can now see that there are two problems with the proposed option of letting the brain wander. First, the course of action most likely to result from this makes little sense and should not be desired by anybody. Secondly, whatever the will desires (even if it wants an unstable brain), the course of action the brain ends up implementing would be determined randomly. On Stapp's view, that is to say that there would be no free choice. It thus appears that the proposed third option is of no help; the will must be patient to earn its metaphysical freedom.

Although the above suggests complications, there does not appear to be any *principled* difficulty with Stapp's general approach. It is true that there is no established theory resolving the measurement problem in today's physics, so one may speculate that conscious will is the missing factor. However, Stapp never gives more details than what I exposed here, except for claims to the effect that collapse could be associated with the direction of attention (1993).[12] In §VIII I argue on the basis of uncontroversial psychological data that the Zeno effect is unlikely to be exploitable in the way he suggests.

[12] In particular, note that Stapp does not give details concerning the critical point discussed here in the recent additions made to his 1993. The following is the most complete statement of the proposed theory as it appears in the 2004 edition of this work:

> … owing to its quantum nature, the brain necessarily generates an amorphous mass of overlapping and conflicting templates for action. Process I acts to extract from this jumbled mass of possibilities a dynamically stable configuration in which all of the quasi-independent modular components of the brain act together in a maximal mutually supportive configuration of nondiscordant harmony that tends to prolong itself into the future and produce a characteristic subsequent feedback. This is the "Yes" state *PSP* that specifies the form of the process I event. But the quantum rules do not assert that this preferred part of the prior state *S* necessarily comes into being: they assert, instead that if this process is activated –say by some sort of "consent"—then this "Yes" component *PSP* will come into being with probability Tr*PSP*/Tr*S*.
>
> The rate at which consents are given is assumed to be increasable by mental effort.
>
> The phenomenon of "will" is understood in terms of this effort control of Process I, which can, by means of the quantum Zeno effect, override strong mechanical forces, and cause a large deviation of brain activity from what it would be if no mental effort were made. (2004, p. 255)

What is missing from this passage, as from the above, is an explanation of how the will can in any interesting sense *choose* what the brain will do if the state the latter ends up into after measurement is randomly selected. The only potential answer that comes to mind is that the will probes the brain until it measures the desired state.

V: Stapp's Account of the Nature of Consciousness

I now turn to the second side of Stapp's project, which is to explain the nature of consciousness with the help of special features of QM. The characteristic of consciousness Stapp concentrates on is its unity. We will see, however, that his solution to this problem depends on a corresponding quantum mechanical explanation of qualia and intentional states. He sketches such an account in his 1993 book.

1. The unity of consciousness

The unity of consciousness has been a popular theme since Kant grounded his critical system on this purported phenomenological datum. The idea that conscious experience is somehow continuous and brings together various isolable phenomena (sounds, colors, shapes, etc.) is very intuitive. It is the basis for the picture of the mind as a blank screen on which experiences are projected – what Dennett (1992) calls the 'Cartesian theatre'. Whether this character of consciousness is illusory or real remains debated, along with the possibility of explaining it in the framework of classical physics, but there is no room here to enter into these considerations. We can only proceed to see what can be done with QM on the assumption that Stapp is correct in claiming that the unity of consciousness requires an objective physical correlate that is not available in classical physics, in which only contact interactions are found (Stapp, 1995; see also Silberstein [2001] for a similar point of view).

Stapp (1993, 1995; 2004, ch. 12) wants to explain the unity of consciousness by appealing to the holistic character of process 1. This holism is best exemplified by the well-known EPR effect: a measurement on one of a pair of entangled particles also affects the other, independently of the distance and obstacles separating them. Stapp argues that the unity of consciousness can be explained by the fact that the physical activity underlying conscious experience cannot be regarded as a collection of separate physical events (e.g. individual neural firings), because conscious events correspond to actual events (collapses) occurring holistically over large portions of one's brain. The scope of collapse in one's brain thus determines the extent of the unity of one's consciousness.

Interestingly, Stapp (1995, 1997, 2000, 2001) maintains that this account does not depend on the possibility of entanglement at the macroscopic level. He claims that entanglement is not a necessary condition for collapse to be holistic; it is only a necessary condition for our being able to perform experiments that reveal this holism (as in the typical EPR setup). I discuss this explanation below (§VI). For now I will follow Stapp and assume that there can be superpositions of macrostates without entanglement.

2. Thoughts and qualia: the content of consciousness

Stapp's writings are focussed on the binding problem and the problem of freewill; however, it is clear that the explanatory value of the solution proposed to the former depends on the reducibility of the content of phenomenal states to the quantum states that result from actual events in the brain.[13] At a minimum, there should be correlations between conscious states and the quantum mechanical states of this category. These may be arbitrarily complex, but the states actualized by process 1 must on Stapp's picture relate in a non-random manner to the content of consciousness. Besides the fact that significant correlations are obvious (e.g. alcohol has a determinate effect on my conscious experience), the holism of actual events in the brain would be no more relevant to the explanation of the nature of consciousness than the holism of actual events occurring anywhere else

[13] As a referee pointed out to me, the reduction could be to actual events, that is, transitions from states that can be in superpositions of the relevant observables to states that cannot. Stapp's wording sometimes suggests this. In his 1996a, for instance, he says that 'the collapse of the wave function ..., is the brain correlate of a corresponding psychological or experiential event.' (p. 202) But there are two ways to interpret this. Take the set S of superpositions which have state A as one of their terms. S includes A, as well as superpositions of A and B, superpositions of A, B and C, etc. On the first interpretation, the experience resulting from a collapse onto A should be the same whatever the pre-collapse state member of S is. That is to say that the *content* of the experience is correlated with the resulting state, although its *timing* is correlated with the occurrence of the transition. On the second interpretation, the content of the experience may vary with both the reduced and pre-collapse state; it is correlated with pairs of pre-collapse and reduced states. In this case a transition from a superposition of A and B to A may not be associated with the same experience as a transition from a superposition of A and C to A. When Stapp says that the physical correlates of experiential events are collapses, he may mean either interpretation. I take it that his referring to collapses instead of quantum states resulting from collapses is intended to emphasize the timing correlation, even though the content of the experience is correlated only with the state resulting from collapse. That is to say that the interpretation I favor is the first. It is suggested by the example Stapp gives after the above quotation. He says that one's experience of wanting to raise one's arm should be correlated with a 'physical command' to that effect, which should be a specific state followed by specific changes, not a pair of states one of which may be a superposition of 'raise arm' and 'don't raise arm.' In his 1999 he also adds precisions with a similar import after making a statement just as ambiguous as the above: he explains that his account 'ties in a practical way into … studies of the correlations between, on the one hand, brain activities of a subject, as measured by instrumental probes and described in physical terms, and, on the other hand, the subjective experiences …' (p. 153). If the tie in question is to traditional neuroscience, no reference should be made to superposed states as correlates of the content of phenomenal states. More general considerations also support the first interpretation: 1) the fact that the second is phenomenologically implausible (it implies that we are somehow aware of superpositions); 2) the fact that Stapp's reductionist account as exposed in the present section is from mental states to quantum states, not transitions.

in the universe if the states they actualize did not somehow relate to the content of consciousness. In other words, the hypothesis rests on the supposition that the particular way in which particular phenomena are united in conscious experience depends on the particular way in which particular brain processes interact: it is supposed that it is because such and such neurons have their state collapsed in a holistic manner that such and such experiences come together in consciousness. This background assumption is made especially clear in light of Stapp's (1995) critique of classical physics as a framework for the explanation of the unity of experience. In essence, he objects that the cluttered nature of the physical substratum underlying consciousness would be reflected in its content. His appeal to evolutionary considerations to justify (and perhaps naturalize) his account of freewill also rests on the existence of lawful correlations between mental events and quantum states, since a causally efficacious conscious will would be a disrupting force if its linkage with the rest of the world were not lawful (see his 1996b and 1999, p. 156).[14]

If phenomenal consciousness must in principle be reducible to quantum states for Stapp's project to succeed, so must all forms of thought (e.g. beliefs, desires, intentions), inasmuch as we can be conscious of those along with more immediate experiences. Stapp also acknowledges this commitment to the reducibility of mental states to quantum states. In fact, he presents his theory as superior to Eccles's precisely for this reason (1993, 2001). His views on the matter are made clearest in his 1993 (p. 155-160). There he tells us that any conscious experience can be regarded as 'the feel of an [actual] event in the top-level process occurring in a human brain.' He describes such events as actualizing top-level 'symbols' made of combinations of smaller symbols that correspond to self-facilitating neural patterns capable of being recalled through partial activation. He takes as a 'basic principle' that 'the compositional structure of the feel [or meaning] of a top-level event is isomorphic to the compositional structure of the symbol actualized by that event,' and draws a key distinction between 'the elemental, or absolute, units of experience, such as the immediate direct experience of redness, or of the pitch of high C,' and 'the meanings of symbols that arise from their compositional structure.' In sum, qualia correspond to non-decomposable patterns of neural activ-

[14] Another opportunity to highlight the reductionist commitments that come with Stapp's approach to the binding problem will arise in light of complications due to entanglement (§IX).

ity, while meanings –intentional states— correspond to combinations of such basic patterns into more complex ‘symbols.’ One feels such meanings and qualia when they are holistically actualized by process 1 occurring in one’s brain.

The quantum mechanical binding mechanism and the connectionist terminology notwithstanding, the present account is very close to Hume’s, on which meanings (ideas) are also combinations of irreducible bits of sensation. It also echoes a well-known brand of phenomenalism due to Carnap. My aim is not to criticize this aspect of Stapp’s theory, but it must be remarked that this kind of conceptual empiricism has long been deemed untenable, even by Carnap himself (see Quine, 1951). What is more important for our immediate purpose is that Stapp’s proposal is a form of reductionism (in the epistemological sense): it is meant to capture the ‘correlations between, on the one hand, brain activities of a subject, as measured by instrumental probes and described in physical terms, and, on the other hand, the subjective experiences, as reported by the subject …’ (1999, p153).

VI: Quantum Decoherence and its Interpretation

‘Decoherence’ has come to be the first word in the mouths of all critics of quantum mechanical theories of consciousness. I consider the relevance of decoherence to Stapp’s theory in this section before raising more serious issues in §VI through §IX.

A conservative, uncontroversial exposition of decoherence should be limited to the formal level: the interaction of a system with its environment can have the consequence of changing its mathematical representation from that of a pure state to that of a mixture of states (Zeh, 1996; Joos, 1996; Zurek, 2003). The relevant kind of interaction turns degrees of freedom of the environment into indicators of degrees of freedom of the system. In other words, the system becomes entangled with its environment following a transfer of (quantum) information (Nielsen and Chuang, 2000). The rules of QM entail that the state of a system must be represented as a mixture in such circumstances if it is considered in isolation from its environment.

The fact that decoherence changes the representation associated with a system from that of a pure state to that of a mixture is often regarded as significant because mixtures can be thought of as classical states. That is to say, a mixture can be interpreted as a representation of a system which is in a state without significant superpositions, albeit an unknown one. This is just what the formalism of mixtures was introduced to model in the first place, and we cannot think of pure states in this way (see Joos, 1996). The two kinds of states are in principle empirically distinguishable, because mixtures do not generate entanglement and interference effects.

Decoherence might pose a problem for Stapp's theory for two reasons: first, it makes the existence of some of the macroscopic quantum effects he wants to rely on more questionable; second, it could remove a great deal of the motivation for supposing that consciousness is the key to the measurement problem. I look at the first issue here and turn to the second later (§VII) in the broader context of an assessment of his commitments with respect to the interpretation of QM.

Tegmark (1999) concludes on the basis of conservative assumptions that the decoherence time of neurons interacting with environing particles in the brain and ambient radiation should be of the order of 10^{-20} seconds, which is very fast compared to the 10^{-3} scale of their maximal firing rate. It follows that even if a single atom (e.g. a calcium ion) were controlling the firing of a neuron and that this atom were in a superposition of positions, by the time the neuron fires it would not be in a pure, but in a mixed state. Hagan, Hameroff and Tuszyński (2000) criticize Tegmark's calculations in the case of sub-neuronal microtubules, but their revised estimates still fall significantly short of the required decoherence delay by three orders of magnitude (Mulhauser, 1995). In any case, long-lasting, neuron-level coherence remains clearly hopeless, because neurons interact much more than microtubules with their environment.

It might seem surprising at first glance that Stapp (2000) welcomes these results, claiming they support his main thesis. I mentioned earlier that he emphasizes that his theory of the unity of consciousness does not rely on entanglement (§V.1), which requires coherent states and is ruined by decoherence. Stapp makes two main claims concerning decoherence: that it is compatible with his theory, and that his account in fact makes good use of it. The second seems plausible: by eliminating the possibility of actual events in unusual bases, decoherence presumably rules out correspondingly unusual conscious experiences (see Zurek [2003] for a good explanation of basis-choice by decoherence). But the first point is more controversial. According to Stapp, his theory can be seen to be compatible with decoherence if we correctly interpret the mathematical representation of a decohered state.

The relevant question of interpretation concerns the classical nature of decohered states. Zeh (1996, p. 23), Mohrhoff (2002), Stamatescu (1996), Pessoa Jr (1998), d'Espagnat (1966), Joos (1996), and Hughes (1989, p. 283) all hold that the mathematical form obtained when a system entangled with its environment is abstracted from it should not be interpreted as the description of a classical state. They point out that there are two possible interpretations of these decohered states. In the terminology introduced by d'Espagnat, we can regard them either as proper or improper mixtures. Improper mixtures are superpositions (e.g. 'live' and 'dead' for Schrödinger's cat) just like normal superposed pure states, but incapable of generating interference effects. Since improper mixtures can, like pure states, be maximally known yet indefinite with respect to macroscopic observables, such states are quite different from proper mixtures, which are classical but unknown states. There are two good reasons to refrain from assuming that decoherence causes collapse into proper mixtures.

The first is that decoherence is relative to what is measured. As Joos (2000) explains, very little seems to be solved by decoherence when we look at the larger system constituted by a system plus its environment. What decoherence does is move the interference terms out of the scope of a decohered subsystem, but these are found again if we look also at the environment: 'The interference terms still exist globally in the total (pure state), although they are unobservable at either system alone –a situation which may be characterized by the statement, `the interference terms still exist, but they are not there.`' (Joos, 2000) It follows that only a subset of the degrees of freedom of the universe could ever be decohered at once, its complement serving as the 'environment.' This prediction of QM concerning the preservation of coherence at the level of system and environment would be contradicted if decoherence caused collapse, so the states it results in must be regarded as improper mixtures.

The other reason not to think of decoherence as a collapse mechanism is that this would be begging the question against the original measurement problem: decoherence occurs through a process of correlation between environmental and systemic degrees of freedom which is formally the same as that obtaining between a measuring apparatus and a measured system. Assuming that decoherence can engender collapse would thus amount to assuming that an apparatus can collapse the state of a system by entangling with it, which would be begging the question against the measurement problem, while at the same time removing any need for decoherence to explain the absence of macroscopic interference effects.

Given that we can and should think of decohered states as improper mixtures, Stapp's accounts of freewill and of the content/unity of consciousness are, on their face, compatible with the results found by Tegmark: improper mixtures can be subject to the Zeno effect (but see Joos [1996] for exceptions which do not seem to apply here), and all that is needed for holistic collapses in the brain is the existence of macroscopic superpositions.

VII: Issues of Interpretation and Coherence

If the theory is compatible with decoherence, Mulhauser (1995) argues that it is not plausible because decoherence works in favor of no-collapse interpretations, which are incompatible with Stapp's views. It has also been argued that the collapse models of the kind it requires are questionable independently of the importance of decoherence (Vaas, 2001). But it is not so clear that Stapp's theory requires a collapse interpretation (although he does endorse such an interpretation). And the exact nature of the collapse model he works with is not even clear – he claims to be taking it from Heisenberg, von Neumann, and Wigner, which makes for a puzzling mixture (see his 1993). I clarify his interpretive commitments in the next subsection, and I criticize this aspect of his theory in the following one.

1. Stapp’s interpretive commitments and their consistency

It should be clear that Stapp’s accounts of freewill and of the unity of consciousness require an interpretation of QM on which simple information-processing devices are not endowed with the capacity to trigger process 1. In his 1993 book he claims to be following Heisenberg in recognizing that collapse occurs in the presence of any measuring apparatus (e.g. a photoelectric plate).[15] But if that were the case, single neurons should also be able to trigger process 1, which would both remove the need for the will to play this role and eliminate the macroscopic superpositions of brain states that are supposed to be collapsed holistically to yield a unified conscious field. Perhaps as a result of noticing this problem, he repudiates this view in his 1999 paper (p. 149), holding instead that collapse occurs only at the level of the human brain, and not at the level of its subcomponents or external objects.

Stapp does not give much more detail concerning the conditions of collapse, but it is worth distinguishing different ways of interpreting his new stance. The hypothesis that collapse occurs at the level of the brain is compatible with four kinds of interpretations corresponding to the combinations of answers we can give to the questions:

(A) Is collapse real or apparent?

(B) Can the conditions resulting in a real or apparent collapse be described quantum mechanically?

It is not hard to see how we must answer the first question. If the Zeno effect is more naturally described in terms of collapse events, it also works with Everett-type interpretations (Joos, 1993). This means that Stapp’s proposal for freewill could be reformulated in the context of such a no-collapse interpretation, in which ‘collapse’ would be replaced by ‘branching’ to yield exactly the same empirical predictions. However, his account of the unity of consciousness is dependent on a collapse interpretation, for it ties conscious events to the occurrence of such collapses.

[15] ‘the actual events associated with human conscious experiences are not presumed to be the only actual events: actual events associated for example with the firing of a Geiger counter are presumed to exist, as Heisenberg assumed. Here it is merely accepted that, under similar conditions, the brain, which also is a physical system, should also be subject to the collapsing action of actual events.’ (1993, p. 167)

It is trickier and more interesting to unpack Stapp's commitments with respect to the second question. If process 1 (real or apparent) takes place only in the neighborhood of human brains, either there are correlations between the quantum states of those systems and the occurrence of process 1, or there are not. The second option is in line with Wigner's earlier interpretation, on which (actual) collapse is caused by a form of consciousness that is irreducible to physical happenings. Consciousness could also be involved on the first view, but its manifestations would be correlated with the occurrence of certain brain-level quantum states, which states could be regarded as causes of collapse. We can thus say that these two lines of interpretation are incompatible in that on the former there should be a correct model describing which quantum states cause collapse, whereas such a model should not be available on the latter.

Interestingly, Stapp's accounts of freewill and of the unity of consciousness appear to commit him to both of these kinds of interpretations at once, while his proposal concerning freewill is clearly incompatible with the second. As I emphasize in §V.2, his account of the unity of consciousness requires that the content of conscious states be correlated with the quantum states that result from actual events in the brain. The content of conscious intentions must thus also be correlated with such quantum states. Take CI, the set of quantum states that have as their correlates conscious intentions. As explained in §IV, the content of such conscious intentions is supposed on Stapp's account of freewill to influence the brain by determining the timing of process 1. If conscious intentions are correlated both with the quantum states of CI and the occurrences of process 1 that are relevant to free choice, it follows that these collapses are correlated with the quantum states of CI as well. There are complications due to the fact that collapses are temporally separated on Stapp's view, but it nevertheless follows that there should be a model (however complex) enabling in principle the prediction of the actual events which are relevant to freewill. Furthermore, it should be possible to tell which empirical questions are asked in these occurrences of process 1, since Stapp already gave us a formula for this (§IV). It thus seems that his account of the unity of consciousness, in conjunction with his account of freewill, implies that we could in principle explain the occurrence of the actual events relevant to the latter in purely quantum mechanical terms. The resulting model would make conscious will superfluous in describing one's behavior, which is just what he thought was the problem with the classical framework (§III). It appears that if we take the path he suggests to explain the unity of consciousness, we have to give up his account of freewill, and vice-versa.

2. The viability of Stapp's interpretive commitments

If the two components of Stapp's theory require incompatible interpretations of QM, the next question we must ask is whether these are viable when taken independently.

The interpretation that comes with Stapp's proposal for freewill gives raise to familiar metaphysical problems. First, the claim that process 1 cannot be given a physical explanation is tantamount to the claim that quantum mechanics is not complete *and* cannot be completed. Wigner's interpretation has often been reproached as implying this dualism between physical laws and 'other' factors. In addition, Stapp's accounts of freewill and consciousness come with commitments to the solipsism of Wigner's interpretation: both require that collapse be triggered at the level of brain-like mechanisms (whether by physical brains or by separate consciousnesses). As a result, the physical world we do not look at is downgraded to the status of meaningless superpositions. Wigner admits that this consequence of his view spells a 'return to the spirit of Descartes's *Cogito ergo sum,* which recognizes the thought, that is, the mind, as primary.' (1962, p. 98) The unity of science and the full existence of unseen objects are still debated points in philosophical circles, so it is hard to blame a theory unconditionally for denying them. But if ever there were good aesthetic and heuristic reasons for setting aside a hypothesis, these are two prime examples. It is worth noting that Wigner himself repudiates this position in his latest writings, expressing the need for 'a less solipsistic theory.' (1973, p. 382)

In addition to the aforementioned solipsistic implications, Stapp's account of the unity of consciousness faces a difficulty due to the requirement that collapse be a real phenomenon. From the general to the more specific, two questions must be addressed. First, are collapse interpretations in general sufficiently credible to support his hypothesis? Secondly, are the collapse models of the particular kind he is committed to plausible?

Stapp systematically attacks the major no-collapse interpretations of QM (many-worlds, many-minds and consistent histories; Stapp 1993, 1996a, 1997). He raises the usual issues against them: the preferred basis problem, the problem of accounting for the continuity of consciousness through branchings, and the problem of making sense of the projection postulate on such interpretations (they make all terms of a superposition equally actual). These problems have been around since Everett's original formulation, and even a strong many-worlder such a Barrett (1999) admits that they might be intractable. Everett's way thus seems controversial enough for Stapp to feel entitled to work on the basis of a collapse interpretation.

Mulhauser (1995) argues that the recent discovery of quantum decoherence represents a significant gain for no-collapse interpretations in general, if not a solution to all their problems, and should thus count in their favor. It is true that decoherence helps no-collapse interpretations explain the absence of macroscopic interference effects and get around the preferred basis problem (see Zurek, 2003). But it does not account for the subjective probabilities given by the empirically correct projection postulate. In spite of the explanatory power of decoherence, most physicists remain neutral concerning the collapse versus no-collapse question (see Giulini et al*, 1996).*

If we cannot at this time dismiss a theory such as Stapp's on the general collapse/no-collapse point, it is important to see whether the type of collapse model it requires is plausible in itself. The currently available explanations that have the potential to completely bridge the gap between process 1 and process 2 are:

a) Heisenberg's (and also Bohr's).

b) Wigner's.

c) Gravitational and spontaneous localization models such as those of Pearle (1990), Ghirardi et al (1990), and Penrose and Hameroff (1996).

Clearly, none of these will support Stapp's theory of the unity and content of consciousness. (a) and (c) will not do for the reason that they locate collapse at very small scales: neuron-scale measurement apparatuses should trigger collapse on Heisenberg's view (see §VII.1), and even smaller systems should automatically collapse on the theories of (c). Neither could Wigner's interpretation lend any support to Stapp's theory. On this interpretation, it is consciousness that triggers collapse. But consciousness is supposed to be reducible to brain states on Stapp's account (§V.2). So to appeal to Wigner's interpretation in defense of Stapp's theory is just to claim that only brain-level physical information processing systems can trigger collapse —precisely the hypothesis that is *prima facie* implausible and needs to be defended. There are many other purported solutions to the measurement problem, but the above are the only credible real collapse models normally mentioned (see Stamatescu, 1996). The other main contenders explain collapse away by various mechanisms (including decoherence), which will not work with Stapp's account of the unity of consciousness (§VII.1).

I now turn to internal problems with Stapp's theory. These are unrelated to the interpretation of QM.

VIII: Freedom Through Paralysis?[16]

Let us again take the example of my wanting to make some simple movement, e.g. lift my little finger. This should be a paradigmatic case of free decision if I have no specific reason or incentive to do it or not. My brain is on Stapp's view 'wandering' as I face this choice. That is to say that its wave function cyclically superposes into states corresponding to the yes and no answers to the question 'am I lifting my little finger?' before collapsing into one of these picked at random. I am supposed to enforce my free choice by holding it onto either the yes or no answer once the one I have chosen arises, this until completion of the action (let us suppose for simplicity that the counter-intuitive and unhelpful option of not stabilizing the brain is not open).

We can analyze the proposed model from the complementary perspectives of spatial and temporal organization. Let us call the relevant degrees of freedom the 'quantum switch' (QS). From the spatial perspective, these could be virtually anything: they could be macroscopic (as Stapp suggests, see §IV), or they could correspond only to microscopic parameters such as the position of a few

[16] This section has been greatly improved as a result of the constructive criticism of an anonymous referee.

calcium ions. Stapp's model entails more constraints in the temporal dimension. Without loss of generality, we can say that two rates of measurement are implicated: one corresponding to the time that separates measurements before the right answer is obtained (ΔT_s), and one corresponding to the time separating them when the switch must be kept in the right state (ΔT_a). The search for the right answer may take a certain time T_s, and once it is found it must be reinforced sufficiently long for the action to reach completion (T_a). This process is illustrated in Figure 3.

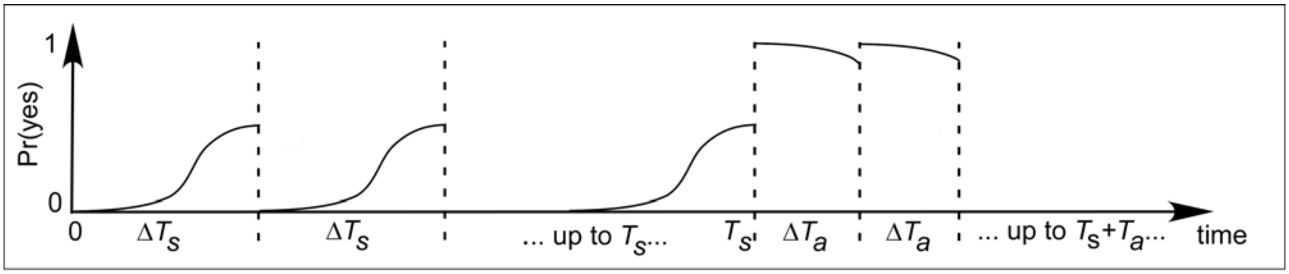

Fig. 3: The process of enforcing a free decision using a quantum switch. The vertical axis represents the probability that the switch collapses onto the 'yes' state when measured, which is the one that is desired in this example. Each dashed line marks an actual event. The first phase of the process lasts until the desired outcome is obtained. The will has up to *Ts* to obtain this outcome. In this example the desired outcome occurs at *Ts*.

Psychological and phenomenological data may impose important restrictions on the preceding temporal variables. We should ask first whether there is any limit to the frequency of collapses in the brain. Each reduction must be a conscious event, for the conscious will must know what its result is. This means that if it turns out that conscious events have a minimal duration, the rate of collapse in the brain will be limited accordingly**.** This question has been studied before, and it has been found that, depending on the stimulus conditions, two events which occasion simple conscious experiences must be separated by approximately 20 to 50 ms to be perceived as separated in time (Kristofferson, 1969; Hirsh and Sherrick, 1961).[17] Also, Varela (1997) suggests that the minimal duration of complex conscious states (e.g. intentional states) is probably considerably longer (in the order of 100 ms). Stapp's account deals with free choices that express complex thoughts, so the latter timeframe is the most relevant. However, it will soon become apparent that long intervals are problematic on his model, so let us use the most conservative estimate. Given that the frequency of collapse must be higher when a course of action is being enforced, we have:

$$\Delta T_s > \Delta T_a \geq 20 \text{ to } 50ms$$

[17] These are results anybody can verify at home, as television standards specify field rates (refresh rates) only a little bit higher than our perception threshold (there are standards at 50hz and 60hz).

We should next ask how much time one normally has at one's disposal to set one's quantum switch in the desired state. Each time a switch collapses, one experiences a specific course of action being initiated (Stapp, 1999). We may characterize this feeling as an urge to act in a certain way. Libet (1983; 1999) found that subjects asked to take a simple action of their choice experience such feelings only about 200 ms before their muscles start contracting to perform the action. Correcting for a typical bias in their reports and the time it takes for a nerve influx to travel from a subject's brain to her hand, Libet (1999, p. 51) estimates that subjects typically have only 100 ms of relevant conscious activity before being irrevocably committed to an action. The time T_s it takes for a quantum switch to collapse onto the intended state must be bounded accordingly:

$$T_s \leq 100ms$$

We can now calculate the probability that I fail to do what I want because I failed to get my quantum switch in the desired state. In the case at hand, my QS tends to evolve into a superposition of the two states corresponding to the two courses of action 'lift little finger' and 'do not lift little finger'. This superposition may make one outcome more likely than the other, but since each course of action must be available to me, the case in which both have probability .5 should be optimal. If I have 100 ms to make a conscious choice and the frequency of collapse in my brain during the search phase is under 20 to 50hz, my QS can be collapsed a maximum of 2 to 5 times before it is too late. If the probability of its falling in the right state is each time .5, the likelihood that it does not before it is too late is between 3% and 25%, depending on what we take the exact value of ΔT_s to be.

It should be clear that even the best case (3%) is unacceptable. Since the state of my QS determines a course of action and its related experience, I should have something like a feeling of not doing what I want if I fail to set my quantum switch according to my conscious decision. In other words, I should literally have an experience of weakness of will.[18] Whether there really are genuine

[18] The phenomenology associated with such circumstances depends on what we take the relevant empirical question to be, but the interpretation suggested here appears to be the only viable one. Besides the form 'am I lifting my little finger?', we may entertain the possibility that the question be formulated as 'do I want to lift my little finger?' or 'is my brain making me lift my little finger?', but neither will do for obvious reasons. In the first case the will becomes subordinated to the experience engendered by the brain, and should therefore never oppose the latter's dynamics. In the second case we simply misconstrue the experience of lifting one's little finger, which is something that one does, not one's brain. Now if the relevant empirical question is of the form used here, it follows that failure to set one's QS in the proper state would result in a feeling of doing something that is against one's will.

cases of akrasia is a question that remains debated. It is clear, however, that the circumstances in which we experience what we may want to describe as weakness of will typically involve conflicting motives, like that to be slim and healthy and that to eat chocolate. For my part, I never had the feeling of not doing what I want in the kind of situation described here; if all I desire is to move a finger, I never fail at that. In general, it seems that one should never have a feeling that would be associated with failure to set one's QS in the desired state for such free choices unless one suffers from a pathological condition impeding motor control. I suggest that anybody can take a few minutes to verify that the lowest average frequency of such events the model allows (3 out of 100) is far too high (and we must keep in mind that the relevant ΔT_s is most likely over 100 ms, which would mean entirely random actions). The problem is not so much that the will turns out not to be completely free, but that the proposed picture is in tension with daily experience.

Let me emphasize that these results are independent of the mechanism that is supposed to implement the quantum switch; they are in fact independent of the quantum nature of Stapp's proposal. They depend solely on the premise that the will must control the brain by forcing it to stay in a state in which it evolves randomly, whether this randomness be relative to its ignorance of deterministic brain processes or due to quantum superpositions of states. I gave a quantum presentation of the issue only because this is Stapp's framework of choice.

Two other difficulties are worth mentioning. The first is that it seems doubtful that there be a quantum switch for each of the huge (if not infinite) number of actions we can elect to perform. In particular, the model appears to run against common wisdom in cognitive science and linguistics if free speech is free action (Wilson [1999] makes a similar point). Another problem arises specifically from Stapp's use of the quantum framework: it seems that the dynamics of the quantum switch would have to be very peculiar in order for measurements occurring at a rate of one every 20 ms to virtually stop its evolution even though measurements occurring every 50 ms yield random results.[19]

[19] Atmanspacher *et al* (forthcoming) provide a model of just this kind of quantum switch. This model would require that T_s be greater than T_a to yield acceptable probabilities of weakness of will, while on Stapp's account T_a can be in the order of several seconds, but T_s must be restricted to 100 ms. Since this example uses a simple oscillatory system, this suggests that no such mechanism could play the role of the quantum switch Stapp's theory requires.

IX: Individual Consciousnesses in an Entangled Universe?

According to Stapp, the scope of the unity of one's consciousness corresponds to the scope of collapse of one's brain's wave function during conscious events. This claim is compatible with decoherence to the extent that the interpretation of decohered states as improper mixtures is plausible. But the facts highlighted in the study of decoherence can have strange implications for the resulting picture, as it appears that any macroscopic object should be heavily entangled with its environment (§VI): this implies that superpositions of brain macrostates can disappear only through actual events that encompass both the brain and its environment. It thus seems that on Stapp's account my conscious experience depends not only on the physical state of my brain, but also on huge numbers of particles which have become entangled with it and are bouncing around randomly far beyond the limits of my body. Further, the same entanglement should also be found between those degrees of freedom in my brain which we may suppose are relevant to conscious experience and those which are not (whatever the relevant degrees of freedom are, it is plausible that not all physical activity in the brain is consciously felt).

The question that arises is, why are only certain subsystems within the scope of an actual event relevant to conscious experience? It might be hoped that a reductive model of conscious states to quantum mechanical states could answer this question. However, Stapp's current proposal does not do this (in addition to being unsatisfactory because of its phenomenalism, as noted in §V.2). The properties he attributes to symbols as physical patterns are certainly not characteristic of the physical processes that are relevant to conscious experience, for they are merely described as patterns of activity in associative networks (§V). In particular, consider that we can have unconscious thoughts, while Stapp's account is of thought in general. A simple, uncontroversial case is that of searching for a word without consciously thinking about it. One moment I cannot remember how to say P in language L. A few words come to mind, but none is appropriate. I stop thinking about it completely (that is, it seems that I do), and a few seconds later the answer springs to mind. Something very close to thinking has been taking place in my brain without my being aware of it. For a more general perspective on the unconscious and the recent revival of pertinent aspects of Freud's view, see Solms (2004). If one does not want to admit unconscious thoughts in the picture, one may compare the patterns of neural interaction described by Stapp with those that can be expected in the cellular signaling between proteins at the molecular level (feedback loops, amplification patterns, etc.). There are certainly brain processes which are irrelevant to conscious experience[20] yet follow patterns of activity of the same type as those found in Stapp's account.

In general, it is hard to see what is the advantage of invoking quantum holism if one has to provide a model that partly negates it. The problem mirrors that of accounting for the unity of consciousness in a world ruled by local interactions: in classical physics we must explain why the locality of physical processes does not affect the unity of consciousness; in quantum physics we must explain why the exceedingly broad holism of physical processes does not affect the locality of consciousness.[21]

X: Summary

Quantum theories of mind are often dismissed by appealing to decoherence and derision. I made a charitable presentation of what seemed to be the most promising such theory. I then outlined the real challenges it faces, as intuitively appealing as it can be at first sight. These are partly related to decoherence, but not in the way normally suggested (e.g. by Mohrhofff, 2002; Mulhauser, 1995; and Vaas, 2001). First, Stapp's theory does not require macroscopic entanglement, and it is plausible to regard decohered states as improper mixtures (§VI). Secondly, the debates over collapse and no-collapse interpretations (often associated with decoherence) do not appear to be close to resolution (§VI.2). The issues I raised are both more serious and more straightforward.

Most importantly, Stapp's accounts of freewill and of the unity of consciousness depend on incompatible interpretations of quantum measurement (§VII.1): the first requires the absence of sufficient conditions expressible in terms of quantum mechanical states for the occurrence of process 1

[20] Except of course as enabling conditions.

[21] Note also that this line of reasoning poses an additional difficulty for Stapp's overall theory, since it calls for a reductive model of precisely the kind which we saw is incompatible with his proposed account of freewill (§VII.1).

when the Zeno effect is supposed relevant to the reinforcement of decisions; the second implies that a correct quantum mechanical model singling out just these conditions could and should be found, because it depends, as the preceding discussion makes particularly clear (§IX), on the reducibility of consciousness (including conscious will) to quantum mechanical brain states. Once these commitments are well in sight, it appears that Stapp and his followers will have to make a choice and retain only one side of the theory.

I also pointed out that the interpretations associated with each of Stapp's proposals are problematic on their own. On the side of freewill, his account hinges on an interpretation of QM that is deeply unsatisfying for its dualism and solipsism (§VII.2). His model for the unity of consciousness also depends on a solipsistic interpretation, but it faces extra difficulties because it makes the presence of brain-like systems conditional to the occurrence of collapse. (§VII.2).

Finally, it should be clear that Stapp's theory faces serious challenges independently of the interpretation of QM. He does not give much details concerning the role of the Zeno effect in brain dynamics, but even then we can see that the general model seems unworkable, for it entails that weakness of will should be commonplace even in the context of simple choices (§VIII). His account of the unity of consciousness (and, indirectly, of its content) does not fare better. The fact that macroscopic objects rapidly become entangled with their environments implies that the model extends the scope of one's experience to one's environment and irrelevant brain processes (§IX). This consequence could perhaps be avoided with the help of the appropriate reductive account. However, it is not clear that any progress has been made by appealing to the holism of quantum measurement: for my part, I am as inclined to doubt that one could provide such an account as I am to puzzle over the capacity of the classical approach to succeed at explaining the unified character of consciousness.

Acknowledgements
This work benefited from the financial support of the University of Toronto and the Fonds de Recherche sur la Société et la Culture, Government of Québec. I am grateful to William Seager, Allan Olley, Rodney Snooks, Julien Beillard, Maite Herrera, and two anonymous referees for their helpful comments and suggestions. Sari Kisilevsky has been invaluable in making this paper more readable.

References

Akert et al (1975), ‘Structural Organization of Motor End Plate and Central Synapses’, in P. G. Waser, *Cholinergic Mechanisms,* pp. 43-59 (Raven Press/Lippincott Williams & Wilkins)

Atmanspacher, H. et al (forthcoming), ‘Quantum Zeno Features of Bistable Perception’, *Biological Cybernetics.*

Barrett, J. A. (1999), *The Quantum Mechanics of Minds and Worlds* (Oxford: Oxford University Press).

Dennett, D.C. (1992), *Consciousness Explained* (Boston: Little, Brown & co).

D’Espagnat, B. (1966), ‘An elementary note about mixtures’, in De-Shalit, Feshbach, Hove (Eds.), *Preludes in theoretical physics,* pp. 185-191 (Amsterdam: North-Holland).

Eccles, J.C. (1994), *How the Self Controls its Brain,* (Springer-Verlag).

Fogelson, A.; Zucker, R. (1985), ‘Presynaptic Calcium Diffusion from Various Arrays of Single Channels: Implications for Transmitter Release and Synaptic Facilitation’, *Biophysics Journal.* **48**, pp. 1003-17.

Ghirardi, G.C., et al (1990), ‘Continuous spontaneous reduction model involving gravity’, *Physical Review A,* **42**, pp. 1057-64.

Giulini, D. et al (1996), *Decoherence and the Appearance of a Classical World in Quantum Theory.* (Springer).

Hagan, S., Hameroff, S.R., and Tuszyński, E. (2000), ‘Quantum Computation in Brain Microtubules: Decoherence and biological feasibility’, *Physical Review E,* **65**.

Hirsh, I.J. and Sherrick, C.E.J. (1961), ‘Perceived order in different sense modalities’, *Journal of Experimental Psychology*, **62**, pp. 423–32

Hodgson, D. (2002), ‘Quantum Physics, Consciousness, and Free Will’, in R. Kane (ed.) *The Oxford Handbook of Freewill* (Oxford University Press).

Hughes, R. I. G. (1989), *The Structure and Interpretation of Quantum Mechanics* (Harvard University Press).

Itano, W.M. et al. (1990), ‘Quantum Zeno effect’, *Physical Review A,* **41**, pp. 1295-1300.

Joos, E. (1996), ‘Decoherence Through Interaction with the Environment’, in Giulini et al (eds.), *Decoherence and the Appearance of a Classical World in Quantum Theory* (Berlin, New York: Springer).

Joos, E. (2000), ‘Elements of Environmental Decoherence’, in P. Blanchard et al (eds.), *Decoherence: theoretical, experimental, and conceptual problems* (Berlin, New York: Springer). Retrieved online at http://xxx.lanl.gov/abs/quant-ph/9908008.

Kane, R. H. (2002), *The Oxford Handbook of Freewill* (Oxford: Oxford University Press).

Kristofferson, A.B. (1967), ‘Successiveness discrimination as a two-state, quantal process’, *Science* **158**, pp. 1337–39.

Libet, B. *et al* (1983), ‘Time of conscious intention to act in relation to onset of cerevral activity (readiness potential): The unconscious initiation of a freely voluntary act’, *Brain,* **106**, p.. 623-42.

Libet, B. (1999), ‘Do We Have Free Will?’, *Journal of Consciousness Studies,* **6** (8-9), pp. 47-57.

Lockwood, M. (1989), *Mind, Brain and Quantum* (Oxford: Oxford University Press).

Misra, B., Sudarshan, E.C.G. (1977), ‘The Zeno’s paradox in quantum theory’, *Journal of Mathematical Physics*, **18**, pp. 756–63

Mohrhoff, U. (2002), ‘The World According to Quantum Mechanics (or the 18 Errors of Henry P. Stapp)’, *Foundations of Physics,* **32** (2), pp. 217-55.

Mulhauser, G. R. (1995), ‘On the End of the Quantum Mechanical Romance’, *Psyche,* **2** (5).

Nielson, M.A., Chuang, I.L. (2000), *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press).

Pearle, P. (1990), ‘Toward a relativistic theory of statevector reduction’, in A. I. Miller (ed.), *Sixty-Two Years of Uncertainty* (New York: Plenum).

Penrose, R., Hameroff, S. R. (1996), ‘Conscious Events as Orchestrated Space-Time Selections’, *Journal of Consciousness Studies,* **3** (1), pp. 36-53.

Pessoa Jr., O. (1998), ‘Can the Decoherence Approach Help to Solve the Measurement Problem?’, *Synthese,* **113**, pp. 323-346.

Principe, J.S., Euliano, N.R., Lefebvre, W.C. (1999), *Neural and Adaptive Systems : fundamentals through simulations* (Wiley & Son).

Quine, W.V.O. (1951) “Two Dogmas of Empiricism”, *The Philosophical Review* 60 (1951): 20-43. Reprinted in W.V.O. Quine, *From a Logical Point of View* (Harvard University Press, 1953; second, revised, edition 1961).

Rosenblum, B. and Kuttner, F. (1999), ‘Consciousness and Quantum Mechanics: The Connections and Analogies’, *The Journal of Mind and Behavior,* **21** (3), pp. 229-56.

Shear, J. (ed.) (1997), *Explaining Consciousness: the hard problem* (MIT Press).

Silberstein, M. (2001), ‘Converging on Emergence: Consciousness, Causation, and Explanation*’, Journal of Consciousness Studies,* **8** (9-10), pp. 61-98.

Solms, M. (2004), ‘Freud Returns’, *Scientific American,* May 2004.

Stamatescu, I.-O. (1996), “Stochastic Collapse Models”, in Giulini et al (eds.), *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer).

Stapp, H.P. (1993), *Mind, Matter, and Quantum Mechanics,* First Edition (Springer-Verlag).

Stapp, H.P. (1995), ‘Why Classical Mechanics Cannot Naturally Accommodate Consciousness but Quantum Mechanics Can’, in J. King and K. Pribram (eds.), *Scale in Conscious Experience: Is the Brain Too Important To Be Left to Specialists to Study?* (Lawrence Erlbaum Mahwah NJ).

Stapp, H.P. (1996a), ‘The Hard Problem: a Quantum Approach”, *Journal of Consciousness Studies,* **3** (3), pp. 194-210.

Stapp, H.P. (1996b), ‘The Evolution of Consciousness’, *Proceedings of Toward a Science of Consciousness,* Tucson, April 8-13, 1996.

Stapp, H.P. (1997), ‘Science of Consciousness and the Hard Problem’, *The Journal of Mind and Behavior,* **18** (2-3), 171-194.

Stapp, H.P. (1999), ‘Attention, Intention, and Will in Quantum Physics’, *Journal of Consciousness Studies,* **6** (8-9), pp. 143-64.

Stapp, H.P. (2000), ‘The importance of Quantum Decoherence in Brain Processes’, Lawrence Berkeley National Laboratory e-print LBNL-46871. Retrieved online at http://www-physics.lbl.gov/~stapp/stappfiles.html

Stapp, H.P. (2001), ‘Quantum Theory and the Role of Mind in Nature”, *Foundations of Physics,* **31**, pp. 1465-1499.

Stapp, H.P. (2002). ‘The Basis Problem in Many-Worlds Theories’, *Canadian Journal of Physics,* **80**, 1043-1052.

Stapp, H.P. (2004), *Mind, Matter, and Quantum Mechanics,* Second Edition (Springer-Verlag).

Tegmark, M. (1999), ‘The importance of Decoherence in Brain Processes’, *Physical Review E,* **61**, pp. 4194-4206.

Thagard, P. (1996), *Introduction to Cognitive Science* (MIT Press).

Vaas, R. (2001), ‘Why Quantum Correlates of Consciousness Are fine, But Not Enough’, *Informação ϵ Cognição* **3** (3). Online: http://www.marilia.unesp.br/atividades/extensao/revista/v3/artigo4.html

van Fraassen, B.C. (1991), *Quantum Mechanics: an empiricist view* (Oxford: Oxford University Press).

Varela, F. J. (1997), ‘The Specious Present: a Neurophenomenology of time Consciousness’, in J. Petitot, et al (eds.), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science* (Stanford: Stanford University Press).

von Neumann, J. (1955), *Mathematical Foundations of Quantum Mechanics* (Princeton University Press).

Wigner,E.P. (1962), 'Remarks on the mind-body question*',* in I. J. Good (ed.), *The Scientist Speculates,* pp.284–301 (London: Heinemann).

Wigner,E.P. (1973), 'Epistemological perspective on quantum theory', in C. A. Hooker (ed.), *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, pp.369–385 (Dordrecht: Reidel).

Wilson, D.L. (1999), 'Mind-Brain Interaction and Violation of Physical Laws', *Journal of Consciousness Studies*, **6** (8-9), pp. 185-200.

Zeh, H.D. (1996). 'The Program of Decoherence: Ideas and Concepts'; in Giulini et al, *Decoherence and the Appearance of a Classical World in Quantum Theory.* (Springer).

Zurek, W.H. (2003), 'Decoherence, Einselection, and the Quantum Origin of the Classical', *Review of Modern Physics,* **75**, 715. Retrieved online at http://xxx.lanl.gov/abs/quant-ph/0105127.