# Learning in Non-superpositional Quantum Neurocomputers

Ronald L. Chrisley
School Of Cognitive & Computing Sciences
University of Sussex, United Kingdom
ronc@cogs.susx.ac.uk

November 19, 1996

**Abstract**

A distinction is made between superpositional and non-superpositional quantum computers. The notion of quantum learning systems – quantum computers that modify themselves in order to improve their performance – is introduced. A particular non-superpositional quantum learning system, a quantum neurocomputer, is described: a conventional neural network implemented in a system which is a variation on the familiar two-slit apparatus from quantum physics. This is followed by a discussion of the advantages that quantum computers in general, and quantum neurocomputers in particular, might bring, not only to our search for more powerful computational systems, but also to our search for greater understanding of the brain, the mind, and quantum physics itself.

## 1 Quantum computers & quantum learning

In both the search for ever smaller and faster computational devices, and the search for a computational understanding of biological systems such as the brain, one is naturally led to consider the possibility of computational devices the size of cells, molecules, atoms, or on even smaller scales. Indeed, it has been pointed out [Braunstein, 1995] that if trends over the last forty years continue, we may reach atomic-scale computation by the year 2010 [Keyes, 1988]. This move down in scale takes us from systems that can be understood (to a good enough approximation) using classical mechanics alone, to those which require a quantum mechanical understanding. Thus, it should not be surprising to find that the idea of *quantum computation* is not new (see, e.g., [Deutsch, 1985] and [Feynman, 1982]). However, most if not all work so far has been understandably speculative.

This paper continues in this speculative vein, but tries to be concrete in describing what an implementation of a quantum computational system might be like. There are two ways in which the focus here differs from other considerations of quantum computation First, the focus is on quantum *learning*: quantum computers that modify themselves in order to improve their performance in some way. The type of learning that is considered here is that family of algorithms loosely known as *neural networks*, *connectionism*, or *parallel distributed processing*. Second, in order to investigate the possibilities for quantum learning, a distinction is made between two types of quantum computation: superpositional and non-superpositional.

## 1.1 Superpositional quantum computation

Superpositional quantum computations exploit the fact that a coherent quantum state is a superposition of $n$ distinct states, $x_i$, each weighted by some complex scalar $\alpha_i$. Under certain conditions, this quantum state decoheres, and the particle adopts one of the $x_i$ as its determinate state, with a probability that is determined by the ratios of the $\alpha_i$. The idea proposed in [Feynman, 1982] and developed in, e.g., [Deutsch, 1985], is that if such superpositional states were used to implement the states of a computer, then various registers or memory locations in the computer would not be conventional bits with a determinate value of 1 or 0, but would instead be quantum bits – *qubits* – which are superpositions of both the 0 and 1 values.

A key advantage to this would be that $n$ qubits could be used to perform $2^n$ computations in parallel, one computation for each combination of values of the superposed states. However, there are two principal difficulties in exploiting this proposal. First, there is the problem of maintaining the coherence (superpositionality) of a qubit while performing computations on it: the danger is that the kind of physical processes necessary to implement the relevant bit operations are such that they would cause the quantum state to collapse or decohere. But even supposing one can perform such operations while maintaining the coherence of the qubit, there is still the difficulty of exploiting the superpositionality of the qubit in a way that can perform effective computational work.

The only specific idea of how this can be done was proposed by Shor [Shor, 1994]. Shor describes how to initialize a superpositional quantum state with a random number $x$ and a number $n$ to be factored into primes. He then describes how to transform that state into another which is such that the probability distribution of the measurements of the state is a simple function of a number $r$ which is a reliable guide to the prime factors of $n$. A few collapses, then, of this system allows one to calculate $r$ and thus factorize $n$. If this algorithm could be implemented in a real quantum computational system, one could then produce the prime factors of large (e.g., 159-digit) numbers in seconds. Since current cryptography technology relies on the fact that such numbers would take a single computer many years to factor, Shor's algorithm has generated much interest in quantum computation. However, it has proven difficult to generalize this exploitation of the qubit to other applications. The general problem of how to use a superpositional state to do computational work remains.

## 1.2 Non-superpositional quantum computation

However, a form of non-superpositional quantum computation is possible. Non-superpositional quantum computation uses quantum systems to implement standard computational architectures in a way that does indeed involve superposed states, but does not exploit them for computational parallelism in the way that, e.g., Shor's algorithm does. Thus, it does not have the disadvantages of that approach: it can proceed even if the superpositional state cannot be maintained, and it can be generalized to implement any computation whatsoever. But so also it does not have the parallel advantages of superpositional quantum computation. Nevertheless, as the rest of this paper will attempt to show, non-superpositional quantum computation has its own advantages and theoretical interest.

First, mention can be made of the more mundane computational advantages of quantum computation in general: advanatages of size and speed. Non-superpositional quantum computers (NSQC's) have the potential to be very small indeed, allowing a lot of computational power in a very small space. This is not just because of the fact that quanta are small; it is also because of the nature of the physical forces involved. The biggest stumbling block, in conventional hardware design, to greater and greater scales of component integration is not the size of the components, but the density of connections. Communication in classical computers is via wires, and as components get smaller, there is geometrically less surface area of the component to which one can attach connecting wires. Also, wires have to be insulated from each other, which takes up more space. We will see below that in an NSQC, not only are the components small, but they communicate, not with wires, but with *forces*. In a conventional computer or network, this communication would require wire connectivity between the relevant components, which would limit the scale of integration.

Furthermore, in NSQC's, the non-locality of quantum interactions means that this communication is instantaneous. This increase in speed may or may not be dramatic for the short distances involved in conventional ways of thinking of nano-scale integration. But one can imagine a macro-spatially extended array of NSQC processors, which could communicate instantaneously (or near-instantaneously if relativistic considerations demand such a restriction) across substantial distances.

### 1.3 Quantum learning

Carver Mead, a visionary of computer hardware design, has pointed out [Mead, 1989] that until lately, advances in hardware have focussed on issues of *scale*; smaller is better, smaller is faster. But he points out that although the brain's hardware is of much larger scale, and much slower than current computer hardware, the brain can perform computations far beyond our fastest supercomputers. His recommendation, then, for advances in hardware design, is to look at how the brain is organized, for inspiration to come up with new forms of computation, rather than just try to make the kinds of computer we have now faster and smaller. Neural network algorithms are part of this search for novel kinds of computation. A computer on the spatial and temporal scale of quanta is bound to have advantages over current hardware, but Mead's point still holds. So why not pursue *both* improvements in scale *and* alternative forms of computation? That is one of the reasons for looking at *quantum learning*; but perhaps there are qualitative advantages of quantum learning that are not just the simple addition of algorithmic and scale advantages.

For the discussion of the advantages of (especially non-superpositional) quantum computation on which I wish to focus, a more concrete notion of an NSQC will be helpful. Since I also wish to stress the further advantages of quantum learning over quantum computation that does not involve learning, the particular NSQC I will consider is an implementation of a neural network architecture.

## 2 A non-superpositional quantum neurocomputer

Hopfield [Hopfield, 1982] popularized the idea that any physical dynamical system could be constrained to serve as a neural network, with fixed points of the system acting as "memories", which could be recalled associatively, in a content-addressable manner. However, the learning algorithms availble for the attractor-based neural networks on which Hopfield focussed are not as powerful as those for feed-forward networks, for which Hopfield's observation also applies. It is for this reason, and because the feed-forward back-propagation network is the one most likely to be familiar to a general audience, that the exmaple here will be based on a feed-forward network architecture. The principles and insights, however, generalize to other kinds of neural network.

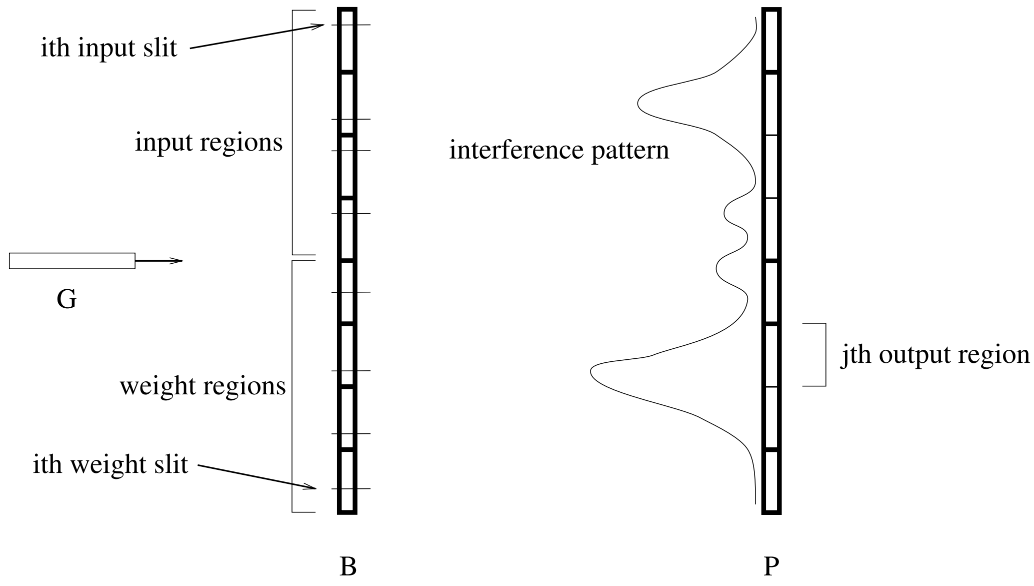


Figure 1: A feed-forward quantum neurocomputer. See text for key.

Similarly, many different quantum situations could be used to implement a learning algorithm; I will stick to the situation of a barrier with slits in front of a photo-sensitive plate, because of its familiarity. It is very likely that such a situation would prove impractical in order to obtain some of the computational advantages (especially those of speed and size) of quantum computing, but as such advantages are relatively obvious, and are not the advantages on which I will be concentrating, this impracticality need not concern us much.

The details of feed-forward networks involve units, activation values, layers, and weights. But what is important at first is that:

- feed-forward networks realize parameterized non-linear functions from an input space to an output space;
- networks modify these parameters in response to interaction with their environments (usually via a training set) so that the function realized by each network better approximates some intended function.

A quantum implementation of this kind of network is illustrated in figure 1. A particle beam $G$ is set up facing a barrier $B$, behind which is a photo-sensitive plate $P$. The barrier has several slits in it, which are responsible for the famous interference patterns on the plate. Some of the slits are designated to be *input slits*, and the rest are the *weight slits*. The interference pattern on the plate $P$ that results from the beam $G$ being aimed at the barrier $B$ with a particular input slit configuration is the output of the system for that input. Thus, use of the system to perform some computation will require two mappings: one ($I$) from inputs (e.g., character strings, images, queries, etc.) to input slit configurations, and one ($O$) from interference patterns to outputs (e.g., classifications, stored data, etc.).

Assume that one already has a mapping from inputs to $n$-vectors, and a mapping from $m$-vectors to outputs. One could map these input $n$-vectors to input slit configurations by, for example, dividing the barrier into $n$ sections, each having an input slit, with the relative position of the $i$th slit within the $i$th section of the barrier indicating the value of the $i$th coordinate of the input vector (this is the type of arrangement depicted in figure 1). For binary input vectors, this mapping could be even simpler: there is a slit in the $n$th section of the barrier if and only if the $n$th coordinate of the input vector is "1".

Less straightforward is the mapping $O$ from interference patterns to outputs. Since interference patterns are of high-dimensionality, any dimension-reducing mapping will do. Perhaps the plate could be divided into $m$ sections, and the average intensity within the $m$th region can serve as the $m$th coordinate of the output vector. For the case of binary output vectors, a soft (sigmoid) thresholding of this value would do (the thresholding must be soft so as to allow differentiation for back-propagation learning; see below).

The *error* of the system, $E$, is defined to be some function of the desired ($\vec{d}$) and actual ($\vec{a}$) output vectors, typically $\sum_i (d_i - a_i)^2$. If $S(\vec{x}, \vec{w}) = \vec{p}$ is the function that yields an interference pattern $\vec{p}$ given an input $\vec{x}$ and a weight slit configuration $\vec{w}$, then $\vec{a} = O(S(\vec{x}, \vec{w}))$.

Given some such setup, the system could learn an associative mapping $f$ in the following way:

- a number of samples $< \vec{x}, \vec{d} = f(\vec{x}) >$ are collected as input/output pairs for training;
- the system's weight slits are randomized;
- for each training sample $< \vec{x}, \vec{d} = f(\vec{x}) >$, the following occurs:
  - the input slits are initialized according to $I(\vec{x})$; the plate is cleared or replaced;
  - the beam is activated, until an interference pattern of sufficient resolution is produced on the plate, and an output is produced according to $\vec{a} = O(\vec{p})$;
  - for each weight slit $w_j$, the partial derivative of $E$ with respect to $w_j$ is estimated. This is done by calculating $\frac{\partial E}{\partial w_j}(\vec{d} - O(S(I(\vec{x}), \vec{w}))^2$;
  - this estimate is used to calculate the change to be made to the weight slit $w_j$ in such a way that gradient descent in error is achieved: the change in $w_j$ is proportional to the negative of the partial of $E$ with respect to $w_j$.

After several passes through the training set, this procedure will ensure that the system settles into a weight configuration that produces minimal error on the training set. In many cases, this will also result in good performance on other samples drawn from the same source as the training set (i.e., generalization).

This is enough to establish a correspondence between neural nets and a quantum system. The correspondence can then be used to suggest how the many variations on connectionist learning (e.g., recurrency, competitive learning, etc.) also could be implemented in a quantum system.

# 3 Implementation issues

## 3.1 Multi-layer quantum networks

It has been shown [Minsky and Papert, 1969] that, even though feed-forward networks might contain a non-linearity, they must have at least two layers of non-linear units if they are to be able to compute non-linearly-separable functions, which can be as simple as the function XOR (exclusive OR). Thus, for any reasonably powerful form of quantum learning, it might be better to think of a quantum beam/barrier/screen apparatus as implementing one unit in a layered network of units. Any units in a layer beyond the input layer would not have their input slits positions determined by the input sample $x$ directly, but rather by the outputs of the units in the previous layer. In such a case, it would be typical to take the output of a unit to be uni-dimensional, usually soft-thresholded.[1]

This would differ substantially from standard feed-forward networks, in that each weight would modulate all inputs to a unit. In standard networks, each weight only modulates the input from one other unit. This has the result of making the derivative computation during learning a relatively local computation. Whether or not this difference can be avoided, or whether it would be disadvantageous if it could not, has yet to be worked out.

### 3.2 Two-way quantum networks

One limitation of the scheme so far is that the quantum system really only implements the forward phase of the network. The back-propagation of error, or learning phase, must be calculated off-line, after which the slits are altered accordingly. It might be better if the learning phase were implemented directly in the quantum system as well, by having the desired interference patterns and actual interference patterns *directly cause* the changes in the weight slits.

Imagine a setup similar to the one already described, but with the following additions. Behind the plate, there is another particle beam, directed back toward the original barrier. Furthermore, there are slits in the plate, which will allow the second beam to pass through and hit the original barrier, which has a photo-sensitive plate mounted on its back. Thus, the second particle beam will cause interference patterns on the back side of the original barrier.

The goal would be to have the setup work like this: the plate itself would calculate the difference between the actual and desired interference patterns (perhaps by having something like the negative of the target pattern projected onto the plate), then this could cause certain slits to open in the plate, causing characteristic interference patterns on the back of the original barrier. These patterns would in turn cause the weight slits in the barrier to move according to whatever learning rule is being employed.

Of course, implementation-dependent speculations such as these may be premature, or irrelevant, since the principal reason for using both the barrier/slit/plate setup and back-propagating feed-forward networks was not for ease of implementation, but ease of explication.

### 3.3 Purely quantum networks

Perhaps it seems even more desirable to eliminate all macroscopic entities except those needed to fix inputs, and read outputs. That is, perhaps it would be better if the control variables and the mechanisms which manipulate them were not macroscopic slits, but themselves quantum

[1] Note that the function that the system realizes must lie between the two extremes of linearity and discontinuity: if the function is merely linear, then it lacks computational power (to avoid this, non-linearities may be introduced in between the barrier and plate); but if it is so non-linear as to be undifferentiable, the gradient cannot be followed during learning.

phenomena.

Perhaps not. Penrose makes some interesting comments [Penrose, 1989, p 403, 171-2] concerning the physics of computation that are relevant here. He argues that we can only have computers built out of macroscopic objects because of the discreteness of the quantum level; if there were no underlying discreteness, then there would be an unacceptable degradation of accuracy within any computational system. Furthermore, at least part of this discreteness is provided by the collapse of the superposed wave packet, so a purely quantum computer which does not have its superpositonal states collapsed now and again by macro objects, may be less powerful than a hybrid classical/quantum one.

On the other hand, this limitation of purely quantum systems may only be an impediment to traditional, von Neumann style computation. Neural networks, in that they are more robust and noise-tolerant, do not require as high a degree of accuracy, and thus quantum implementations of them may be able to function adequately without frequent "observations", which collapse the superpositional states.

### 3.4 Computing with individual quanta as opposed to aggregates

A issue related to the above is this: could one get more computational power by not using aggregates of quantum phenomena, but by using individual quantum events? In the network I have described, the mapping is from *ensembles* of quanta hitting the plate to outputs. One could instead imagine a faster-scale form of computation in which individual quanta hitting the plate are interpreted as outputs. As part of an inherently stochastic process, each quantum hitting the plate conveys not determinate information about the slit configuration, but probabilistic information: what the likely configuration of the slits is.[2] This kind of information may be used on its own during actual computation, but during learning, it seems most likely that many samples will have to be used in order for the network to learn the proper statistics, regardless of whether or not the weight changes occur after each quantum or only after an ensemble.

## 4 The impact on Physics, Brain & Mind

I'll consider the impact of quantum computation and quantum learning on the issues mentioned in the title of this collection in the order: physics, brain, mind.

### 4.1 Physics: The interpretation of quantum mechanics

Deutsch [Deutsch, 1985] mentions some ways in which superpositonal quantum computation can help us understand other physical phenomena:

**Complexity measure** Traditional computation-based complexity measures (e.g., the complexity of a string of digits is the length of the shortest computer program that can print that

[2] For example, suppose slit configuration $S$ produces an interference pattern that implies a zero probabilty for a quantum hitting at point $p$. Then a quantum actually hitting the plate at point $p$ tells us that the slits are not in configuration $S$.

string) have the problem that they classify noise as complex. The stochastic nature of quantum computation allows one to use it to provide a complexity measure that will classify noise as non-complex, since it can be generated by a very simple program on a quantum computer.

**Foundations** Deutsch suggests that this complexity measure could be used in the following way: one can postulate that the universe moves from the quantum-simple to the quantum-complex, and derive the third law of thermodynamics, and the psychological arrow of time from that.

**Experimentation** Given that a quantum computer is a true quantum system, one could program it so that its operation actually tests various physical hypotheses.

However, the aspect of the relevance of quantum computation to physics on which I wish to concentrate has to do with the interpretation of quantum mechanics. The standard interpretation of quantum phenomena is in terms of wave/particle duality: each quantum is a system with both particle and wave aspects. In the two slit experiment, for example, the particle-like aspect of the system is realized in, among other things, there being a hit on the plate at a particular point; at the same time, the wave-like aspect of the system is realized in the fact that the probability of there being a hit at any given point is [roughly] the value of the wave function at that point. These aspects are thought to be complementary: both are necessary, but in a sense they are also mutually exclusive.

There are other interpretations, however. The two to be mentioned briefly here are Everett's *many worlds* interpretation [Everett, 1983], in which the superpositional state is actually a superposition of universes, one for each possible value of the observable; and Bohm's *ontological* or *causal* interpretation [Bohm and Hiley, 1993], in which there are particles, but they are always accompanied by a new type of field (the *quantum field*), which in turn yields a potential (the *quantum potential*). The quantum field (and thus the quantum potential) is shaped by the experimental configuration, and thus can affect the trajectory of the particle in such a way as to generate the kind of behaviour (apparent indeterminacy, non- local sensitivity to slit configuration, etc.) that the standard interpretation attempts to explain in terms of the wave-like aspect of a quantum.

Deutsch sees quantum computation as implying the many worlds interpretation of quantum mechanics. He uses the many-worlds interpretation freely in explaining his ideas, and although he admits that these explanations could be re-formulated for other interpretations, he feels this can only be done with some loss of explanatory power.

On Deutsch's view, a superposition of $n$ quantum states can be used to perform parallel processing, although only one of the $n$ results will be accessible in a given world. Although the expected *mean* running time of a superpositional quantum computation is no better than a classical parallel version, Deutsch claims that some of the time the computation may take much less time than the fastest possible classical implementation. He reasons as follows: assume that a quantum computer has been set up to compute a task which classically takes at least two days; assume that there is a program that extracts the info from the superpositional state in negligible time, with a certain probability of success per unit time (per day, suppose). Then there is a non-zero probability that the information will be extracted from the superpositional state in just one day, faster than the classical limit. One can just check the halt bit to see if a two-day computation has occurred in one day. Deutsch uses the illustration of a Stock Exchange simulation program that predicts activity one day in advance, but classically takes two days to run; if run on a quantum computer, there will be lucky days where one manages to run the simulation in only one day, so one can actually use the predictions to invest successfully.

Deutsch suggests that the many-worlds interpretation is necessary for understanding such a program when he asks: "On the days when the computer succeeds in performing two processor-days of computation, how would the conventional interpretations explain the presence of the correct answer? *Where was it computed?*" [Deutsch, 1985, p 114, emphasis his]. I'm not convinced that quantum computation supports the many-worlds interpretation, mainly becaused I am not convinced that Deutsch's account of the Stock Exchange simulator is correct.

For one thing, the presence of a halt bit would seem to destroy the coherence of the state being used in the computation. As Deutsch himself points out, a quantum computer "must not be observed before the computation has ended, since this would, in general, alter its relative state" [Deutsch, 1985, p 104]. So he requires that there be a halt bit that can be observed, without affecting the operation of the quantum computer. But this seems paradoxical: if the halt bit depends on the computational state, then surely observing it will collapse the superpositional state, just as observing a light that indicates the presence of poison in Schrödinger's Box will either kill or save Schrödinger's Cat. On the other hand, if the halt bit *is* independent of the computation, then it isn't really a halt bit, any more than a flip of a coin would be: if the halt bit goes on, it can only be an accident that the machine has in fact halted. But even if the halt bit *could* indicate to one that the computation had been achieved in a shorter time than a classical computation, there seems to be no guarantee that this computational result is the one available in our world; the most we can know is that the correct result was obtained in *some* world. Whether that knowledge could be used in the way Deutsch intends has not been made clear.

Furthermore, even if one is convinced (as Penrose seem to be) that quantum parallelism can do the work that Deutsch claims it can, it seems that one can dispute (as Penrose does) Deutsch's claim that this argues in favour for the many-worlds interpretation [Penrose, 1989, p 401, fn 9].

Despite these disagreements, I am intrigued by Deutsch's explicit endorsement of the idea that the various interpretations of quantum mechanics can be distinguished experimentally. Whether or not the ontological interpretation is required in order to properly explain, think about and design quantum learning systems remains to be seen.

## 4.2 Brain: Quantum learning in real neural networks?

Another possible use of quantum implementations of neural networks might be as a way to understand what the *brain* is doing. The hypothesis would be that we might get a better idea of the function of some of the brain's features if we view them as implementing a quantum learning machine. But is there any evidence so far that the brain is sensitive to quantum effects? Not really. There is the well-known study [Baylor et al., 1979] that shows that a single photon striking the retina of a toad is sometimes sufficient to trigger a nerve impulse, but in humans, this phenomenon seems to be suppressed by noise filtering [Hecht et al., 1941].

But as Penrose [Penrose, 1989, p 400] points out, this does show that there are some cells in the human body that are sensitive to individual quanta, and therefore the possibility of quantum-mechanical effects in the brain is still tenable. But we would be making the task unnecessarily difficult if we, like Penrose, required that we find neurons that are sensitive to a single quantum. As discussed in §3.4, quantum computation can still occur in the cases where an aggregate of phenomena (an entire interference pattern, rather than one quantum) is required to yield an output. Many of the advantages of quantum computation would still apply in such a situation, such as communication via instantaneous forces, rather than wires.

If quantum learning networks are a more plausible model of brain activity than mere quantum computation, it may have little to do with the fact that the learning algorithm is currently called a *neural* network learning algorithm. Most likely, the bits of the brain that would correspond to the "neurons" in the algorithm would be sub-cellular phenomena. The extra advantage of quantum learning as a model of brain activity would rather derive from its sub-symbolism, and from the fact that it emphasizes learning.

## 4.3 Mind: Quantum learning and active information

One of the central obstacles to a complete, unified understanding of the world is the *mind-body problem*, which in recent years has been generalized to the problem of *naturalizing intentionality*: how can meaning or aboutness be seen to be part of the natural world? One traditional approach to solving this problem, one that is implicit in much work in cognitive science, is to close the gap by making the mind more like the physical, the mechanical. Thus, computers have played an important role. An alternative hope is this: perhaps we can make the problem easier, not by seeing the mind as physical, as mechanical; but by seeing the physical as having mental aspects, even (or especially) at the lowest (quantum) level.

This kind of idea has been advocated by Bohm [Bohm, 1990], and has been furthered by Pylkkänen [Pylkkänen, 1992]. Both are concerned with the idea of *active information*, the kind of information about its environment that a quantum particle purportedly carries. On Bohm's ontological interpretation, there is no traditional wave-particle duality; rather, there is a particle, with a determinate position and momentum, accompanied by a new type of field which gives rise to a new potential, the *quantum potential*. The interference patterns are caused by subtle variations in the quantum potential through which the particle moves, which in turn arises from the quantum field which is influenced by the experimental configuration (e.g., number and position of slits, etc.) It is claimed that the trajectory of the particle is such that the particle can be said to *carry information* about its environmental configuration. It is argued that the particle must *know* what the slit configuration is in order to avoid hitting points on the plate for which the wave mechanics indicate a zero probability of a hit. The particle hitting the plate at a place that has a zero probability of a hit under a particular slit configuration *means* that the slits are not in that configuration. And, the idea goes, if even a lowly particle can be seen to involve such mental phenomena as " meaning", "carrying information" and "knowing", then perhaps the physical/mental chasm can be crossed.

However, this idea has a fundamental difficulty: causal or statistical correlation is not the same as knowledge or having meaning. Otherwise, we would have to say that a broken window means something about the stone that smashed into it. And if we have to say that, then it looks like meaning is everywhere, in every physical interaction. But to say that is to water down the notion of "meaning" to the point where the chasm opens up again (to mix metaphors a little). For now it will be difficult to explain how the special case of meaning in cognitive systems arises out of the more inclusive, almost trivial notion of meaning as any kind of causal interaction.

However, both Bohm and Pylkkänen attempt to establish a difference between the effect of the quantum field and that of a classical field by claiming that the former, and not the latter, is dependent upon the field's *form*, rather than merely its *intensity*. Nevertheless, I do not see how this alone can save the notion of active information. Their explication still ignores the principal difference between our notion of the physical, and our notion of the mental: the mental is *normative*. Thoughts can be correct or incorrect, right or wrong, true or false. But the kind of information that quantum systems seem to have is of the everday, non-normative kind; quantum systems can't be *incorrect*. A quantum state "means" just whatever caused it; there is no room for falsity or error. Contrast this with thoughts: I might have a thought "That is a cow", that was caused by me seeing something across a field. As it happens, the thing that caused my thought was in fact a horse. This does not mean that my thought *means* that there was a horse there. Rather, it continues to mean that there was a cow there, and is therefore false. Unless we can make sense of such notions in a quantum system, then we will still have the large dualistic gap between mind and world. We will still wonder how a non-normative physical system can be the same as a normative mind.

Pylkkänen replies to this objection [Pylkkänen, 1992, p 96]:

> It would, however, seem that it is not possible to speak about misrepresentation (an important aspect of human intentionality) in connection with the quantum field as long as we have not been able to discover a sub-quantum level. But if there were such a sub-quantum level, it would then be in principle possible that processes in this level could sometimes interfere with the functioning of the quantum field, and "fool" the electron in[to] believing, say, that the second slit is open even if in fact it isn't. The electron would then "mistakenly" act as if both slits were open. This would require that there be some other way of giving the necessary form to the quantum field than what we now know to be possible (i.e. we can presently give the quantum field the form required for the two- slit behaviour only by having both slits open).

Even if the (seemingly desperate) request for belief in the existence of a sub-quantum level – a level for which we have no evidence – is granted, the above response must face the *disjunction problem* [Fodor, 1991]. Pylkkänen wants to say that the meaning of the quantum field is that there are two slits open, even though there is only one slit open, because some sub-quantum process $p$ affects the quantum field in such a way as to give it the shape that it usually has when two slits are open. But on what grounds do we single out this as the meaning of the quantum field? One could just as well say that the meaning of the field is a disjunction: either there are two slits open, or there is one slit open and $p$ is occurring. This meaning, which is just as valid as the one Pylkkänen favours, is not false. Thus we have no reason to believe that active information can misrepresent, and thus we have no reason to believe that it can help explain human intentionality.

This is where quantum learning might be able to help. Dretske [Dretske, 1986, p 35-6] has attempted to naturalize intentionality with the notion of learning. As said before, we can't get a notion of falsity going for a state if we just equate its meaning with whatever causes it. But suppose that we equate the meaning of a state $s$ with some $x$ which comes to cause $s$ during a *learning situation*, a situation in which a system could possibly learn a relationship between $x$ and some relevant behaviour. Then even if, later on, after the learning sitaution, some $y$ different from $x$ causes the state, the state will not mean $y$, because $y$ was not a cause of the state in a learning situation. A biological example: on this account, a rat's brain state $B$, typically caused by a bell, means, for a rat that has undergone conditioning, that there is food present because food caused (or shared a common cause with) the bell during the learning situation. So if, after the learning situation, a bell rings because someone hit it accidentally, and causes state $B$ (as is likely), then that state still means there is food present; it does not *mean* that someone hit it accidentally, even though that was the cause. Since there need not be food present in this case, we have the possibility of falsity. Therefore, a quantum learning system might similarly acquire some form of intentionality, and begin the bridging of the physical/mental gap.

A similar response to the disjunction problem might be given, that avoids all talk of learning. That is, all one has to do is give a principled way of determining which situations are meaning- fixing situations, and one can then solve the disjunction problem. My above solution, following Dretske, proposes that the meaning-fixing situations are learning situations. But another suggestion is this:

the meaning-fixing situations for a quantum system are the ones in which the quantum level gives a complete, correct characterization of the system. This would mean that the quantum field in Pylkkänen's example does indeed mean that both slits are open, since in the purely quantum cases, cases in which the sub-quantum does not interfere to yield a violation of quantum-mechanical laws, the quantum field in question is indeed the result of only one configuration: two slits being open. The situations in which the sub-quantum process $p$, together with one slit being open, yields that very same quantum field, is not a meaning-fixing situation, so the disjunctive meaning cannot be ascribed to the field. Thus, in the latter situations, the field is indeed false: it means something other than what is actually the case.

Although I have no direct rebuttal to this learning-free way of grounding quantum meaning, I do find it to be less satisfactory than my proposal involving learning. First, distinguishing meaning-fixing situations solely on the basis of whether quantum mechanics correctly describes the situation seems arbitrary and ad hoc. Learning has a non-arbitrary connection with intentionality and meaning; but why should respecting some quantum mechanical regularity be a source of normativity? We are left with a puzzle at least as great as the one with which we began. Furthermore, this principle could be applied at any level, resulting in an unacceptable ubiquity of meaning and intentionality. For if the quantum/sub-quantum reply is correct, then there would be no reason to reject a similar reply, one that appeals to a classical/quantum dichotomy. That is, one could say that all macroscopic events have meaning and intentionality as well: the meaning-fixing situations are ones in which classical physics gives a complete, correct account, and the possibility of misrepresentation is provided for by the possibility of situations involving macroscopic events that are influenced by the quantum level. It seems impossible to reject this proposal, and still support the quantum/sub-quantum proposal for grounding intentionality. A response based on learning, however, would not have such a problem, since it restricts meaning-provision to only those situations that involve learning.

Other approaches to naturalizing intentionality may suggest other forms of quantum computation for those who wish to see quantum systems as intentional. For example, the evolutionary approach, e.g. [Millikan, 1984], proposes that a state $s$ means that $p$ if $s$ is the product of a process of natural selection, and the explanation for why $s$ was selected for was because it was present when $p$ was true. Thus, a quantum implementation of *genetic algorithms*, with their computational version of natural selection, might be another way to get intentionality in at the most fundamental level.[3]

Once one has a notion of quantum information, one might use this show how information can be *implicit*, yet causally potent. The information that a particle has about its environment is not explicitly represented in the particle (there is no structure to the particle to provide an explicit representation), but it does have causal effect: it causes the particle to move in a particular way. Now one might think that one does not have to talk of information in this context at all; rather, we have action at a distance. The represented environment itself *directly causes* the action, so there is no need to invoke implicit information as a causal agent. But if we can make sense of a system being correct or incorrect, then we will not be able to say that what is represented is the direct cause, because the world might not be the way the particle's information takes it to be; the represented might not even *exist*. This idea of implicit but causally potent information has strong resonances with Bohm's idea of the *implicate order* [Bohm, 1980].

[3] Although there has not, to my knowledge, been any prior mention of quantum implementations of genetic algorithms, [Narayanan and Moore, 1995] introduces the idea of classical implementations of genetic algorithms that are quantum-inspired.

## 5 Acknowledgements

An earlier version of this paper was presented at the Brain, Mind & Physics workshop, Prague, September 14-17, 1993, organized by the Centre For Theoretical Studies at Charles University. I wish to thank Ivan Havel and CTS for making my participation there possible. That earlier version appeared as [Chrisley, 1995].

Many of the ideas in this paper were developed during very fruitful discussions with Paavo Pylkkänen, Larry Gould, Ivan Havel, Ajit Narayanan, and with the partcipants in the symposium on "Physicalism, Connectionism, and Consciousness" held in Siuntio, Finland in September of 1992. I am very grateful, then, for the support of the Finnish Academy and the University of Helsinki, which made these interactions possible.

## References

Baylor, D., Lamb, T., and Yau, K. (1979). Responses of retinal rods to single photons. *Journal of Physiology*, 288:613–34.

Bohm, D. (1980). *Wholeness and the Implicate Order.* Ark, London.

Bohm, D. (1990). A new theory of the relationship of mind and matter. *Philosophical Psychology*, 3(2):271–286.

Bohm, D. and Hiley, B. (1993). *The Undivided Universe.* Routledge, London.

Braunstein, S. (1995). Quantum computation: a tutorial. Unpublished draft available at http://chemphys.weizmann.ac.il/ schmuel/comp/comp.html.

Chrisley, R. (1995). Quantum learning. In Pylkkänen, P. and Pylkkö, P., editors, *New directions in cognitive science: Proceedings of the international symposium, Saariselka, 4-9 August 1995, Lapland, Finland*, pages 77–89, Helsinki. Finnish Association of Artificial Intelligence.

Deutsch, D. (1985). Quantum theory, the church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London A*, 400:97–117.

Dretske, F. (1986). Misrepresentation. In Bogdan, R., editor, *Belief.* Oxford University Press, Oxford.

Everett, H. (1983). 'Relative state' formulation of quantum mechanics. In Wheeler, J. and Zurek, W., editors, *Quantum Theory and Measurement.* Princeton University Press, Trenton.

Feynman, R. (1982). Simulating physics with computers. *International Journal of Theoretical Physics*, 21:467–488.

Fodor, J. (1991). *A Theory of Content and Other Essays.* MIT Press, Cambridge.

Hecht, S., Shlaer, S., and Pirenne, M. (1941). Energy, quanta and vision. *Journal of General Physiology*, 25:891–40.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, 9(2554).

Keyes, R. W. (1988). *IBM J. Res. Develop.*, 24.

Mead, C. (1989). *Analog VLSI and Neural Systems.* Addison Wesley, New York.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories.* MIT Press, Cambridge.

Minsky, M. and Papert, S. (1969). *Perceptrons.* MIT Press, Cambridge.

Narayanan, A. and Moore, M. (1995). Quantum-inspired genetic algorithms. Technical report, Department of Computer Science, University of Exeter.

Penrose, R. (1989). *The Emperor's New Mind.* Oxford University Press, Oxford.

Pylkkänen, P. (1992). *Mind, Matter and Active Information: The Relevance Of David Bohm's Interpretation of Quantum Theory to Cognitive Science.* Department of Philosophy of the University of Helsinki, Helsinki. Ph.D. thesis.

Shor, P. W. (1994). Algorithms for quantum computation: Discrete log and factoring. In Goldwasser, S., editor, *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, pages 124–134. IEEE Computer Society Press.