A Quantum Mechanical Approach to Cognition and Representation[1]

Joseph C.H. Chen (joseph@nats.informatik.uni-hamburg.de)
Computer Science Department, University of Hamburg
Vogt-Kölln-Strasse 30, D-22527 Hamburg, Germany

Abstract

In this article, a novel approach to cognition inspired by quantum mechanical principles is proposed. It is based on an analogy between the physical objects at the quantum level and the human's mental states. Specifically, symbols in various cognitive functions are to be treated as eigenstates with respect to a particular quantum experimental arrangement. The arrangement is associated with an *operator* (called a *formulation*) corresponding to an *observable* of quantum mechanics. A *state of affairs* is treated as a superposition of eigenstates. Based on these assumptions, the quantum mechanical formalism can be straightforwardly applied to cognitive processes. As in quantum mechanics, there is an analogous *Uncertainty Principle* in representation. Furthermore, it is shown that *non-monotonicity* and *counterfactual conditionals* can be accommodated in a quantum mechanical framework.

Introduction

Since the advent of modern natural science, any serious theory of mind in general and cognition in particular has to explain the fundamental observation that *there is significant* mesh *between physical brain and mental states.* Specifically, to answer the question, there are at least two *incompatible* philosophical positions — *physicalism* and *idealism / dualism*.

While dualism may account for many fundamental questions, it needs to fall back to "mysterious" arguments (e.g. God [Descartes, 1641]) earlier than physicalism. We will therefore concentrate our discussion to physicalism. Indeed, for many it seems unnecessary to argue for a distinct *mind* which is independent of the physical substrate in order to account for our high level mental abilities. For them, mind is nothing but activities of a physical brain. In this regard, many cognitive scientists are, implicitly or explicitly, *monistic physicalists*.

But this stance is not without question. For one thing, the modern natural science is strongly influenced by, if not strictly dependent on, the Cartesian-Newtonian world view. In this classical view, Nature is to be understood "objectively" and "coherently" by scientists. From here we can see an unfortunate consequence for classical physicalism, because the classical view of Nature *needs* an independent *subject* to observe and / or evaluate the *object*. Furthermore, the classical view takes Nature — this includes the human mind if a physicalist stance is taken — as an articulately designed but *passive machine*, qualitatively identical to a clockwork. It becomes therefore very implausible to accommodate some crucial human mental phenomena, such as *consciousness*, *intention*, *aboutness*, or *responsibility*, for these are excluded from Nature right at the beginning in the classical view. One way or another, one has to resort to the "magic" of "emergence" to prevent physicalism from collapsing. This irony is not surprising, however, because the facts on which naive monistic physicalism is based can be valid only if a dualist stance is taken.

The advent of quantum mechanics (see, e.g. [Dirac, 1958]) has profoundly changed the classical view in physics. Similarly, it may imply a profound change in physicalist cognitive science. In fact, it can be argued that the micro objects at the quantum level are prototypically *mind-like* and *active.* This, in a way, has alleviated the justified pessimism towards mechanistic cognition. After all, quantum mechanics is one of the most triumphant theories of physics, it has subsumed chemistry and eventually offered an account bridging fundamental natural phenomena to biology and the human body.

Nevertheless, the dilemma of dualism / monism haunts quantum mechanics, as it does the classical view. This is because the basis of quantum mechanics is dualist, if not totally idealist[2]. For many, it seems that the final *realist* substance, if any, in modern quantum mechanics can only be its mathematical formalism, which is an abstract language that begs for a logical mind. In fact, the unfortunate impact of quantum mechanics goes beyond that, since the physical reality to which the symbols refer is classically ill-defined. At best, one can predict the probability of the properties of a micro object and this probability depends on the experimental arrangement.

Indeed, most of the difficulties of quantum mechanics can be traced back to an intrinsic philosophical paradox that quantum mechanics needs classical objects that are free from the Uncertainty Principle for measuring the micro objects that obey this very principle. For some, the thorny paradox of quantum mechanics can be solved

[1]Poster prestened to *Consciousness and its Place in Nature, Toward a Science of Consciousness*, Skövde, Sweden, 7 - 11 August, 2001

[2]An example is the semi-orthodox Copenhagen interpretation of quantum mechanics.

only if its linguistic problem is nailed down and solved (e.g. see [Pyllkkö, 1997]). Interestingly, this also suggests that there may be a way to achieve this goal without betraying physicalism of cognitive science. This is done by making an *analogy* between language and quantum mechanics. This will be described in more detail in the following sections.

Quantum mechanical cognition

When we recognize something, or think of a matter and infer new results based on a state of affairs, we are mostly doing this using our *memory* of these subject matters (or "things") rather than using their direct sensorial data. In fact, it is very unlikely that these "raw" sensorial data are stored as verbatim neuronal activities. Rather, memory functions more like a system of compact *codes* — like a *language*. Whenever a state of affairs is needed for thinking, similar "raw" or "cooked" sensorial data have to be regenerated. It is therefore obvious why *representation* is placed at the center of modern cognitive science.

A physicalist may contend that all representations are to be physically "implemented" in the brain [Newell and Simon, 1976]. Acknowledging this, the topic is quickly shifted to the dispute over whether these representations are local and discrete symbols, as conventional artificial intelligence (AI) may conceive; or distributed and continuous activation, as connectionism may suggest. Despite this disagreement, there is a common denominator in both AI and connectionism: both regard representations as *classical* physical objects. In AI, they are like boxes that can contain other boxes or objects; in connectionism, they are activations on the neuronal membrane. Classical physical objects have *well-defined* properties and follow the law *passively*. This reveals how AI and connectionism may face the same difficulty in accommodating crucial mental phenomena such as intention, consciousness, and qualia.

Now if quantum mechanics is to be taken into account, we have to treat physical representations as quantum objects that *do not* have well-defined properties until a measurement is actually performed. *Heisenberg's Uncertainty Principle* must apply to representations in general.

At first sight, this seems to be a drawback. However, if we consider that in quantum mechanics the experiment outcomes depend on the arrangement of the measuring instruments, we can think of the measurement as a sort of *meaning*-giving procedure. Furthermore, if we assume that the experimental arrangement is associated with a *formulation* in a language-like system, we realize that *qualia* may be envisaged as an eigenstate that corresponds to a specific arrangement. In the following paragraphs, these ideas will be detailed with the help of the formalism of quantum mechanics.

Representation as a vector in a Hilbert space

The idea of quantum mechanical cognition is to treat a cognitive agent as a quantum physical system. Specifically, there is a symbol inventory of a cognitive agent that consists of *eigenstates* (symbols) of a particular *formulation* corresponding to an *observable* S (an Hermitian operator, i.e. $S^\dagger \equiv \{S^t\}^* = S$, where ^{t} and * are the transpose and complex conjugate operation, respectively). Furthermore, these eigenstates form a complete *orthonormal* basis of a Hilbert space. Generally speaking, a state of affairs (denoted as $|m\rangle$ [Dirac, 1958]) is a *superposition* of these basic kets,

$$|m\rangle = \sum c_n |s_n\rangle \tag{1}$$

where $c_n \in \mathbb{C}$ is a complex number with $c_n = \langle s_n \mid m\rangle$ being the *projection* of $|m\rangle$ on the eigenstate (or eigenket) $|s_n\rangle$; $|s_n\rangle$ fulfills the property: $S|s_n\rangle = \lambda_n |s_n\rangle$ with $\lambda_n \in \mathbb{R}$ being an *eigenvalue*. For a state of affair, the probability of finding a particular symbol corresponding to $|s_i\rangle$ is $P(s_i) = |c_i|^2/\Sigma_n |c_n|^2$. For convenience, we can normalize $|m\rangle$ so that it has always the length of unity. An *inference* based on a state of affairs can be then regarded as an undisturbed evolution of the system, which can be described by a *unitary* operator U (i.e. $U^\dagger = U^{-1}$). In a quantum system with constant energy and given an initial state of affairs $|\phi_0\rangle$, the state of affairs at time t is,

$$|\phi(t)\rangle = U\,|\phi_0\rangle = e^{-iHt/\hbar}\,|\phi_0\rangle\,, \tag{2}$$

where H is the *Hamiltonian* operator, which is Hermitian; $\hbar$ is the Planck constant divided by 2π. For brevity, we can write a state of affairs in terms of a complex-valued vector and an operator in terms of a complex-valued matrix. The goal is then to find a suitable Hamiltonian so that a particular reasoning can transform the initial state of affairs ($|\phi_0\rangle$) to the desired end state of affairs. The result is taken at an arbitrary but fixed time point. In this case, $\hbar$ and t can be absorbed into a matrix H'. That is, $U = e^{-iH'} = e^{-iHt/\hbar}$. For convenience and without losing generality, we will discuss H' in place of Hamiltonian and denote it as H hereafter.

Since H is Hermitian (i.e. $H^\dagger = H$), there are a total of n^2 free real parameters to be determined, provided the size of the vocabulary is n. The optimal parameters can be found using a standard optimization algorithm. We use the conjugate gradient method [Press et al., 1992] with the cost function defined as,

$$err(H) = \sum_{(\phi_t,\phi_i)\in T} \left|\left\langle \phi_t^k \,\middle|\, \phi_o^k \right\rangle\right|^2, \tag{3}$$

where T is a set of training pairs (ϕ_t, ϕ_i); $|\phi_t\rangle$ and $|\phi_i\rangle$ are the target and input state of affairs respectively. Moreover, $|\phi_o\rangle$ is related to $|\phi_i\rangle$ with $|\phi_o\rangle = U|\phi_i\rangle = e^{-iH}|\phi_i\rangle$.

The uncertainty principle

There is an important implication of our treating a symbol inventory as the eigenbasis of a particular language formulation operator. In fact, a Hilbert space can be decomposed in different ways. Suppose there are two operators S_1 and S_2, each of which can be used as the operator to decompose the space of states of affairs, S_1 and S_2 may not *commute*. That is, there may exist a state of affairs $|\phi\rangle$ such that $S_1 S_2 |\phi\rangle \neq S_2 S_1 |\phi\rangle$, or denoted as the *commutator* $[S_1, S_2] \equiv S_1 S_2 - S_2 S_1 \neq 0$. According to quantum mechanics, we must have $[S_1, S_2] = \pm i\hbar$. Then we have, using straight-forward application of Schwartz's inequality,

$$\Delta S_1 \Delta S_2 \geq \hbar/2. \tag{4}$$

This is the *uncertainty principle* of language formulation.

A corollary of Equation 4 is an uncertainty relation between *symbols* (in spoken or written language) and *concepts* (as the "real-world referents" of symbols). This can be argued by treating the non-verbal *memory*, whatever it may be, as a patterned representation system embedded in the neuronal quantum mechanical substrate.

In the following sections, preliminary applications of quantum mechanical principles on *non-monotonic* and *counterfactual* reasoning will be presented.

Non-monotonic reasoning

Suppose I had an appointment at 8 a.m. on a certain Sunday and I took a look at my watch. The scenario is: (1) *my watch shows 7:30, so I still have enough time.* However, on my way to the meeting I quickly learned that it was the Sunday when daylight-saving time takes effect, so I was certainly too late for the appointment. The scenario changes to: (2) *my watch shows 7:30* and *today is the first day of DST, so I am too late.*

In formal terms, asserting both arguments as sound is to assert that the following formula is true:

$$(p \rightarrow q) \wedge (p \wedge r \rightarrow \neg q) \tag{5}$$

where p is the proposition "my watch shows 7:30;" q, "my having enough time to be on time;" r, "it is the first day of DST." This is obviously incompatible with classical logic. Specifically, a non-monotonic reasoning is:

$$\mathfrak{C} \vdash \phi \text{ and } \mathfrak{C} \cup \psi \vdash \neg\phi \tag{6}$$

where $\neg\phi$ is the negation of a statement ϕ; $\mathfrak{C}$ is a collection of premises; ψ is an additional premise. In non-monotonic reasoning, additional knowledge may *falsify* some facts that could have been derived without the additional knowledge. In the classical context, this simply indicates that $\mathfrak{C}$ and / or the newly formed premises collection $\{\mathfrak{C}, \psi\}$ is *inconsistent.*

Non-monotonicity is a well-known problem and has been discussed by many. For instance, if $\vdash$ is used to denote a (formal) derivation in a statistical "knowledge-base" approach, the classical consistency could be saved because statistics (as a branch of monotonic mathematics) is itself consistent. For instance, the "truthfulness" of argument (1) can be transformed into a formula of conditional probability: $P(q|p) = P(q, p)/P(p)$; and that of argument (2): $P(q|p, r) = P(q, p, r)/P(p, r)$, where $P(\cdot)$ is the probability of proposition(s) $\cdot$ being true. Thus the apparent contradiction to classical logic can be technically explained away. Unfortunately, a statistical approach like this cannot account for all types of non-monotonicity. This can be shown by considering the following *non-monotonicity of strong kind.*

Definition 1 *Non-monotonicity of strong kind is a reasoning process in which ψ is statistically* independent *of the original premises collection $\mathfrak{C}$ and Equation 6 holds.*

If a simplified scenario described by Equation 5 is a case of non-monotonicity of strong kind, we have,

$$P(q|p, r) = \frac{P(q, p, r)}{P(p, r)} = \frac{P(q, p)P(r)}{P(p)P(r)} = \frac{P(q, p)}{P(p)} = P(q|p).$$

This indicates that the conclusion of classical logic should hold, because $P(q|p, r) = P(q|p)$ means that both assertions have the same "truthfulness."

Intuitively, the statistical independence of ψ and $\mathfrak{C}$ indicates that they are "compatible" and consistent. So in the non-monotonic reasoning of strong kind the newly introduced knowledge may *actively* change the reasoning structure. As a consequence, *novel* facts can be derived and / or established facts may be falsified.

Indeed, according to modern physics, these situations are everywhere in Nature. For example, in the electron two-slit experiment [Feynman et al., 1963], the consistent and independent additional knowledge (about through which slit an electron has passed) *changes* the experimental results. These situations prevent atoms, out of which we are made, from collapsing. In fact, we may find even *more* instances of non-monotonicity of strong kind in everyday life, where our knowledge may change the outcomes *actively*. These situations are those in which the *intention* of the reasoner plays a crucial role.

With these justifications, we can now see what quantum mechanics has to say about non-monotonicity as in the situation described above. Specifically, an inference is treated as a quantum mechanical experiment setup with a total of six eigenstates ($|p-\rangle$, $|p+\rangle$, $|q-\rangle$, $|q+\rangle$, $|r-\rangle$, and $|r+\rangle$), in which the plus sign ($+$) following a proposition symbol indicates that the proposition is asserted while a minus sign ($-$) indicates that it is refuted.

Classically speaking, for each proposition, there can be a third situation (*unknown* or X) where the proposition is neither asserted nor refuted. In fact, an unknown status of a situation σ is nevertheless a state *known* at a higher level. An agent has to assert the unknown status of σ. So he can consider the consequence based on this *explicit unknown* status. If, however, he does not even know that he does not know σ, he cannot consider the consequence of the unknown status as such. This is an important source of non-monotonicity. That said, the following training data should be considered as taken from a spatial vantage point of *another* observer or a temporal vantage point of the reasoner, which should not be confused with the explicit unknown status of the reasoner. These arguments are listed in the following table,

p	r	q	p	r	q
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{T}$	X	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	X
$\mathbb{F}$	$\mathbb{F}$	X	$\mathbb{F}$	X	X
X	$\mathbb{T}$	X	X	$\mathbb{F}$	X

In fact, non-monotonicity appears only when I learned that it was the first day of DST and *I had not* noticed that fact. In essence, the above table may eliminate non-monotonicity. But this is true only if the agent considers the issue from a temporal vantage point. And this is what the reasoner *cannot* afford during the inference. In this regard, the above table should be regarded as a "classicization" of non-monotonic reasoning. The ninth possibility $(p, r) = (X, X)$ is excluded from the table, because this situation is represented by a zero vector which always yields a null output (it is correct, though).

In a quantum mechanical framework, it is easy to express the unknown status of a proposition *without* introducing an artificial unknown status. This can be done by simply leaving out both the eigenstates pertaining to this proposition. Specifically, an input state of affairs is prepared according to Equation 1 and subject to a unitary *inference* operator U. The architecture is trained with the states of affairs listed in the table above. These states of affairs are prepared with phases (arguments of complex components) being zero.

In typical experiments, the goal can be achieved within a margin of 3% contingent fluctuation. That is, in three out of a hundred tests, the system gives a "wrong" answer owing to the statistical nature of quantum mechanics. If a threshold is applied to the output ensemble, an accuracy of 100% can be achieved. In short, a quantum mechanical architecture "implements" a prototypical everyday "classicalized" non-monotonic reasoning.

Nevertheless, a quantum mechanical architecture is richer in properties. For instance, it is common in everyday reasoning that we cannot be sure of how "true" or how "false" the antecedents are. This situation can be easily represented by a mixed state of affairs. For example, if proposition p is known to be true but r is *refuted* to a certain degree, certain conclusions can still be drawn. Specifically, the input state can be prepared as follows,

$$\frac{|p+\rangle + \rho e^{i\theta}|r-\rangle}{\sqrt{1+\rho^2}},$$

where $\rho \in \mathbb{R}$ with $0 < \rho \leq 1$. The deviation to the targeted output is shown in Figure 1. As can be seen in the figure, proposition q remains largely asserted if the phase of p and that of r are far away from π. However, if the phase difference of p and r happens to be near π and ρ is near 1, the output is flipped ($q \approx \mathbb{F}$).

Similarly, if proposition p is known to be true but r is *asserted* to a certain degree, the input state can be prepared as follows,

$$\frac{|p+\rangle + \rho e^{i\theta}|r+\rangle}{\sqrt{1+\rho^2}},$$

where $\rho \in \mathbb{R}$ with $0 < \rho \leq 1$. If the situation is correspondingly prepared, the relative probability of the output state is shown in Figure 2. It is *not* the complementary picture of Figure 1. This should not come as a surprise, for in many everyday arguments, we do not seem to treat "being refuted to a certain degree" as the logical

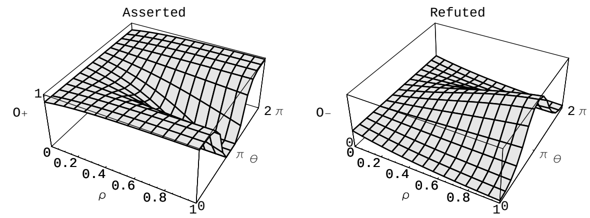

Figure 1: Relationship between the output and the argument (θ) / absolute value (ρ) of the refuted second antecedent.

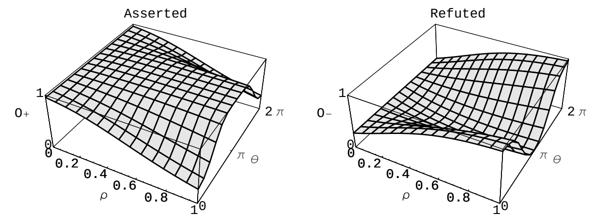

Figure 2: Relationship between the output and the argument (θ) / absolute value (ρ) of the asserted second antecedent.

complement of "being asserted to a certain degree," as a quasi-classical statistical framework may suggest.

Another example is when p is known to be true but r is at the same time asserted and refuted to a certain degree. To pursue this issue further, the input is prepared as

$$\frac{|p+\rangle + \rho e^{i\theta}|r+\rangle + (1-\rho)e^{i\theta}|r-\rangle}{\sqrt{1+\rho^2+(1-\rho)^2}},$$

where $\rho \in \mathbb{R}$ with $0 < \rho \leq 1$. The deviation to the targeted output is shown in Figure 3. This figure shows the complexity of this situation. All in all, it seems that the phase plays an important role. A conjecture on the meaning of phases will be provided after we discuss counterfactual conditionals in the following section.

Counterfactual conditional

Counterfactual reasoning is a thorny problem that interests many logicians (cf. [Lewis, 1986], for example). Roughly speaking, counterfactual reasoning is

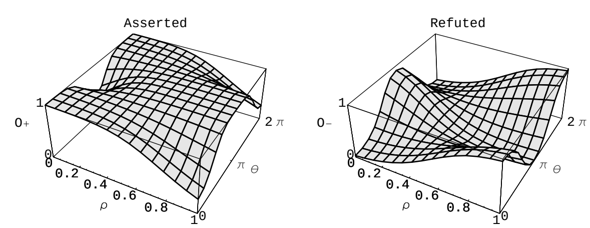

Figure 3: Relationship between the output and the argument (θ) / absolute value (ρ) of the second antecedent being at the same time asserted and refuted.

about drawing conclusion based on antecedents that are *not* (or not yet) actual. This definition is somewhat misleading, however, since many seemingly counterfactual reasonings such as "if I took this cyanide compound, I would die" are problems of knowledge-based decision-making and can be tackled accordingly. Nevertheless, there are counterfactual reasonings that are much more problematic. For example, consider the following (cf. [Barwise, 1989]): Jack and Jim are old friends. Under normal circumstances they will help each other. But Jim is very proud, so he will never ask for help from someone with whom he had recently quarreled. Jack, on the other hand, is very unforgiving. So he will never help someone with whom he just had a quarrel. Now Jack and Jim had a quarrel. A problematic counterfactual statement is:

Example 1 *If Jim had asked Jack for help, Jack would have helped him.*

Now if they had not had a quarrel, Jim would have asked Jack for help and Jack would have helped him. So the statement is true. But if Jim were not proud, Jack wouldn't have helped him since Jack is unforgiving. So the statement is false. This statement can be both true and false! Although the situation appears familiar in an everyday scenario, it is very difficult to find a coherent view of it. In the following paragraphs, we will try to construct a quantum mechanical model of this scenario.

Let p be the proposition "Jim is very proud," q be "Jack is very unforgiving," and r be "Jim and Jack had a quarrel." Our goal is a unitary *inference* operator U, with which an unequivocal answer s (whether Jack helps Jim) can be delivered under unambiguous circumstances. In the experiment, each proposition is associated with one *assertion* eigenstate and one *refutation* eigenstate, which are denoted by a plus sign $(+)$ or a minus sign $(-)$ attached to the proposition symbols respectively. Furthermore, we assume $\{p\pm, q\pm, r\pm, s\pm\}$ is a complete eigenbasis for representing the states of affairs. Any input state of affairs can be represented according to Equation 1.

The training set is constructed according to a simplified classical model of *possible worlds* [Lewis, 1986] and summarized in the following table. Specifically, we construct a set of situations where the inference rule can be unquestionably applied and may lead to coherent answers. The architecture is then trained according to the scheme as in the previous section.

p	q	r	s	p	q	r	s
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$

Among all eight combinations of situations, there are two questionable ones which are not included in the above table: (1) *Jim is very proud and Jack is* not *unforgiving and they had a quarrel.* (2) *Jim is very proud and Jack is unforgiving and they had a quarrel.* The first questionable assertion is in fact not very problematic, for it leads to an identical conclusion. We will concentrate on the experimental results of the second situation.

In typical experiments, the input states of affairs are prepared with phases being zero. The training goal can be achieved. Roughly speaking, a quantum mechanical architecture has acquired the "common sense" based on its experience of coherent day-to-day situations. We then constructed a state of affairs corresponding to the questionable situation and subjected it to the trained *inference* operator, with which we find out that the absolute square of the assertion-component is 0.24. That is, among repeating quantum measurements, about one fourth of the cases come up with s being true — the outcomes jump back and forth between true and false. Moreover, we found out that the phases of p and q play an important part. To show this, the inputs are prepared as follows:

$$\frac{e^{i\theta_1}|p+\rangle}{\sqrt{3}} + \frac{e^{i\theta_2}|q+\rangle}{\sqrt{3}} + \frac{|r+\rangle}{\sqrt{3}}.$$

The corresponding assertion state of s is shown in Figure 4 left. As can be seen in the figure, if the phase of $|q+\rangle$ is near π (relative to the phase of p and r), s is almost always asserted. This phenomenon seems puzzling at first sight. However, if we take a phase difference of two asserting eigenstates as a measure of "relevance" of these two propositions, we may regard $\theta_2 = \pi$ as indicating that q is "irrelevant" to s. Thus we have an explanation of why s is almost always asserted: if q is taken as irrelevant to the state of affairs, whether Jack is unforgiving will no longer play a crucial part in determining whether Jack helps Jim in a counterfactual situation (i.e. that Jim were *not* proud and that they had not had a quarrel). This hypothesis seems to offer an account for the graphics in the previous section, in which an irrelevant proposition can flip the outcomes.

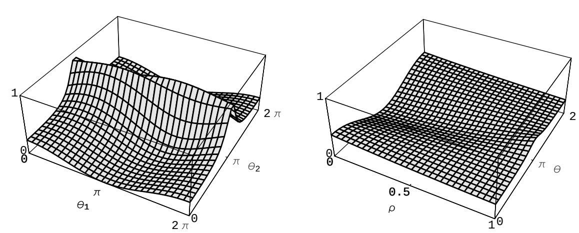

Figure 4: Left: The probability of s being asserted when $(p, q, r) = (\mathbb{T}, \mathbb{T}, \mathbb{T})$. The input states are prepared with different phases (θ_1, θ_2). Right: The probability of s being asserted when p is partially asserted.

Additionally, there are situations in which Jim and Jack did have a quarrel but the fact that Jim is proud is asserted to a certain degree and so is the fact that Jack is unforgiving. Specifically, the input is prepared as,

$$\frac{\rho e^{i\theta}|p-\rangle + (1-\rho)e^{i\theta}|p+\rangle + |q+\rangle + |r+\rangle}{\sqrt{2+\rho^2+(1-\rho)^2}},$$

where $\rho \in \mathbb{R}$ with $0 < \rho \leq 1$. The corresponding assertion state of s is shown in Figure 4 right. It shows that the “refutation-degree” of whether Jim is proud seems to have little influence on proposition *s*. This agrees with our intuition, for whether Jim had asked for help does not pretty much influence whether Jack would help him (Jack is unforgiving, so he will not help Jim anyway).

Alternatively, the input can be prepared as,

$$\frac{\rho e^{i\theta}|q-\rangle + (1-\rho)e^{i\theta}|q+\rangle + |p+\rangle + |r+\rangle}{\sqrt{2+\rho^2+(1-\rho)^2}},$$

where $\rho \in \mathbb{R}$ with $0 < \rho \leq 1$. The corresponding assertion state of *s* is shown in Figure 5 left. As can be seen in this figure, *p* depends heavily on both the “refutation-degree” (ρ) and phase (θ) of *q*. If the phase difference is kept small, the assertion of *s* is roughly a monotonous increasing function of ρ. This is not surprising. However, if the phase difference is about π, the degree of assertion becomes a very strange function of ρ. The detailed function when $\theta = \pi$ is illustrated in Figure 5 right. When $\rho = 0.258609$ there is a minimum. When $\rho = 0.343384$ there is a maximum of 0.998. At this moment, there is no intuitive explanation for this puzzling phenomenon.

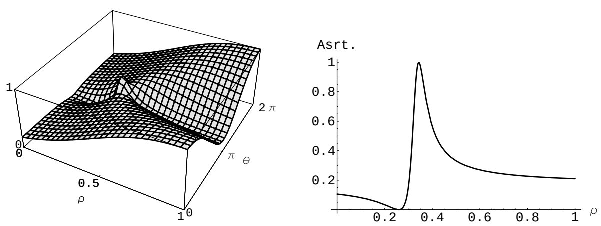

Figure 5: Left: The probability of proposition *s* being asserted when *q* is partially asserted. Right:The detailed relation between ρ and the assertion of *s* when $\theta = \pi$.

Discussion and conclusion

The treatment of cognition as a quantum system is motivated more or less by philosophical discontent with the classical mechanistic view of AI. In fact, although there are disputes over whether quantum effects should be taken into account in neurology [Hameroff, 1998], a system can still be built to show what effect this model may have. Indeed, the modeling described in this article shows that a quantum mechanical architecture can accommodate non-monotonic and counterfactual reasoning, albeit in a quite simplistic and miniature form. These are cognition problems which are very difficult, if not impossible, to account for with classical frameworks. Despite being quantum mechanical, this approach squares well with classical principles as far as the measurement outcomes are concerned. For one thing, the *law of exclusive middle* still holds, since either $|\mathbb{F}\rangle$ or $|\mathbb{T}\rangle$ (but not both) may manifest itself as output. In this sense, all quantum assertions (or refutations) are XOR-type assertions (or refutations), therefore two-valued. Strictly speaking, there is *no* unknown state in quantum mechanics, only the *absence* of certain eigenstates.

Moreover, quantum mechanics offers an adequate account for *actuality* and *potentiality*, thanks to complex numbers and superpositions. In fact, the *modality* (*necessity* and / or *possibility*) of a situation is crucial for us to understand the world. Thus quantum mechanics may remedy an important drawback of classical AI, for modality is totally neglected or at best “emulated” by stochastic procedures in classical approaches. Most importantly, in the realm of possibilities, contradictory situations can coexist. In such schemes, we have a superposition with mutually contradictory states of affairs, each of which has a corresponding complex coefficient.

Another important advantage of the quantum mechanical approach to cognition is that it accommodates the physical substrate and mental activity in a unified framework. In this way the bottom-up *quantitative* and *continuous* physical properties (manifesting as complex numbers) can be bridged to *qualitative* and *discrete* (manifesting as eigenstates) computation or logic of higher level intelligence.

Finally, it may be worthwhile to mention that although the approach proposed in this paper shares a similar physicalist stance with connectionism, it points out that we should look at the neuronal hardware deeper at the quantum level [Hameroff, 1998] instead of staying with the approximated classical explanation (chemistry or classical electromagnetism). In fact, if language in general is to be treated as quantum phenomena, thorny problems of cognitive science or linguistics such as *intention*, *qualia*, or *meaning* can be better accommodated [Chen, 2001].

References

[Barwise, 1989] Barwise, J. (1989). *The Situation in Logic*. Cener for the Study of Language and Information, 333 Ravenswood Avenue, Menlo Park, CA 94025.

[Chen, 2001] Chen, J. C. (2001). *Quantum Computation and Natural Language Processing*. University of Hamburg, Germany. Ph.D. thesis in preparation.

[Descartes, 1641] Descartes, R. (1641). *Meditations on first philosophy*. 1986 Cambridge University Press, Cambridge [Cambridgeshire] ; New York.

[Dirac, 1958] Dirac, P. A. M. (1958). *The principles of quantum mechanics*. International series of monographs on physics. Clarendon Press, Oxford,, 4th edition.

[Feynman et al., 1963] Feynman, R. P., Leighton, R. B., and Sands, M. L. (1963). *The Feynman lectures on physics*. Addison-Wesley Pub. Co., Reading, Mass.

[Hameroff, 1998] Hameroff, S. (1998). Quantum computation in brain microtubules? the Penrose-Hameroff “Orch OR” model of consciousness. *Philosophical Transactions Royal Society London (A)*, (356):1869–1896.

[Lewis, 1986] Lewis, D. K. (1986). *Counterfactuals*. Basil Blackwell, Oxford, reprint with corrections edition.

[Newell and Simon, 1976] Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126.

[Press et al., 1992] Press, W. H., Teukolsky, S., Vetterling, W., and annery, B. F. (1992). *Numerical Recipes in C*. Cambridge University Press.

[Pyllkkö, 1997] Pyllkkö, P. (1997). Eliminative holism as a solution to Bohr's puzzle of two languages. In Pylkkänen, P., Pylkkö, P., and Hautamäki, A., editors, *Brain, mind and physics*, pages 140–156. IOS Press, Amsterdam.