City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Quantum Cognition and Decision Theories: A Tutorial

James M. Yearsley[a,b], Jerome R. Busemeyer[c]

[a]*Department of Psychology, Vanderbilt University, Nashville, TN 37240, USA*
[b]*Department of Psychology, School of Arts and Social Sciences, City University London, Whiskin Street, London EC1R 0JD, UK*
[c]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA*

Abstract

Models of cognition and decision making based on quantum theory have been the subject of much interest recently. Quantum theory provides an alternative probabilistic framework for modelling decision making compared with classical probability theory, and has been successfully used to address behaviour considered paradoxical or irrational from a classical point of view.

The purpose of this tutorial is to give an introduction to quantum models, with a particular emphasis on how to build these models in practice. Examples are provided by the study of order effects on judgments, and we will show how order effects arise from the structure of the theory. In particular, we show how to derive the recent discovery of a particular constraint on order effects implied by quantum models, called the Quantum Question (QQ) Equality, which does not appear to be derivable from alternative accounts, and which has been experimentally verified to high precision. However the general theory and methods of model construction we will describe are applicable to any quantum cognitive model. Our hope is that this tutorial will give researchers the confidence to construct simple quantum models of their own, particularly with a view to testing these against existing cognitive theories.

Keywords: quantum theory, order effects, similarity

1. Introduction

Models of decision making based on the mathematics of quantum theory have attracted a large amount of interest recently (Busemeyer and Bruza, 2014; Aerts, 2009; Mogiliansky et.al., 2009; Yukalov and Sornette, 2011; Khrennikov, 2010; Pothos and Busemeyer, 2013; Wang et al., 2013). These models have arisen in part as a response to the empirical challenges faced by 'rational' decision-making models, such as those based on Bayesian probability theory (such examples are mostly associated with the famous Tversky-Khanaman research tradition. See e.g. Tversky and Kahneman (1974); Chater et al. (2006).) These quantum models posit that, at least in some circumstances, human behaviour does not align well with classical probability theory or expected utility maximisation. However unlike, for example, the fast and frugal heuristics programme (see, e.g. Gigerenzer et al., 2011), quantum cognition aims not to do away with the idea of a formal structure underlying decision-making, but simply to replace the structure of classical probability theory with an alternative theory of probabilities. This new probability theory has features, such as context effects, interference effects and constructive judgments, which align well with psychological intuition about human decision-making. Initial research involving quantum models tended to focus mainly on explaining results previously seen as paradoxical from the point of view of classical probability theory, and there have been a number of successes in this area (Pothos and Busemeyer, 2013; Wang et al., 2013; Trueblood and Busemeyer, 2011; White et al., 2014; Pothos and Busemeyer, 2009; Aerts et al., 2013; Bruza et al., 2015; Blutner at al., 2013). More recently, the focus has switched to some extent to testing new predictions arising from quantum models, and designing better tests of quantum vs classical decision theories (Atmanspacher and Filk, 2010; Yearsley and Pothos, 2014, in preparation).

One key success of quantum models of cognition has been the treatment of question order effects (Moore, 2002). 'Order effects' here describes a phenomenon where, for example, given two particular questions, each with a number of possible responses, the expected distribution of responses to a particular question depends on whether it was asked first or second in the series. In other words, asking a prior question can influence the outcome of a subsequent one. We will explain in more detail below exactly how to characterise this effect.

As we shall see, order effects arise naturally in quantum theory, and thus they can be accounted for by quantum cognitive models. However what is more remarkable is that quantum theory also predicts particular constraints on the probabilities that can be generated by these models, most notably in the form of the Quantum Question (QQ) Equality (Wang and Busemeyer, 2013). These constraints seem to be extremely well satisfied in the data from real world experiments and surveys (Wang et al., 2014). Thus as well as being a natural application of quantum theory, question order effects also represent a striking empirical confirmation of the idea of using quantum theory to model decisions.

Although the mathematical machinery of quantum theory is not inherently more complex than that required by many other cognitive models, essentially linear algebra and a small amount of calculus, it is rather unfamiliar to most cognitive scientists. Our aim in this tutorial paper is therefore to introduce readers to the ideas and machinery of basic quantum theory, such that after

working their way through this tutorial readers will feel more confident making use of quantum models in their research.

As well as existing cognitive scientists, we hope this tutorial may find a secondary audience in those researchers who already have a background in quantum theory gained from studying the physical sciences, who are interested in the application of these ideas in social science [1]. To help these readers we have structured our discussion of the basic formalism of quantum theory in a way which should feel familiar to anyone who first encountered it in the context of the physical sciences (see for example Isham (1995) or the notes by Plenio available online). Hopefully this should enable those already familiar with quantum theory to quickly grasp how to apply their existing knowledge to the construction of cognitive models.

The material we will cover in this tutorial is essential background to any application of quantum theory in judgment and decision making. We will pay particular attention to two important but sometimes overlooked issues; first how exactly does one choose a particular framework of Hilbert space, basis vectors, initial state etc to suit the problem at hand, and what do these choices mean? Second, how are the various calculations actually carried out? Grasping both of these issues is essential for any student of the field, and we hope this tutorial will help researchers bridge the gap between reading about quantum models and actually constructing them for themselves.

There are a number of things we will not cover in this tutorial, which may be worth stating now. First, although we will mention it, we will not cover the dynamics of quantum systems in any detailed way; this is mainly a tutorial on quantum statics. Quantum dynamics are relatively simple to grasp once one understands the material in this tutorial. Second, we will not touch upon 'entanglement' or issues around quantum information. Finally, some advanced topics, such as CP-Maps and the full theory of POVMs will not be covered, as they are best learnt about once one is familiar with the basics. They will be covered in a subsequent tutorial (Yearsley, in preparation).

We will assume the reader has a good familiarity with linear algebra in the usual form of vectors, matrices etc, but for reference, and to set notation, we give a brief summary of some important ideas in an appendix.

The rest of this tutorial is structured as follows; in Section 2 we introduce the basic elements of quantum cognition. In Section 3 we then expand upon some points, with the aim of guiding readers through the process of constructing a quantum model in practice rather than in theory. In Section 4 we give a brief introduction to order effects in quantum theory, and in Section 5 we expand on this to include a derivation of the QQ Equality. In Section 6 we give a brief introduction to POVMs, which can be used to represent noisy or imperfect measurements, and in Section 7 we apply these in the setting of order effects, our goal being to see to what extent the QQ Equality generalises to the case of more realistic noisy measurements. In Section 8 we briefly talk about another application of quantum theory to modelling similarity judgments. We summarise

in Section 9. A number of mathematical details are contained in the Appendix.

2. The basics of quantum cognition

The aim of this section is to present the basic formalism of quantum cognition, including information about the state, the dynamics, and the description of measurements. Our goal here is to give a reasonably concise account of the essentials; in the next section we will return to each element in turn and ask in more detail what it means and how it may be specified for a particular model. We hope this format will make it easy for readers to grasp the essential structure of quantum models. All of the material in this section is standard, and we will not give references for individual results/definitions. For a more compete account see Isham (1995) or for an alternative description with a more cognitive focus see Busemeyer and Bruza (2014).

2.1. What is quantum cognition?

Quantum cognition is a framework for constructing cognitive models based on the mathematics of quantum probability theory, which is itself a mathematical framework for assigning probabilities to events, much like classical probability theory (for a full account see Busemeyer and Bruza (2014)). For a given event, usually thought of as the outcome of some judgment process, and specification of the decision maker by means of a cognitive state, quantum cognition gives a real number between 0 and 1 which is to be interpreted as the probability that the decision maker will make that particular choice. Quantum cognition also includes information about the set of possible dynamics, state transformations and measurements that can be performed on a system, although to a large extent this follows directly from the basic probabilistic structure.

In its most conservative form, quantum cognition is simply an algorithm for computing probabilities, without any claim to reflect the underlying way decisions are made in the brain. In this way of thinking, the success or otherwise of the approach is to be judged purely by the empirical success of its predictions. However steps are being taken towards viewing quantum cognitive theories as *process* models, that do reflect in some way the process of arriving at a given decision (Kvam et al., 2015). The attraction of quantum models in this case stems in part from the fact they have features, such as contextuality, interference effects, order effects and constructive judgments, to name a few, that seem to align well with the way we think human decision makers process information. However an important feature of quantum models of decision is that they generally do not assume the underlying processing that happens in the brain is represented by quantum physics. In this sense that are distinct from so called *quantum brain* models (Hameroff and Penrose, 1996; Tegmark, 2000; Hagan et al., 2002; McKemmish et al., 2009).

The ingredients of any quantum cognitive model are a space of possible thoughts/judgment outcomes etc with subspaces corresponding to particular beliefs/opinions/choices, together with a cognitive state that keeps track of a decision maker's current state of mind, and a dynamics that specifies how this state

---

[1] Indeed one of the present authors (JMY) has such a background.

of mind changes with time/evidence presentation etc. From these ingredients we can compute probabilities for any relevant judgment outcome.

In the following subsections we will introduce each element in turn.

2.2. The Space

The most basic ingredient in quantum cognition is a space of possible thoughts/beliefs/etc. This space is a finite dimensional Hilbert space[2], which is essentially just a vector space equipped with an inner product $\langle \cdot | \cdot \rangle$. In quantum theory the Hilbert spaces we consider are complex spaces, which means they can be spanned by vectors whose entries are complex numbers. It is usual to denote the Hilbert space by the symbol $\mathcal{H}$.

We will write the elements of $\mathcal{H}$ as $|u\rangle$. $|u\rangle$ is known as a 'ket', and is roughly analogous to a column vector in standard linear algebra. The inner product of two vectors in $\mathcal{H}$ is denoted as

$$\langle |v\rangle, |u\rangle \rangle \equiv \langle v|u\rangle \tag{1}$$

which is roughly analogous to the dot product of two vectors in linear algebra. The inner product induces a norm $|\cdot|$ on $\mathcal{H}$ via,

$$|\cdot| : \mathcal{H} \to \mathbb{R}$$
$$|u\rangle \to \sqrt{\langle u|u\rangle}. \tag{2}$$

Technically our Hilbert space is an *abstract* space, i.e. it can be spanned by many different possible sets of basis states. One important point about this is that the object, $|u\rangle$, while a 'vector' in the technical sense, is *not* equal to a column vector, i.e.

$$|u\rangle \neq \begin{pmatrix} a \\ b \end{pmatrix}. \tag{3}$$

The reason for this is that the right hand side of Eq.(3) is defined relative to a particular basis, whereas the left hand side is basis independent. We will often abuse notation to write a ket as equal to a column vector etc, but when we do so we will always take care to specify the basis in which we are computing the vector.

One particularly striking example of the difference between vectors in $\mathcal{H}$ and vectors/functions in a particular basis occurs when dealing with states of a particle moving along a line (Isham, 1995). Let $|\psi\rangle$ be the state of a particle in 1D. Then $\psi(x) = \langle x|\psi\rangle$ is the wave function of the particle in position space, and $\tilde{\psi}(p) = \langle p|\psi\rangle$ is the wave function in momentum space. $\psi(x)$ and $\tilde{\psi}(p)$ are equivalent representations of $|\psi\rangle$ (one is the Fourier transform of the other, and they carry the same information), but they live in different spaces and have different units, so they are certainly not equal, and neither are they equal to the abstract vector $|\psi\rangle$.

An important concept is the idea of an operator on a Hilbert space. An operator on a space is a map which sends an element of the space to another element of the same space. e.g.,

$$A : \mathcal{H} \to \mathcal{H}$$
$$|u\rangle \to A |u\rangle \tag{4}$$

In a particular basis, the operator A will be represented by a matrix. Note though that in a similar way to states, an operator is not *equal* to a matrix, since the particular matrix depends on the choice of basis.

An important property of operators in Hilbert space is that they need not *commute*. Let A and B be two operators on $\mathcal{H}$. The *commutator* $[A, B]$ is defined as,

$$[\cdot, \cdot] : \mathcal{H} \times \mathcal{H} \to \mathcal{H}$$
$$[A, B] \to AB - BA \tag{5}$$

i.e. the commutator represents the difference between the two possible orders in which A and B could be applied to a state. A and B are said to *commute* if $[A, B] = 0$. Otherwise they are said to be *non-commuting*. That many important operators do not commute with each other, such as those representing position and momentum, is what lies behind many interesting features of quantum theory, such as the uncertainty principle.

A basic overview of operators on Hilbert spaces is given in the appendix.

Aside on Infinite Dimensional Spaces

In this tutorial we will be working mainly with finite dimensional vector spaces, such as are appropriate for modelling finite sets of possible question answers. However we may sometimes need to model situations where the variables can take any value in a continuous range, such as the position of a slider, or situations where the set of possible outcomes is so large it may be more convenient to think of them as continuous, for example confidence judgments on a 0-100% scale, where the corresponding finite dimensional representation has 101 dimensions. In these cases we can extend the formalism to a infinite dimensional Hilbert space. Instead of vectors our space is now spanned by complex valued functions. Most of what we say in this tutorial still holds, but the maths becomes slightly more complex. Interested readers should consult any good book on quantum theory (e.g. Isham, 1995), or for the full mathematical glory, see Reed and Simon (1980).

2.3. The Knowledge State

The second ingredient in any quantum model is the specification of the initial knowledge state of a participant, or group of participants. This is given in the simplest case by a vector $|\psi\rangle$ in $\mathcal{H}$. This state vector encodes information about the cognitive state of participants prior to any computation. This state can also be written as an operator on the Hilbert space by taking the outer product of $|\psi\rangle$ with itself, to give the *density operator*,

$$\rho = |\psi\rangle \langle \psi| \tag{6}$$

---

[2]The restriction to finite dimensions is not strictly necessary. The extension to infinite dimensional spaces such as those required to model motion in a continuous space requires only some extra mathematical precision. However the spaces most often used in quantum cognition are finite, so we will mostly restrict to this case in this tutorial. See below or Isham (1995); Reed and Simon (1980) for more.

The main advantage of using density matrices to represent knowledge states, rather than just vectors, is that it allows us to deal with situations where we have some (classical) uncertainty about the state. Suppose we have an ensemble of different groups of participants and each group can be characterised by the state vector $|\psi_i\rangle$ for $i = 1, 2...N$. Let the proportion of participants in each group be ω_i. Then the state of the whole ensemble can be written as,

$$\rho = \sum_i \omega_i |\psi_i\rangle \langle\psi_i| \tag{7}$$

It is cumbersome to represent this type of ensemble using state vectors alone.

Let us introduce some simple nomenclature. States that can be written in the form $\rho = |\psi\rangle \langle\psi|$ for a single $|\psi\rangle$, i.e. that have $\omega_i = 1$ for some i and $\omega_{j \neq i} = 0$ are called *pure states*. States that cannot be written in this form, i.e. that have $\omega_i \neq 0$ for more than one i, are called *mixed*.

The knowledge state plays a role roughly analogous to a prior probability distribution in classical theories. If the states $|\psi_i\rangle$ are orthonormal then the ω_i play the role of the probabilities for the initial state to be one of the $|\psi_i\rangle$. The following properties of ρ guarantee that the ω_i can be interpreted as probabilities in this way;

- $\rho = \rho^\dagger$. This guarantees the ω_i are real.

- $\text{Tr}(\rho) = 1$. This guarantees $\sum_i \omega_i = 1$.

- $\langle\psi| \rho |\psi\rangle \geq 0, \quad \forall |\psi\rangle \in \mathcal{H}$. This property, known as positivity, guarantees the ω_i are non-negative.

A particularly important example of a knowledge state is the case where the $\{|\psi_i\rangle\}$ form an orthonormal basis for $\mathcal{H}$ and $\omega_1 = \omega_2 \ldots = 1/D$, where D is the dimension of the space. The knowledge state can be written in this case as,

$$\rho = \frac{1}{D} \mathbb{1} \tag{8}$$

where $\mathbb{1}$ is the identity operator, i.e. the operator that has the property that, for any A, $\mathbb{1}A = A\mathbb{1} = A$. This state corresponds to a state of maximum (classical) uncertainty, and it plays a similar role to a uniform prior in classical models. Note that it is possible to define the notion of the 'entropy' of a quantum state, in a similar way to the entropy of a classical probability distribution. This is beyond the scope of this tutorial, see Neilsen and Chuang (2000) or Yearsley (in preparation) for details, however it can be shown that $\rho = \frac{1}{D} \mathbb{1}$ does indeed give the maximum value of the entropy, equal to $\ln(D)$.

We note that it is usual in quantum cognitive models to see the initial knowledge state represented as a pure state (e.g. in Busemeyer and Bruza (2014) density matrices are introduced only at the very end of the book). In the interests of completeness we will mainly work in this tutorial with density matrices, so that the initial state may be either pure or mixed. Where appropriate we will show how our expressions simplify in the case of pure states.

2.4. Observables and Measurements

The term *observable* in quantum theory refers to any quantity whose value can be measured by a suitable experiment. In quantum cognition observables are generally preferences or choices, such as whether a suspect if guilty or innocent, which of two gambles is preferred, or how confident a decision maker feels about a particular choice. The observables in a quantum cognitive theory are ultimately the same as they would be in a classical theory.

Observables in quantum theory are represented by *Hermitian* operators on $\mathcal{H}$. An operator is Hermitian if

$$A = A^\dagger \equiv (A^T)^* \tag{9}$$

The operation $\dagger$ is known as Hermitian conjugation. Hermitian operators are an important class because they have real eigenvalues, which makes sense, since any measurement must have a result that is a real number.

The expected value of the observable A for a participant with a cognitive state given by ρ is equal to,

$$\langle A \rangle = \text{Tr}(A\rho) = \langle\psi| A |\psi\rangle \tag{10}$$

where the final equality holds if the cognitive state is pure. However if we perform a measurement of A we do not get the result $\langle A \rangle$ in general. An important feature of quantum theory is that the possible results of a measurement of any observable are one of its eigenvalues.

Because the operators representing observables are self-adjoint, the spectral theorem for self adjoint operators (see Appendix) tells us that A may be written as,

$$A = \sum_i a_i P_i^A \tag{11}$$

where the a_i are the eigenvalues of A and P_i^A are the projection operators onto the subspaces corresponding to these eigenvalues. The probability of measuring A and getting the result a_i is given by,

$$p(a_i) = \text{Tr}(P_i^A \rho) = \langle\psi| P_i^A |\psi\rangle \tag{12}$$

where again the last equality holds if the cognitive state is pure.

A crucial feature of quantum theory is what happens to the state when this measurement is carried out. Suppose we measure the observable A and we get the answer a_i. Then the state *collapses* to the new state,

$$\rho \rightarrow \frac{P_i^A \rho P_i^A}{\text{Tr}(P_i^A \rho)} \tag{13}$$

or

$$|\psi\rangle \rightarrow \frac{P_i^A |\psi\rangle}{|P_i^A |\psi\rangle|} \tag{14}$$

if the state is pure. This is known as the collapse postulate in quantum theory and is responsible for much of the strange non-classical behaviour of the theory. Note that if we take the collapsed state and immediately preform another measurement of A then we will get the answer a_i again with certainty, since,

$$\text{Tr}\left(P_j^A \frac{P_i^A \rho P_i^A}{\text{Tr}(P_i^A \rho)} \right) = \delta_{ij} \tag{15}$$

4

This means that measurements represented by projection operators can also be used to prepare systems to be in a known state. These are often known as 'ideal first kind' measurements. That such operations can actually be carried out in practice is a critical assumption in quantum cognition[3].

It is worth saying a little more about the role of the collapse postulate, since anecdotally it seems to cause some concern amongst cognitive scientists. On the surface it looks a lot like Bayesian updating. Imagine that the initial state is a density operator ρ which can be written as a weighted sum of projection operators onto eigenvectors of a particular observable A, i.e.,

$$\rho = \sum_i p(a_i) |a_i\rangle \langle a_i| \tag{16}$$

with $\{p(a_i)\}$ a probability distribution over the possible a_i. Such a decomposition is not guaranteed to be possible, because although the $|a_i\rangle$ form an orthonormal basis ρ need not be diagonal in this basis. However suppose it is possible for the particular state under consideration. Then if we measure the observable A and get the result a_j then the state collapses to,

$$\rho \rightarrow |a_j\rangle \langle a_j| \tag{17}$$

In other words, if one thinks of the ω_i in Eq.(16) as defining a probability distribution over the set of states $|a_i\rangle$, then the collapse looks exactly like a Bayesian updating of this probability distribution.

The reason this interpretation doesn't work in general is that a Bayesian update happens when one gets some new information about the state as a result of performing a measurement. Because of this, a Bayesian update may, at least in some cases, decrease the uncertainty, or entropy, associated with our description of the system. In contrast, at least for pure quantum states, increasing our knowledge of an observable A will decrease our knowledge of any observable that doesn't commute with A. For pure states this means state collapse does not reduce our uncertainty about the state, in contrast to a classical Bayesian update. That being said, there is an approach to understanding the interpretation of quantum theory which emphasises the connection between collapse and Bayesian updating, called Quantum Bayesianism (See, for example, Caves et al. (2002).) We will not discuss this further here since the interpretation of quantum theory is beyond the scope of this tutorial. The important point to take away from this discussion is that although the collapse postulate is a fundamentally quantum feature, it does have some similarities with more familiar ideas in classical probability theory.

2.5. Evolution

We will not be especially concerned in this tutorial with the evolution of quantum states. Nevertheless, for the sake of completeness we note here the basics of quantum evolution.

For a closed quantum system, evolution is generated by an Hermitian operator H known as the Hamiltonian[4]. The evolution of a quantum state is governed by,

$$\rho(t) = U(t)\rho(0)U^\dagger(t) \tag{18}$$

or

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle \tag{19}$$

where, if H is time independent, $U(t)$ is given by,

$$U(t) = e^{-iHt} \tag{20}$$

The operator $U(t)$ is *unitary* which turns out to be an extremely important property. Unitary operators preserve the length of the state vector, or more generally the trace of the density operator, which means they conserve total probability. They also have an inverse, $U^{-1}(t) = U^\dagger(t) = e^{iHt}$, so that,

$$U^\dagger(t)U(t) = U(t)U^\dagger(t) = \mathbb{1}. \tag{21}$$

The reason this is important is the following: In practice we generate evolution of a cognitive state by the presentation of some stimuli. If we assume this evolution can be modelled by a Hamiltonian operator leading to a unitary evolution, then we also assume that this evolution is reversible. That means we are implicitly assuming there is some second set of stimuli that, if presented to a participant, could *undo* the change in cognitive state caused by the first set of stimuli (assuming no measurement intervenes). Whether this is reasonable or not depends on the experimental set up, but it is frequently not realistic.

The solution is to model the cognitive state not as a closed system, but as an *open* one, where information can flow between the cognitive system and some external system or environment. This allows for the possibility of *irreversible* evolutions not represented by unitary operators. It is also beyond the scope of this tutorial (Yearsley, in preparation). Concerns about reversibility notwithstanding, unitary evolutions are the most widely used in quantum models of decision, and have indeed proven to be very successful. However one should recognise that such a choice may not always be appropriate.

There are two further points we want to make about evolution. First we want to quickly discuss how to write expressions for the probabilities associated with measurements at different times, second we want to give two simple examples of useful Hamiltonians.

If we start with an initial cognitive state ρ_0, evolve for a time t and then perform a measurement of A, the probability we will get the outcome a_i is given by any of the following expressions,

$$p(a_i, t) = \text{Tr}(P_i^A U(t)\rho_0 U^\dagger(t)) \tag{22}$$
$$= \text{Tr}(P_i^A \rho(t)) \tag{23}$$
$$= \text{Tr}(P_i^A(t)\rho_0) \tag{24}$$

where,

$$\begin{aligned}\rho(t) &= U(t)\rho_0 U^\dagger(t) \\ P_i^A(t) &= U^\dagger(t)P_i^A U(t)\end{aligned} \tag{25}$$

---

[3]Although one can argue that it is a critical assumption in any theory of cognition, quantum or otherwise.

[4]Since H is Hermitian, it is an observable. In physics H corresponds to the *energy* of a system, however its significance in cognition is less clear.

Eq.(23) and Eq.(24) refer to the *Schrödinger* and *Heisenberg* pictures of evolution respectively. They differ in terms of whether one regards the evolution as changing the cognitive state, or the observables. For the most part we tend to work in the Schrödinger picture in quantum cognition, but both representations are equally valid.

Now suppose we measure A at time t_1, then we continue to evolve our state, and now we also measure B at time t_2. The probability that we get the outcome a_i at t_1 and b_j at t_2 is,

$$p(a_i, t_1; b_j, t_2) = \mathrm{Tr}(P_j^B U(t_2 - t_1) P_i^A U(t_1) \rho_0 U^\dagger(t_1) P_i^A U^\dagger(t_2 - t_1))$$
$$= \mathrm{Tr}(P_j^B(t_2) P_i^A(t_1) \rho_0 P_i^{\dagger A}(t_1)) \tag{26}$$

This can easily be extended to further evolutions and measurements. The Heisenberg picture of evolution comes into its own here.

We want to conclude this section on evolution with a couple of examples of useful Hamiltonians. The first is the Hamiltonian for a single particle evolving in some potential $V(x)$, such as might be appropriate for modelling a set of beliefs which could take any value in a given range. In the x basis we have,

$$H = -\frac{1}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{27}$$

Here m is some constant (the mass in physical terms) which gives rise to a sort of 'inertia' of our cognitive state, in other words it determines how quickly the state changes as a result of applying some force determined by $V(x)$. This sort of Hamiltonian has been used in studies of confidence judgments (Kvam et al., 2015). Good choices for the potential might be linear ($V(x) = ax + b$) which generates a constant force in a particular direction, or a quadratic ($V(x) = ax^2$) which produces a state which oscillates around $x = 0$.

Second, the Hamiltonian for a simple two state system can be written (like any other Hermitian operator on this space) as a weighted sum of the identity operator and the three Pauli matrices,

$$H = a\mathbb{1} + b\sigma_x + c\sigma_y + d\sigma_z \tag{28}$$

where

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{29}$$

and a, b, c, d are real numbers. Adding an overall scale to the numbers just speeds up or slows down the evolution, which is equivalent to rescaling the time. Thus only the ratios of the different coefficient are important. Now, the term involving the identity matrix just adds an overall phase to the state. Since the absolute phase is unobservable we can ignore this term. The terms involving $\sigma_x, \sigma_y, \sigma_z$ generate rotations about the x, y, z directions respectively, depending on the set up one might be able to further simplify this expression by considering rotation around only a single axis. Thus, for example, in Yearsley and Pothos (2014) it was argued that the Hamiltonian,

$$H = b\sigma_x \tag{30}$$

gives a reasonable general dynamics if one is interested only in the observable σ_z.

3. What does it all mean?

3.1. Introduction

In the previous section we introduced the basics of quantum theory, as applicable to constructing theories of decision making. In theory this gives all the information you need to go and study existing quantum models, and even in principle to build models of your own.

However much was left unsaid about the meaning of the various objects introduced. As a consequence, while it should now be obvious that, for example, to build a quantum model of order effects one first needs to identify the Hilbert space, it is probably less clear exactly how to do this in practice. What we want to do in this section is to fill in some of the gaps regarding the meaning of the various pieces of the quantum formalism. Our aim will be to give some clues about how to build a quantum model for a particular set of judgments.

There are two possible positions one might find oneself in when attempting to construct a quantum model of a particular decision making process. First we might have in mind some existing classical model, and want to explore a quantum analogue, either because the classical model is already empirically inadequate, or simply because one is interested in whether a quantum model might produce novel predictions. The second position is that we might have little idea about any existing treatment of the problem, and we have to start modelling from scratch. In practice we tend to encounter he latter more often than the former, but we will look at both scenarios here.

3.2. Quantisation?

Suppose we already have a classical model of a decision making problem, and we are interested in creating a quantum analogue. This is very similar to the typical situation in the physics literature, and physicists already have a reasonably straightforward process for constructing quantum models from classical ones. This process is known as 'quantisation' (The classic reference is Dirac, 1982), and we will describe it briefly.

It is easiest to take a specific example, so we will look at confidence judgments than can be expressed on a $0 - 100\%$ scale. We will assume the range of possible values is continuous, even if the set of possible responses is discrete. Assume we have some initial distribution of confidence levels, dynamics that might be generated by some evidence presentation, and then a judgment process that corresponds to splitting the scale into some intervals, e.g. $0\% \le x < 10\%$, or $20\% \le x < 30\%$ etc. The precise details are unimportant, but the key is the classical model gives us four things,

1. The space of states. In this case they are probability distributions (functions) in the interval $(0, 100)$.
2. The initial state. Or at least what that allowable class of initial states looks like.
3. The dynamics. Which we will assume can be written in terms of a classical Hamiltonian function.
4. The observables. In this case the probabilities of lying in various intervals of the form $[a, b]$.

Given these four features, we can now construct a fairly unique quantum analogue of this model. Let us explain how to do this[5],

- The Hilbert space is given by the set of complex valued functions on the interval (0,100), together with the condition that these functions are normalisable. Strictly $\mathcal{H} = \mathcal{L}^2([0, 100])$. These are just the infinite dimensional analogues of column vectors.

- The initial state is a normalised state $|\psi\rangle \in \mathcal{H}$.

- The dynamics follow from the classical Hamiltonian. Suppose we have $H = f(x, p)$ where x an p are the classical position and momentum of the state. A typical example might be a particle evolving in according to some potential $V(x)$ (representing the effect of evidence presented maybe) $H = p^2/2m + V(x)$. To get the equivalent quantum dynamics one makes the replacement[6] $x \rightarrow \hat{x}$, $p \rightarrow \hat{p}$, so $H \rightarrow \hat{H} = \hat{p}^2/2m + V(\hat{x})$. Here $\hat{p}$ is the operator in Hilbert spaces corresponding to momentum, in the x basis it is given by $-i\frac{\partial}{\partial x}$. The dynamics are then given by the Schrödinger equation,

$$-i\frac{\partial}{\partial t}|\psi\rangle = \hat{H}|\psi\rangle \tag{31}$$

- Finally, the observables are given by self adjoint operators on $\mathcal{H}$, but the most important one is simply $\hat{x}$, which is the operator corresponding to the confidence judgment. Typical judgments might be whether a participant's confidence was between $a\%$ and $b\%$, which corresponds to the projection operator,

$$\hat{P}_{[a,b]} = \int_a^b dx\,|x\rangle\langle x| \tag{32}$$

so that,

$$p(a, b) = \langle\psi|\,\hat{P}_{[a,b]}\,|\psi\rangle = \int_a^b dx|\psi(x)|^2 \tag{33}$$

So much for this ideal situation. In practice we are unlikely to have a concrete classical model from which to work, and we have to be considerably more inventive to set up our quantum models. We will explain how to do this below.

3.3. Building quantum models from scratch

In the last section we saw that if one has a well defined classical model of decision making, setting up a quantum analogue is a fairly straightforward process. However it is much more common in practice to approach a modelling problem with little or no prior idea about how to model it. In this case one cannot start

with an existing classical model, and we need to find a different way to proceed.

The difficulty we face is not so much that it is hard to build a quantum model, but that it is hard to constrain the possible choices. For simple models this is rarely a problem, but as the number of, say, responses grows so to does the number of possible ways of encoding these responses as vectors in a Hilbert space. For this reason what we will describe below is a reasonably algorithmic way to construct quantum models, but it may often be the case that the modeller has some extra information that suggests a particular structure to use. Remember that, provided a few simple rules are followed, it is hard to produce a model that is 'wrong' in the sense of not being a valid quantum system. Some degree of experimentation when building models is therefore fine, and may indeed be necessary to get a model which works well.

Here is a rough outline of how you might go about building a quantum model of a particular decision making process.

3.4. Identify the observables

Identify all the observables/measurements/judgments you wish to include in your model. An observable must be defined for any state. Decide which of them you expect to be able to define simultaneously, and for each set of simultaneously well defined observables decide how many distinct states there are. In other words, identify the commuting subsets of your observables, and the associated eigenstates. If all observables commute you have a classical model.

As an example, suppose we have three variables, A, α and B. Suppose A can take values a_1, a_2, α can take values α_1, α_2 and B can take values b_1, b_2, b_3, b_4. Suppose we expect to be able to define A and α simultaneously, but we don't expect to be able to define either together with B. So our two sets of commuting observables are $\{A, \alpha\}$ and $\{B\}$.

Now an important question is whether states with the same value of A might have different values of α or vice versa. If this isn't the case then A and α are simple functions of one another and that's not very interesting. Suppose instead that states with the same value of A can indeed have different values of α. Then there are four possible states defined by the set $\{A, \alpha\}$, let's denote them $\{|a_1, \alpha_1\rangle, |a_1, \alpha_2\rangle, |a_2, \alpha_1\rangle, |a_2, \alpha_2\rangle\}$, where e.g. state $|a_1, \alpha_1\rangle$ has value a_1 for observable A and value α_1 for observable α.

Once we have identified the sets of observables we can use these to determine $\mathcal{H}$.

Aside on degeneracy

One possibility we are ignoring here is that all the commuting subsets we generate may have degenerate eigenspaces.

It might happen that all eigenstates are degenerate because of some observable we have failed to account for. For example, as well as the operators A, α, B there may be some fourth operator C which commutes with all the other operators and gives multiple possible flavours for all existing states, e.g. $|b_1, c_1\rangle, |b_1, c_2\rangle$ .... This could happen if there is a mismatch between the way a judgment is elicited and the way that observable is represented cognitively, e.g. we may ask participants to

---

[5]To avoid confusion, in this subsection we use hats to denote operators.

[6]This hides a multitude of sins. If the classical Hamiltonian contains terms like $x^n p^m$ there is no unique quantisation, since classically this term does not depend on the order of x and p, whereas of course in quantum theory it does.

7

give preference judgments on a 1-9 scale when their preferences are actually encoded on a 0-100 scale in the brain.

Although it seems like this might be a problem, it is actually unlikely to cause difficulty as long as the dynamics do not contain interactions between C and other observables. However a good general rule is to choose your observables to match the way you think the quantities are encoded in the brain, and then if necessary represent the measurements in a coarse grained way as projection operators onto large dimensional subspaces.

3.4.1. Identifying $\mathcal{H}$

For the sets of simultaneously definable observables, find the set with the largest set of associated states. If two or more sets of commuting observables tie for the largest number of states then pick either, call this number D. Your Hilbert space then must have D complex dimensions, or in other words $\mathcal{H} = \mathbb{C}^D$.

To go back to our example, it turns out our two sets of commuting observables $\{A, \alpha\}$ and $\{B\}$ have equal numbers of distinct states, four each, so we can work with either. Since the set $\{B\}$ has four distinct states, we will need a 4D space for them to live in. This means we need a Hilbert space of four complex dimensions, so we have determined that $\mathcal{H} = \mathbb{C}^4$.

What we have done here is used the spectral theorem for Hermitian operators in reverse (see the appendix). We know that the eigenstates of an Hermitian operator span the Hilbert space in which it lives. Thus if we know that we have an observable with D distinct eigenvalues, that means it must live in a complex space with at least dimension D. The trick is then to look at our observables and find the one with the largest number of eigenvalues, this gives the minimum dimension of our Hilbert space. Care needs to be taken if we have commuting subsets of observables, like A and α above, since the number distinct states now depends on the degeneracy. (e.g. if A and α can each take one of three possible values, the number of vectors may vary between 3 and 9.)

So in our example $\mathcal{H} = \mathbb{C}^4$ and the states $\{|b_1\rangle, |b_2\rangle, |b_3\rangle, |b_4\rangle\}$ and $\{|a_1, \alpha_1\rangle, |a_1, \alpha_2\rangle, |a_2, \alpha_1\rangle, |a_2, \alpha_2\rangle\}$ form two orthonormal bases for $\mathcal{H}$.

3.4.2. The relationship between the observables

Now by construction the eigenstates of B and those of the pair $\{A, \alpha\}$ span $\mathcal{H}$, so we already know two different bases for our Hilbert space. The question is, what is the relationship between the sets of bases? Or in other words, we started by assuming that A and B could not be defined simultaneously, which means that they won't commute as operators. So the question is, what is their commutator?

Well the short answer is that we can't tell you! This is essentially where the maths ends and the psychologist's intuition has to take over. Quantum cognitive theories, like any other theory, require certain inputs from the modeller, and one of these is the relationship between the observable quantities. However although we can't tell you exactly how to proceed, we can give you some clues about how to implement different choices. One comes from the fact that since the two sets of vectors $\{|a_1, \alpha_1\rangle, |a_1, \alpha_2\rangle, |a_2, \alpha_1\rangle, |\alpha_2, \alpha_2\rangle\}$, and $\{|b_1\rangle, |b_2\rangle, |b_3\rangle, |b_4\rangle\}$ each form an orthonormal basis for $\mathcal{H}$,

they must be related to each other by a unitary transform. A unitary transformation in the space $\mathbb{C}^D$ has D^2 parameters, so that gives us the maximum number of parameters to be fixed. However this number will often be much lower in practice, because of simplifying assumptions we can make

Another is that we may sometimes be able to make simplifying assumptions about the relationship between the states. Suppose for example we are dealing with emotion states from two different points of view, perhaps two different people, so that our states are $\{|Happy\rangle, |Sad\rangle, |Excited\rangle, |Scared\rangle\}$ from one point of view, and $\{|Happy'\rangle, |Sad'\rangle, |Excited'\rangle, |Scared'\rangle\}$ from a different point of view. It might be that we have reason to believe that e.g. $|Happy'\rangle$ is a mixture of $|Happy\rangle$ and $|Sad\rangle$ and $|Excited'\rangle$ is a mixture of $|Excited\rangle$ and $|Scared\rangle$, but that the change of perspective doesn't mix e.g. $|Happy\rangle$ and $|Excited\rangle$. The unitary transformation can then be split into two separate transformations, each operating on a 2 dimensional subspace, and the number of parameters drops from $D^2 = 16$ to $2 * 2^2 = 8$. Another simplification occurs when we have a small number of observables compared with the dimension of $\mathcal{H}$. We may then be able to pick all our basis vectors to be real

A further special case is when we think the different bases are maximally mixed, in the sense that a basis state from one basis is a mixture of all basis states from the other basis, so that there is equal probability to measure this state to be in any of the basis states for the other basis. In other words, if we have two bases $\{|e_i\rangle\}$ and $\{|f_j\rangle\}$ then,

$$\left| \langle e_i | f_j \rangle \right|^2 = \frac{1}{D}, \quad \forall i, j \in 1, 2...D \tag{34}$$

If our two bases have this property then they are called *mutually unbiased* (mutually unbiased bases were introduced in, Schwinger (1960). For a more recent review see Bengtsson (2006)). There are only a fixed number of mutually unbiased bases in a given Hilbert space, which means only a fixed number of options for choosing $\{|f_j\rangle\}$ given $\{|e_i\rangle\}$. If the dimension of the Hilbert space, D, is a power of a prime number, then the number of mutually unbiased bases is $D + 1$. It is not known how many mutually unbiased bases there are for general D, but it is always less than or equal to $D + 1$ (Durt, 2010). Thus in this case there are only a finite set of possible choices for the basis vectors.

What exactly does the choice of relation between observables mean for a model? In a classical model all observable are *compatible*, which means they can be defined at the same time. In quantum models observables are generally not compatible, and they cannot be defined simultaneously. Assuming one wants to use a quantum model in the first place, then presumably one is looking for incompatibility. The relationship between observables in a particular model therefore determines two things; how much incompatibility one has, and whether the incompatibility occurs equally across all states of those observables.

As we will show for a specific example in Section 4, the degree of incompatibility, measured through non-commutation or order effects, depends on the overall size of the unitary rotation between bases for A and B. However if the unitary transformation does not effect all basis states equally, it might be the case

that, for example, some subset of the possible outcomes of A commutes with a subset of the possible outcomes of B, so that although the observables are technically incompatible this may not be the case when acting on all possible states. In general deciding on the relationship between a given set of observables is a hard problem, that needs a lot of psychological insight to solve successfully.

3.4.3. Identifying the initial state

The initial state is straightforward to identify, at least in theory. The initial state can be fixed from the probabilities for the outcomes of measurements. So if one knows the initial measurement probabilities, one can identify the state.

The problem is that there are $2D - 2$ free parameters in a pure quantum state, and $D^2 - 1$ in a general density operator. By contrast, a complete set of measurements can fix at most $D - 1$ parameters (see appendix for information on how to compute these). Thus to fix all the parameters of a pure state we need probabilities for at least two non-commuting observables, and we will need many more than that to fix a density operator.

However, this is almost never the way things are done in practice. Usually we set the initial state to be a particular eigenstate of one observable. How can we get away with this? The answer lies in recognising that because measurements change the state via the collapse postulate, measurements can be thought of a equivalent to state preparation (Peres, 1998).

Suppose we have a D dimensional Hilbert space with some unknown initial cognitive state $|\psi\rangle$ and an observable A with eigenvectors $|a_1\rangle, |a_2\rangle$ .... If we perform a measurement of the observable A on the initial cognitive state and get the result a_1, then we know that the cognitive state has now collapsed to $|a_1\rangle$. Thus we can use an initial measurement as a priming, to fix the cognitive state. Assuming not all participants give the same answer to this initial question, they will not all be in the same state afterwards. We can deal with this in one of two ways; if there is an extremely high probability of participants giving one answer, say a_1, then we can simply screen out any participant who answers otherwise. Our cognitive state is then $\rho = |a_1\rangle \langle a_1|$.

If, on the other hand, there are a spread of likely answers, we can either model each participant separately with a cognitive state corresponding to his/her initial answer, or we can use the density operator,

$$\rho = \sum_i \omega_i |a_i\rangle \langle a_i| \tag{35}$$

where the $\{\omega_i\}$ are equal to the probabilities that participants give each of the possible answers $\{a_i\}$ to the initial question. A special case of this is when there is equal likelihood of participants being in any of the $|a_i\rangle$, in which case the most appropriate initial state is the completely mixed state, corresponding to maximum ignorance, $\rho = \frac{1}{D}\mathbb{1}$.

We can then take this idea further. Suppose we have good reason to believe participants would, with extremely high probability, answer the initial question in a particular way. One example might be if our experimental set up was something like a jury trial, and the initial question was prior to the presentation of any evidence, since participants should overwhelmingly prefer innocent to guilty in the absence of any evidence. In this case since we are close to certain that participants would respond innocent if questioned, we can say their initial state must be very close to the 'innocent' eigenvector. However this rapidly becomes unreliable as our confidence in what the initial judgment would be drops from 100%. This is, in its simplest form, because it is impossible to tell the difference between a superposition of the form $\sqrt{a}|0\rangle + \sqrt{1-a}|1\rangle$ and a mixed state of the form $a|0\rangle\langle 0| + (1-a)|1\rangle\langle 1|$ by only measuring in the $\{|0\rangle, |1\rangle\}$ basis.

Nevertheless, the key is that the application of some quantum know how together with some psychological intuition can usually help us to identify the initial cognitive state in a straightforward manner.

3.4.4. Identifying the dynamics

We shall briefly talk though how the dynamics might be specified. In the same way as for the initial state above we shall begin with the technical story, and then we shall show how the problem can be simplified with some psychological intuition.

Evolution is generated by a Hermitian operator, the Hamiltonian. For finite dimensional systems this is an $D \times D$ Hermitian matrix, and thus has D^2 parameters. We can fix some of these parameters by using the transition probabilities,

$$p(i \to j; t) = |\langle j|U(t)|i\rangle|^2 \tag{36}$$

There are $D^2/2$ of these transition probabilities, so we either need them for two different bases, or for the same basis at two different times. In theory we know this information if we know the expected dynamics.

In practice, in much the same way as we could assume a simplified form for the initial state, we can probably assume a simplified form for the dynamics. A standard assumption is that the dynamics of a quantum system functions like a transformation between two known bases. e.g. it might rotate the states $\{|0\rangle, |1\rangle\}$ to the states $\{|+\rangle, |-\rangle\}$ after some total time T. This means,

$$U(T)|0\rangle = |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{37}$$

so we can take,

$$U(T) = \exp\left(-i\frac{\pi}{4}\sigma_y\right) \tag{38}$$

thus we conclude,

$$U(t) = \exp\left(-i\frac{\pi t}{4T}\sigma_y\right) \tag{39}$$

Again, we have seen that with some psychology intuition about the expected dynamics, we can fix the unitary operator controlling evolution fairly simply.

3.5. The interpretation of quantum theory

Despite all that we have said above, it is reasonable to feel like we have not really scratched the surface of what quantum theory *means* as a theory of cognition. What does superposition really mean? Why are certain variables incompatible and what does that mean for the way they are actually encoded in

9

the brain? If wave function collapse is not exactly Bayesian up-
dating then what is it? These questions about the meaning of
the theory are in some ways the most interesting and important.

Unfortunately we cannot cover these questions in this tuto-
rial. The reason is essentially twofold; first, the primary pur-
pose of this tutorial is to explain how to use quantum theory to
model certain effects. Furthermore, addressing the meaning of
the theory is generally best done once one has a good grasp of
the way the theory works in practice, and so it is more prof-
itable for the reader to return to this question once they have
more experience with how the theory is used.

Second, there is in fact no general agreement on how to in-
terpret quantum theory as a theory of cognition. The confu-
sion is partly inherited from physics, since the interpretation of
quantum theory in physics is hardly without controversy, but
it is also worsened by the fact that quantum theory is gener-
ally regarded as an *effective* theory of cognition, where the ac-
tually neuroprocessing may take place in an essentially classi-
cal way. Thus, in cognition, quantum theory is regarded as an
effective theory possibly arising from an underlying classical
process. For various reasons this cannot be the case in physics
(Isham, 1995; Bell, 2004), and thus it is not obvious whether
the interpretative framework associated with quantum theory
can be directly imported to quantum cognition in the same way
as the mathematics (see Blutner and beim Graben (2014); beim
Graben and Atmanspacher (2006) for some attempts to inter-
pret quantum cognition as arising directly from classical neuro-
physics). We hope that further research will help us to under-
stand why quantum theory works, and what it tells us about the
underlying process of cognition.

4. Order Effects I

The basic framework we have outlined so far can be used
to build quantum models for any particular set up one wishes.
However it is useful to look at specific examples to see how the
theory works in practice. For the rest of this tutorial we will
therefore focus on the particular example of order effects, but
bear in mind this is just one possible application of these ideas.

Order effects are a central feature of quantum theory because
of the non-commutation of operators. Recall; if A and B are
two operators on $\mathcal{H}$ then the commutator $[A, B]$ is defined as,

$$[\cdot, \cdot] : \mathcal{H} \times \mathcal{H} \to \mathcal{H}$$
$$[A, B] \to AB - BA \tag{40}$$

A and B are said to *commute* if $[A, B] = 0$. Otherwise they are
said to be non-commuting.

Most operators in quantum theory do not commute with each
other, and thus the order in which they are evaluated matters. If
the two operators represent two possible survey questions, for
example, then a lack of commutation means that the expected
answers these questions depends on the order in which they are
asked. Thus quantum theory can naturally incorporate order ef-
fects via non-commuting operators (Atmanspacher and Römer,
2012).

As an example, let us consider a two dimensional Hilbert
space spanned by $\{|0\rangle, |1\rangle\}$, and two projection operators given
by

$$P_0 = |0\rangle \langle 0|, \quad P_+ = \frac{1}{2}(|0\rangle + |1\rangle)(\langle 0| + \langle 1|). \tag{41}$$

A striking example of order effects in this case occurs when the
initial state is $|\psi\rangle = |1\rangle$. We have

$$p(+, 0) = |P_0 P_+ |1\rangle|^2 = |\frac{1}{2}(|0\rangle \langle 0| + |0\rangle \langle 1|)|1\rangle|^2 = \frac{1}{4} \tag{42}$$

but,

$$p(0, +) = |P_+ P_0 |1\rangle|^2 = |\frac{1}{2}(|0\rangle \langle 0| + |1\rangle \langle 0|)|1\rangle|^2 = 0. \tag{43}$$

If we examine the commutator, we see,

$$P_0 P_+ = \frac{1}{2}(|0\rangle \langle 0| + |0\rangle \langle 1|)$$
$$P_+ P_0 = \frac{1}{2}(|0\rangle \langle 0| + |1\rangle \langle 0|) \tag{44}$$

so

$$[P_0, P_+] = \frac{1}{2}(|0\rangle \langle 1| - |1\rangle \langle 0|) \neq 0 \tag{45}$$

However this masks a more complicated structure. The com-
mutator of two operators is itself an operator. There may be
some states for which the expectation value of this operator is
0, and thus for which the order of operations does not matter,
even though the commutator does not vanish identically. In-
deed, it is easy to see that,

$$\langle 0| [P_0, P_+] |0\rangle = \langle 1| [P_0, P_+] |1\rangle = 0 \tag{46}$$

On the other hand, we can show that the largest order effects
come from the two eigenstates of the commutator, $|\pm i\rangle$, where,

$$|+i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), \quad |-i\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) \tag{47}$$

and we have [7],

$$\langle \pm i| [P_0, P_+] |\pm i\rangle = \pm \frac{i}{2}. \tag{48}$$

In addition we have the following very important property:
Consider any two operators A and B, and look at the average
value of their commutator across a complete basis of $\mathcal{H}$. This
is equivalent to computing the expectation value of the com-
mutator in a state where $\rho = \frac{1}{D}\mathbb{1}$. Such a state corresponds to
an ensemble where participants are equally likely to be in any
initial knowledge state, i.e. there are no privileged states. We
see,

$$\text{Tr}([A, B]\mathbb{1}) = \text{Tr}(AB - BA) = 0. \tag{49}$$

In other words, this type of order effect averages out across
Hilbert space. The presence of a particular order effect is there-
fore as much about the initial state as it is about the operators

---

[7]It is easy to see that for two Hermitian operators A and B, $[A, B]$ is *anti-
Hermitian*, i.e. $[A, B]^\dagger = -[A, B]$. Anti-Hermitian operators have pure imagi-
nary eigenvalues, a fact that follows from a simple modification of Theorem 1
of the Appendix.

in question. (We will see in a later section an example of a different type of order effect which does not average out in this way.)

The discussion so far might lead you to conclude that commutation is a binary property, two operators either commute or they do not. This is technically true, but it is more helpful to think of commutation as a continuum. To explore this let's consider the projection operator P_θ, defined by

$$P_\theta = U(\theta)P_0 U(\theta)^\dagger \qquad (50)$$

where, in the basis $\{|0\rangle, |1\rangle\}$,

$$U(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \qquad (51)$$

so P_θ is a unitary rotation of P_0. In the basis $\{|0\rangle, |1\rangle\}$ we have,

$$P_\theta = \begin{pmatrix} \cos^2(\theta) & \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) & \sin^2(\theta) \end{pmatrix} \qquad (52)$$

We can show,

$$[P_0, P_\theta] = \frac{1}{2}\sin(2\theta)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad (53)$$

The eigenvectors of $[P_0, P_\theta]$ are equal to the eigenvectors of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, i.e. they are the same as $|\pm i\rangle$ above, but the eigenvalues are now given by $\lambda_\pm = \pm\frac{i\sin(2\theta)}{2}$ The maximum order effects that we can observe are therefore given by,

$$\langle \pm i | [P_0, P_\theta] | \pm i \rangle = \frac{\pm i}{2}\sin(2\theta). \qquad (54)$$

Therefore although strictly P_0 and P_θ do not commute unless θ is a multiple of $\pi/2$, there is a sense in which they *approximately* commute for values of θ close to these. In a realistic experiment with some low level of noise in the experimental data, for these values of θ the maximum possible order effects may be too small to observe against the noise.

Now thus far we have been interested in the expected value of commutators as a measure of order effects. However when we perform a sequence of measurements of two observables, A and B this is not in fact what we obtain. Rather the probability of getting a positive result for the measurement of A followed by a positive one for the measurement of B is given by,

$$p(A \text{ and then } B) = \text{Tr}(P_B P_A \rho P_A) = |P_B P_A |\psi\rangle|^2 \qquad (55)$$

where the second equality holds when the initial knowledge state is pure. The probability for the same measurements in the opposite order is given by,

$$p(B \text{ and then } A) = \text{Tr}(P_A P_B \rho P_B) = |P_A P_B |\psi\rangle|^2 \qquad (56)$$

We are therefore interested in the quantity,

$$\begin{aligned} p(A &\text{ and then } B) - p(B \text{ and then } A) \\ &= \text{Tr}(P_B P_A \rho P_A) - \text{Tr}(P_A P_B \rho P_B) \qquad (57) \\ &= \text{Tr}(\{P_A P_B P_A - P_B P_A P_B\}\rho) \end{aligned}$$

The order effect we will observe in a measurement therefore depends on the operator $P_A P_B P_A - P_B P_A P_B$. Note that this operator is Hermitian. Using the same 2D Hilbert space as above, and taking $P_A = P_0$ and $P_B = P_\theta$ we can show, in the $\{|0\rangle, |1\rangle\}$ basis,

$$P_0 P_\theta P_0 - P_\theta P_0 P_\theta = \cos^2(\theta)\sin(\theta)\begin{pmatrix} \sin(\theta) & -\cos(\theta) \\ -\cos(\theta) & -\sin(\theta) \end{pmatrix} \qquad (58)$$

The eigenvalues of this operator can be easily, if somewhat tediously, shown to be,

$$\lambda_\pm = \pm\cos^2(\theta)\sin(\theta) \qquad (59)$$

The eigenvectors can be computed if required, but they are not necessary here. The key is that the maximum possible order effects that we can observe are given by the eigenvalues of this operator.

It is worth looking in more detail at how these eigenvalues behave with different values of θ. Firstly they are zero for $\theta = 0, \pi/2$ as expected. The maximum value is,

$$\begin{aligned} max(\cos^2(\theta)&\sin(\theta)) = \\ &\sin\left(2\tan^{-1}\left(\sqrt{5-2\sqrt{6}}\right)\right)\cos^2\left(2\tan^{-1}\left(\sqrt{5-2\sqrt{6}}\right)\right) \\ &\approx 0.385 \end{aligned}$$
$$(60)$$

Which occurs at

$$\theta = 2\tan^{-1}\left(\sqrt{5-2\sqrt{6}}\right) \approx 35.2° \qquad (61)$$

Somewhat surprisingly then, the angle between the projectors P_0 and P_θ at which the measured order effect is maximised is not the same as the angle at which the commutator attains its maximum value, which recall was $\pi/4$ radians or $45°$. Also, the maximum possible order effect, in the sense of the difference between the probabilities $p(A \text{ and then } B)$ and $p(B \text{ and then } A)$ is limited to about .385. This is significant because logically, since the probabilities themselves are bounded by $0, 1$, one might have expected to be able to produce order effects of any size from 0 to 1.

So far in this section, we have glimpsed how order effects may be modelled, and we have seen that there are some non-trivial restrictions on the type of effects that can be accounted for. In the next section we will introduce a precise test of the quantum account of order effects which allows the quantum account to be compared with other models.

5. Order Effects 2

Designing a theory to fit existing data is relatively easy. What is much harder is to develop an a priori or parameter free test of a theory which would allow the theory to be falsified. As we will now show, quantum theory is very constrained with respect to the types of order effects it can account for. This leads to a remarkable equality between various probabilities which must

be satisfied if the probabilities are to be explained by quantum theory (Wang and Busemeyer, 2013; Wang et al., 2014).

Our aim in this section is to show in detail how this constraint, which is known as the Quantum Question, or QQ Equality, is derived. The derivation in the original papers was rather long, here we will make use of the machinery of quantum theory we have discussed in this tutorial to give a simpler and hopefully more transparent derivation.

Suppose we have two different measurements we can perform, A or B, and crucially these measurements have two possible outcomes, $A, \bar{A}$ and $B, \bar{B}$. Denote the projection operators corresponding to these measurement outcomes as $P_A, P_{\bar{A}}, P_B, P_{\bar{B}}$, with $P_A + P_{\bar{A}} = P_B + P_{\bar{B}} = \mathbb{1}$. Let's write the probabilities associated with these measurements as $p(A, B), p(\bar{A}, B)$... where $p(A, B) = Tr(P_B P_A \rho P_A)$ etc.

Now since these measurements don't necessarily commute,

$$p(A, B) \neq p(B, A) \qquad (62)$$

However, slightly unexpectedly, the following is true,

$$p(A, B) + p(\bar{A}, \bar{B}) - p(B, A) - p(\bar{B}, \bar{A}) = 0 \qquad (63)$$

This is the QQ Equality (Wang and Busemeyer, 2013). (This equality also holds if we relabel $A \leftrightarrow \bar{A}$, or $B \leftrightarrow \bar{B}$, and the original form of the equality given in Wang and Busemeyer (2013); Wang et al. (2014) had $B \leftrightarrow \bar{B}$, but the form given above is better to see the essential symmetry of the equality.)

The QQ Equality holds because the following operator is identically zero,

$$P_A P_B P_A + P_{\bar{A}} P_{\bar{B}} P_{\bar{A}} - P_B P_A P_B - P_{\bar{B}} P_{\bar{A}} P_{\bar{B}} = 0 \qquad (64)$$

We give a proof of this below.

This derivation makes use of the properties of commutators and projection operators, which together imply,

$$[P_A, P_B] = [P_A, (1 - P_{\bar{B}})] = -[P_A, P_{\bar{B}}] = [P_{\bar{B}}, P_A]$$
$$= -[P_{\bar{B}}, P_{\bar{A}}] = [P_{\bar{A}}, P_{\bar{B}}] \text{ etc.} \qquad (65)$$

Note however that the use of commutators is just a mathematical convenience.

Now for the derivation. We begin with,

$$[P_A, P_B] - [P_A, P_B] = 0 \qquad (66)$$

Inserting two copies of the identity gives,

$$[P_A, P_B]P_A + [P_A, P_B]P_{\bar{A}} - P_B[P_A, P_B] - P_{\bar{B}}[P_A, P_B] = 0 \quad (67)$$

Now we use the property of the commutator noted above, to get,

$$[P_A, P_B]P_A + [P_{\bar{A}}, P_{\bar{B}}]P_{\bar{A}} - P_B[P_A, P_B] - P_{\bar{B}}[P_{\bar{A}}, P_{\bar{B}}] = 0 \quad (68)$$

Expanding out the commutators gives,

$$P_A P_B P_A + P_{\bar{A}} P_{\bar{B}} P_{\bar{A}} - P_B P_A P_B - P_{\bar{B}} P_{\bar{A}} P_{\bar{B}} = 0 \qquad (69)$$

which is Eq.(64). Since this operator is identically zero, it follows that for any density operator ρ,

$$Tr(\{P_A P_B P_A + P_{\bar{A}} P_{\bar{B}} P_{\bar{A}} - P_B P_A P_B - P_{\bar{B}} P_{\bar{A}} P_{\bar{B}}\}\rho) = 0 \quad (70)$$

By the linearity and cyclic property of the trace this gives,

$$p(A, B) + p(\bar{A}, \bar{B}) - p(B, A) - p(\bar{B}, \bar{A}) = 0 \qquad (71)$$

Where $p(A, B) = Tr(P_B P_A \rho P_A)$ etc.

This QQ Equality was first introduced by Wang and Busemeyer (2013), and it was strongly supported experimentally in a very compelling paper by Wang et al. (2014). That the QQ Equality is obeyed by real world data is surprising in view of the fact that it does not follow from general contextual probability theories, but can only (apparently) be derived assuming the particular structures of quantum theory. This means observing that the QQ Equality is obeyed in the wild provides very good evidence that human decision-making can be described by quantum theory in particular, rather than just a contextual probability theory in general.

It is worth asking exactly what properties of quantum theory the derivation relies on. That is, exactly what properties does the QQ Equality test? The key ones appear to be the properties of the projection operators representing measurements, specifically the fact that they are orthogonal and idempotent. This raises the question of whether the QQ Equality still holds if the measurement operators do not have these 'ideal' properties, such as one might expect to be appropriate for more realistic measurements. We will address this question in a later section.

Another property that this derivation relied upon was that the measurements had binary outcomes. Despite extensive work, we have not been able to prove a generalisation of this result for measurements with more than two possible outcomes. It remains unclear whether any such analogue exists, but it would obviously be of huge interest to resolve this question.

6. A brief introduction to POVMs

The measurements we have been dealing with so far are *ideal* measurements, in the sense that they are perfectly idempotent and perfectly orthogonal. This means if the cognitive state of a participant happens to be an eigenstate $|a_i\rangle$ of an operator A, if we perform a measurement of the observable represented by A we will get the result a_i with probability 1.

Realistic experiments are never this straightforward, and there is always some irreducible noise that means that real measurements are not exactly ideal. The reason for this could be as simple as the presence of response errors; for example if an experiment is performed under conditions where speed is emphasised. Alternatively noise could enter because different response options are not viewed as orthogonal, or because it is impossible to separate out the cognitive variable one is interested in from others which are also effectively 'measured' when a decision is made. In order to account for this we need to generalise the ideal measurements we have been using so far. The correct way to do this in quantum theory is via the use of what

are known as *Positive Operator Valued Measures* or POVMs for short.

This tutorial is not the place to give a full introduction to POVM measurements[8]. However the noise inherent in realistic measurements can play an important part in determining whether we observe order effects in experiments, and it is therefore worth presenting a bare bones account that will enable us to see how noise in the measurements interacts with order effects.

Suppose we wish to model an experiment where we have participants express a preference for one of two alternatives, *A* or *B*, and that these are exhaustive and exclusive alternatives. In an ideal measurement these would be represented by projection operators $P_0 = |0\rangle\langle 0|$, $P_1 = |1\rangle\langle 1|$. Suppose instead our measurement isn't ideal but, intentionally or otherwise, is subject to some noise. This means some participants who really prefer 0 will select option 1, and vice versa.

Let's see how we might model this. What we want is an operator E_0, whose expectation value in the state $|0\rangle$ is close to one, but which also has a non-zero expectation value in the state $|1\rangle$, and likewise for E_1. That is,

$$\langle 0| E_0 |0\rangle = 1 - \epsilon, \quad \langle 1| E_0 |1\rangle = \epsilon,$$
$$\langle 0| E_1 |0\rangle = \epsilon, \qquad \langle 1| E_1 |1\rangle = 1 - \epsilon. \quad (72)$$

Where $0 \leq \epsilon \leq 1$ is some small error probability [9]. Let us also assume,

$$\langle 0| E_0 |1\rangle = 0, \text{ etc.} \quad (73)$$

In the basis $\{|0\rangle, |1\rangle\}$ these operators can therefore be written as,

$$E_0 = \begin{pmatrix} 1 - \epsilon & 0 \\ 0 & \epsilon \end{pmatrix}, \quad E_1 = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 - \epsilon \end{pmatrix}. \quad (74)$$

Can we use these operators to describe a measurement process? It is easily seen that they are not projection operators, nevertheless they satisfy the following properties,

- They are *positive* operators, which means they have positive eigenvalues.

- They are *complete*, in the sense that $E_0 + E_1 = \mathbb{1}$, where, recall, $\mathbb{1}$ is the identity operator over $\mathcal{H}$.

These properties mean that for any density operator,

$$0 \leq \text{Tr}(E_i\rho) \leq 1 \quad (75)$$

and

$$\sum_{i=0,1} \text{Tr}(E_i\rho) = 1 \quad (76)$$

The quantities $\text{Tr}(E_i\rho)$ can thus be interpreted as probabilities, and so E_0 and E_1 are good candidates to describe a measurement process.

But what measurement process do they describe? Well there are many ways to think about this, but probably the easiest is to note that we can write,

$$E_0 = (1 - \epsilon)P_0 + \epsilon P_1, \quad E_1 = \epsilon P_0 + (1 - \epsilon)P_1. \quad (77)$$

In other words, we can write these operators like,

$$E_0 = \sum_i p_0(i)P_i \quad (78)$$

where $p_0(i)$ have (loosely) the interpretation of probabilities. So one way to think about these measurements is that instead of performing a measurement P_0, we instead perform one of the possible measurements P_i with some probabilities $p_0(i)$. So these measurements look like noisy versions of ideal ones.

E_0 and E_1 are specific examples of elements of POVMs. The probability that a measurement described by E_i gives a positive answer is then given by,

$$p(i) = \text{Tr}(E_i\rho). \quad (79)$$

Suppose we perform a measurement of E_i, what is the resulting state after the measurement? Writing,

$$E_i = M_i^\dagger M_i \quad (80)$$

we can show that if we get a positive answer to the measurement described by E_i then the state collapses to,

$$\rho' = \frac{M_i\rho M_i^\dagger}{\text{Tr}(E_i\rho)}. \quad (81)$$

The M_i are often called 'measurement operators'. The decomposition of E_i into M_i is not unique; there in fact many different *realisations* of a given POVM in terms of a set of measurement operators M_i. We ignore this issue here, and simply take the simplest, $M_i = \sqrt{E_i}$. For more details see Yearsley (in preparation); Busch et al. (1995).

To return to our example above, in the basis $\{|0\rangle, |1\rangle\}$ the measurement operator associated with E_0 will be,

$$M_0 = \begin{pmatrix} \sqrt{1 - \epsilon} & 0 \\ 0 & \sqrt{\epsilon} \end{pmatrix} \quad (82)$$

which is nice and simple.

7. Order Effects and POVMs

We are now ready to explore the extent to which the presence and size of order effects depends on whether the measurements are perfect. We will start with the simple case, and then move on to consider whether the QQ Equality still holds for imperfect measurements.

---

[8]The best book JMY has read on POVMs is Busch et al. (1995), however this is currently out of print. A popular alternative is Neilsen and Chuang (2000). See also Yearsley (in preparation).

[9]An important point here is that we are assuming the errors are *unbiased*, i.e. the probability of incorrectly answering 1 when the state is $|0\rangle$ is the same as the probability of incorrectly answering 0 when the state is $|1\rangle$. This need not hold in general, but this would describe a something other than pure *noise*, so we will ignore this possibility here.

13

We will replace the projection operators P_A, P_B, with the following POVMs [10],

$$P_B \rightarrow E_B = \begin{pmatrix} \epsilon & 0 \\ 0 & 1-\epsilon \end{pmatrix}$$

$$P_+ \rightarrow E_+ = \begin{pmatrix} \frac{1}{2} & \frac{1-2\epsilon}{2} \\ \frac{1-2\epsilon}{2} & \frac{1}{2} \end{pmatrix} \tag{83}$$

These have associated measurement operators,

$$M_B = \begin{pmatrix} \sqrt{\epsilon} & 0 \\ 0 & \sqrt{1-\epsilon} \end{pmatrix}$$

$$M_+ = \begin{pmatrix} \frac{\sqrt{1-\epsilon}+\sqrt{\epsilon}}{2} & \frac{\sqrt{1-\epsilon}-\sqrt{\epsilon}}{2} \\ \frac{\sqrt{1-\epsilon}-\sqrt{\epsilon}}{2} & \frac{\sqrt{1-\epsilon}+\sqrt{\epsilon}}{2} \end{pmatrix} \tag{84}$$

Now we can see that,

$$p_\epsilon(+ \text{ and then } B) = \mathrm{Tr}(E_B M_+ \rho M_+)$$
$$= \frac{1}{4}\left(1 - 2(1-2\epsilon)\sqrt{\epsilon}\sqrt{1-\epsilon}\right) \tag{85}$$

and

$$p_\epsilon(B \text{ and then } +) = \mathrm{Tr}(E_+ M_B \rho M_B)$$
$$= \frac{\epsilon}{2} \tag{86}$$

We plot these results against the value of ϵ in Fig.(1). The results are interesting. The key is that the difference in the values of the probabilities (plotted as the dotted line) decreases sharply with increasing ϵ, i.e. with increasing noise. Note however that the value of ϵ is interpretable in terms of the 'error' probability of the measurement. Realistic experiments would probably have values of ϵ in the range 1-5%, and so order effects are still likely to be visible in such experiments, although they might appear smaller than one might have expected.
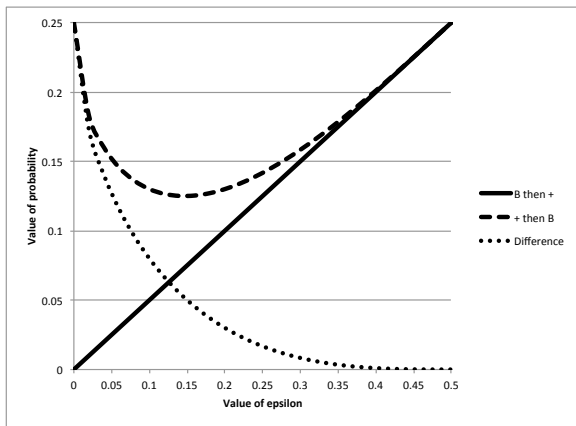

Figure 1: $p_\epsilon(B \text{ and then } +)$, $p_\epsilon(+ \text{ and then } B)$, and their difference, plotted against ϵ.

We don't have space here to pursue this further, but it is clear that small amounts of noise will still allow order effects to be observed, even though very large amounts of noise rapidly kill off such effects. This has important implications for studies looking for these effects in the wild.

If the noise inherent in realistic measurements can kill off order effects, can it also disrupt the important QQ Equality that we introduced previously to test the quantum explanation for such effects? It is obviously vital to know the answer to this if we want to understand whether we expect to see this equality obeyed under realistic conditions.

What we want to do is to examine what happens to the operator in Eq.(64) when we replace the projectors with POVMs. Clearly the derivation that led to Eq.(64) will no longer hold, but it is unclear whether this means the equality is no longer valid.

We are going to restrict ourselves in what follows to a very simple class of POVMs, in particular we will take, in the basis $\{|A\rangle, |\bar{A}\rangle\}$

$$E_A = \begin{pmatrix} 1-\epsilon & 0 \\ 0 & \epsilon \end{pmatrix}, \quad E_{\bar{A}} = \begin{pmatrix} \epsilon & 0 \\ 0 & 1-\epsilon \end{pmatrix}. \tag{87}$$

and in the basis $\{|B\rangle, |\bar{B}\rangle\}$

$$E_B = \begin{pmatrix} 1-\eta & 0 \\ 0 & \eta \end{pmatrix}, \quad E_{\bar{B}} = \begin{pmatrix} \eta & 0 \\ 0 & 1-\eta \end{pmatrix}. \tag{88}$$

We will also chose the measurement operators to have a simple form. Eg in the $\{|A\rangle, |\bar{A}\rangle\}$ basis

$$M_A = \begin{pmatrix} \sqrt{1-\epsilon} & 0 \\ 0 & \sqrt{\epsilon} \end{pmatrix}. \tag{89}$$

This is useful because it means we can write,

$$E_A = (1-\epsilon)P_A + \epsilon P_{\bar{A}}, \quad M_A = \sqrt{1-\epsilon}P_A + \sqrt{\epsilon}P_{\bar{A}} \tag{90}$$

etc. This choice of POVM is special because of the fact that the E_A etc can be written as sums of commuting projection operators. This means they are in some sense very close in behaviour to the more usual projective measurements[11]. The significance of this is discussed in Busch et al. (1995).

We will now see what happens when we replace the projectors with POVMs in Eq.(64). We will use the trick of replacing the POVMs with sums of projection operators, as above. This then lets us write,

$$M_A E_B M_A + M_{\bar{A}} E_{\bar{B}} M_{\bar{A}} - M_B E_A M_B - M_{\bar{B}} E_{\bar{A}} M_{\bar{B}} =$$
$$(\sqrt{1-\epsilon}P_A + \sqrt{\epsilon}P_{\bar{A}})((1-\eta)P_B + \eta P_{\bar{B}})(\sqrt{1-\epsilon}P_A + \sqrt{\epsilon}P_{\bar{A}})$$
$$+ (\sqrt{\epsilon}P_A + \sqrt{1-\epsilon}P_{\bar{A}})(\eta P_B + (1-\eta)P_{\bar{B}})(\sqrt{\epsilon}P_A + \sqrt{1-\epsilon}P_{\bar{A}})$$
$$- (\sqrt{1-\eta}P_B + \sqrt{\eta}P_{\bar{B}})((1-\epsilon)P_A + \epsilon P_{\bar{A}})(\sqrt{1-\eta}P_B + \sqrt{\eta}P_{\bar{B}})$$
$$- (\sqrt{\eta}P_B + \sqrt{1-\eta}P_{\bar{B}})(\epsilon P_A + (1-\epsilon)P_{\bar{A}})(\sqrt{\eta}P_B + \sqrt{1-\eta}P_{\bar{B}}) \tag{91}$$

14

Some rather tedious algebra gives,

$$M_A E_B M_A + M_{\bar{A}} E_{\bar{B}} M_{\bar{A}} - M_B E_A M_B - M_{\bar{B}} E_{\bar{A}} M_{\bar{B}} =$$
$$(P_A P_B P_A + P_{\bar{A}} P_{\bar{B}} P_{\bar{A}} - P_B P_A P_B - P_{\bar{B}} P_{\bar{A}} P_{\bar{B}})((1-\epsilon)(1-\eta) + \epsilon\eta)$$
$$+ (P_A P_{\bar{B}} P_A + P_{\bar{A}} P_B P_{\bar{A}} - P_B P_{\bar{A}} P_B - P_{\bar{B}} P_A P_{\bar{B}})((1-\epsilon)\eta + \epsilon(1-\eta))$$
$$+ (P_A P_B P_{\bar{A}} + P_{\bar{A}} P_B P_A + P_A P_{\bar{B}} P_{\bar{A}} + P_{\bar{A}} P_{\bar{B}} P_A) \sqrt{\epsilon}\sqrt{1-\epsilon}$$
$$- (P_B P_A P_{\bar{B}} + P_{\bar{B}} P_A P_B + P_B P_{\bar{A}} P_{\bar{B}} + P_{\bar{B}} P_{\bar{A}} P_B) \sqrt{\eta}\sqrt{1-\eta} \tag{92}$$

The first term on the right hand side of this equation is proportional to the original QQ operator Eq.(64) and thus vanishes. The second term is also proportional to the QQ operator, with B and $\bar{B}$ interchanged, and thus this also vanishes. What we are left with is,

$$M_A E_B M_A + M_{\bar{A}} E_{\bar{B}} M_{\bar{A}} - M_B E_A M_B - M_{\bar{B}} E_{\bar{A}} M_{\bar{B}} =$$
$$(P_A P_B P_{\bar{A}} + P_{\bar{A}} P_B P_A + P_A P_{\bar{B}} P_{\bar{A}} + P_{\bar{A}} P_{\bar{B}} P_A) \sqrt{\epsilon}\sqrt{1-\epsilon} \tag{93}$$
$$- (P_B P_A P_{\bar{B}} + P_{\bar{B}} P_A P_B + P_B P_{\bar{A}} P_{\bar{B}} + P_{\bar{B}} P_{\bar{A}} P_B) \sqrt{\eta}\sqrt{1-\eta}$$

By using the properties of projection operators, we see e.g.,

$$P_A P_{\bar{B}} P_{\bar{A}} = P_A (1 - P_{\bar{B}}) P_{\bar{A}} = -P_A P_{\bar{B}} P_{\bar{A}} \tag{94}$$

etc. From this we can see that the both the terms on the right hand side of Eq.(93) also vanish, and so finally we see,

$$M_A E_B M_A + M_{\bar{A}} E_{\bar{B}} M_{\bar{A}} - M_B E_A M_B - M_{\bar{B}} E_{\bar{A}} M_{\bar{B}} = 0 \tag{95}$$

Therefore for any density operator ρ we have,

$$\text{Tr}(\{M_A E_B M_A + M_{\bar{A}} E_{\bar{B}} M_{\bar{A}} - M_B E_A M_B - M_{\bar{B}} E_{\bar{A}} M_{\bar{B}}\}\rho) = 0 \tag{96}$$

and thus,

$$p_{\epsilon,\eta}(A,B) + p_{\epsilon,\eta}(\bar{A},\bar{B}) - p_{\epsilon,\eta}(B,A) - p_{\epsilon,\eta}(\bar{B},\bar{A}) = 0 \tag{97}$$

Where $p_{\epsilon,\eta}(A,B) = \text{Tr}(E_B M_A \rho M_A)$ etc.

Therefore we have shown the QQ Equality still holds when the measurements are given by these simple types of POVM. This is an important result, since it helps to explain why the equality seems to hold so well in real world data sets (Wang et al., 2014).

Note that this result relied on the POVMs having unbiased errors. If we relax this condition then the QQ equality no longer holds. This corresponds, for example, to having systematic error that bias the result towards or away from A, and similarly for B. We will not explore this possibility further here.

In summary, using realistic noisy measurements does effect the size and presence of order effects. In general adding noise to a quantum system tends to make it look more classical, and this case provides a good example of that. (For a good introduction to the quantum to classical transition see Halliwell (2005). For a classic review we recommend Zurek (1991).) However for realistic levels of noise order effects should still be visible, and crucially the QQ Equality will still hold provided the noisy measurements remain unbiased.

8. Order Effects in Similarity Judgments

In this final section we want to explore the presence of order effects in a slightly different context, namely that of similarity judgments. Similarity judgments are an important part of many other more complicated cognitive processes such as categorisation. In a classic paper Tversky (1977) demonstrated that human similarity judgments have a number of properties that one would not expect from the simplest classical models. The most important for us is that similarity can sometimes display asymmetry, in other words, if $Sim(A,B)$ denotes the similarity of A to B, then,

$$Sim(A,B) \neq Sim(B,A) \tag{98}$$

In particular, so the theory goes (Tversky, 1977), $Sim(A,B) \geq Sim(B,A)$ if we have more knowledge about B than about A.

In an attempt to model this interesting finding, Pothos et al. (2013) came up with a model of similarity judgments based on quantum theory (see also Pothos et al. (2015)). Their model works as follows, suppose we have a space of concepts $\mathcal{H}$, with subspaces representing the concepts A and B defined by the projection operators[12] P_A and P_B. Then the judged similarity of A to B, given our initial knowledge state is ρ, is given by,

$$Sim(A,B) = \text{Tr}(P_B P_A \rho P_A)/\text{Tr}(P_A \rho) \tag{99}$$

or

$$Sim(A,B) = \frac{|P_B P_A |\psi\rangle|^2}{|P_A |\psi\rangle|^2} \tag{100}$$

if the state is pure[13]. In actual fact Pothos et al didn't include the normalisation factor in the denominator of Eq.(99). However they left the question of normalisation open, so we may explore any reasonable choice of normalisation.

The initial knowledge state is to be set such that the state is 'unbiased'. Specifically "in the absence of priming manipulation or contextual influence, we require the state vector to be neutral between the compared concepts" (Pothos et al., 2013). One might expect that this means the initial state is set such that,

$$\text{Tr}(P_A \rho) = \text{Tr}(P_B \rho) \tag{101}$$

and this is indeed what Pothos et al had in mind. However there are other ways to interpret this. They go on to say "Such an assumption is equivalent to that of a uniform prior in a Bayesian model." Now the concept of a uniform prior is rather slippery, because it depends on what exactly it is uniform over. In this case there are two obvious choices; the prior could be uniform over the whole knowledge space, so one is equally likely to be thinking about any thought in the space, or it could be uniform over the concepts under consideration, which, recall, are represented by subspaces of dimension potentially greater than

---

[12] One might also use smeared projectors or POVMs instead of projection operators if there is some uncertainty about the concept boundary.

[13] This assumes that the similarity between a stimulus and itself is always judged to be 1. If desired, this assumption can be dropped by using POVMs in place of the projection operators.

one. A uniform prior over the whole knowledge space would be given by the density operator $\rho = \frac{1}{D}\mathbb{1}$, whereas a uniform prior over the concepts under consideration would be given by a density operator satisfying Eq.(101). One advantage of the former is that it is guaranteed to exist, whereas the latter is not. At any rate, in the absence of further information we will explore both possibilities.

This similarity measure looks like a quantum version of a conditional probability. It has two nice properties, it is normalised in the sense that,

$$S\,im(A, A) = 1 \tag{102}$$

and also in the sense that,

$$\sum_i S\,im(A, B_i) = 1 \tag{103}$$

if the set $\{B_i\}$ is complete, i.e. $\sum_i B_i = \mathbb{1}$.

Now the key factor responsible for asymmetry in this model is that P_A and P_B may have different dimensions. This corresponds in cognitive terms to a participant having greater knowledge about one of the stimuli. Greater knowledge implies more possible thoughts connected with that stimuli, which means a higher dimensional subspace for that stimuli. What we want to do is explore how the asymmetry depends on the relationship between the different subspaces corresponding to P_A and P_B. A full analysis is given in Pothos et al. (2013), what we want to do is explore a simple analytically tractable example.

As our example, we will take a three dimensional Hilbert space of possible thoughts, and P_A and P_B to be one and two dimensional respectively. It is useful to choose a basis of eigenstates of B, so that,

$$P_B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{104}$$

The bases formed by the eigenstates of A and B are related by a unitary transform U, so in the basis of eigenstates of B we can also write,

$$P_A = U^\dagger \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} U = U^\dagger P_{\bar{B}} U \tag{105}$$

The relationship between P_A and P_B, and therefore the asymmetry, depends on the unitary transformation U.

There appear to be several possible choices for U. However the only important component of any transformation is that which rotates P_A towards the plane defined by P_B. Anything else is equivalent to a relabelling of the coordinates. Therefore without loss of generality we can choose the transformation to simply rotate P_A towards one of the states in P_B. Since the set up is symmetric we can choose,

$$U = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{106}$$

this gives,

$$P_A = \begin{pmatrix} \cos^2(\theta) & -\cos(\theta)\sin(\theta) & 0 \\ -\cos(\theta)\sin(\theta) & \sin^2(\theta) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{107}$$

Now we need to ensure we have an allowable initial state. Recall we have two choices, either our initial state is proportional to the identity, or we must choose it so that $\langle P_A \rangle = \langle P_B \rangle$.

Let us start with the simpler case, of $\rho = \frac{1}{3}\mathbb{1}$. Clearly,

$$\text{Tr}(P_B P_A \rho P_A) = \frac{1}{3}\text{Tr}(P_A P_B) = \text{Tr}(P_A P_B \rho P_B) \tag{108}$$

so the numerators of the two similarities, Eq.(99), are equal. However we can also see,

$$\text{Tr}(P_A \rho) = \frac{1}{3}\text{Tr}(P_A) = \frac{1}{3}, \quad \text{Tr}(P_B \rho) = \frac{1}{3}\text{Tr}(P_B) = \frac{2}{3}. \tag{109}$$

Thus we have,

$$\frac{S\,im(A, B)}{S\,im(B, A)} = \frac{\text{Tr}(P_B P_A \rho P_A)\text{Tr}(P_B \rho)}{\text{Tr}(P_A P_B \rho P_B)\text{Tr}(P_A \rho)} = 2. \tag{110}$$

We see therefore that in the case $S\,im(A, B)$ is always larger than $S\,im(B, A)$ by a factor of two. In general the ratio of the similarities is the inverse of the ratio of the dimension of the subspaces, and thus the inverse of the ratio of the 'degree of knowledge' we have about A and B, as desired [14]. Turning back to our specific example, $S\,im(A, B)$ is easily computed to be $\sin^2(\theta)$.

In summary, this choice of initial state is interesting because it always gives rise to the expected asymmetry. It also produces a particularly simple expression for the similarities. However it is an open question whether the observed asymmetry shares this behaviour.

Now let us turn to the more challenging case. It is easiest to start by looking at the eigenvectors of $A - B$. Once we have these we can take a combinations that give zero expectation value for $A - B$, and which are thus unbiased. In the B basis the eigenvectors are given by,

$$|v_1\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad |v_2\rangle = N_2 \begin{pmatrix} \frac{(\cos(\theta)-1)}{\sin(\theta)} \\ 1 \\ 0 \end{pmatrix}, \quad |v_3\rangle = N_3 \begin{pmatrix} \frac{(\cos(\theta)+1)}{\sin(\theta)} \\ 1 \\ 0 \end{pmatrix}. \tag{111}$$

with eigenvalues, $-1, -\cos(\theta)$ and $\cos(\theta)$ respectively. Here N_2 and N_3 are normalisation factors given by,

$$N_2 = \sqrt{\frac{\sin^2(\theta)}{2(1 - \cos(\theta))}}, \quad N_3 = \sqrt{\frac{\sin^2(\theta)}{2(1 + \cos(\theta))}} \tag{112}$$

We will choose a pure initial state for simplicity, and we will also choose the state to be real. Then the following (unnormalised) set of states have zero expectation value with respect to the operators $A - B$,

$$|\psi_a\rangle = a\sqrt{\cos(\theta)}\,|v_1\rangle + \sqrt{1 - a^2}\,|v_2\rangle + |v_3\rangle \tag{113}$$

---

[14]This way of achieving asymmetries can also be implemented classically. Pothos et al (2013) argue that this might not represent a good solution. However we are only interested here in exploring the structure of the theory.
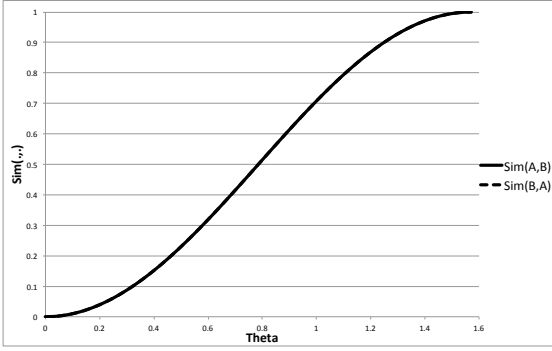
(a) Similarities for the choice $a = 0$. The two similarities coincide in this case, so there is no asymmetry.
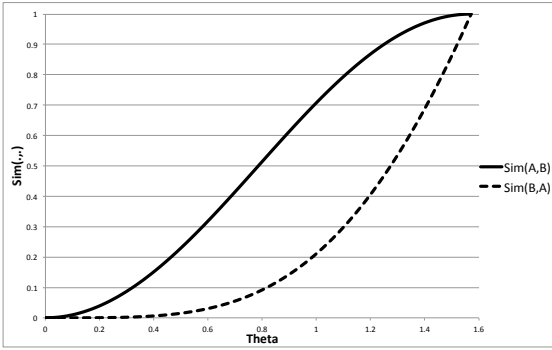
(b) Similarities for the choice $a = 1$. $Sim(A, B)$ is always greater than $Sim(B, A)$, apart from at the end points $\theta = 0, \pi/2$.

Figure 2: Plots of $Sim(A, B)$ and $Sim(B, A)$ against θ for two choices of a.

where $0 \leq a \leq 1$. Thus these states have the same expectation value for the operators A and B and are 'unbiased' in the sense of Pothos et al. Note that there is not a single state with this property, but crucially rather a whole family of possible states.

Now we can compute the similarities $Sim(A, B)$ etc according to Eq.(99). We save the reader the algebra involved in this, but note it is quite straightforward to run in MatLab or similar. We plot the special cases of $a = 0$ and $a = 1$ in Fig.(2), the general case interpolates between these two extremes. The important point is that we generally have $Sim(A, B) \geq Sim(B, A)$ with equality in the case $a = 0$ and for the values $\theta = 0, \pi/2$.

We see that with this choice of initial state the similarities show a rather more complex behaviour. The ratio of the similarities is now a function of θ, as well as of the new parameter a. Thus in this case the degree of asymmetry predicted depends on how incompatible A and B are.

The purpose of this section was to introduce a different type of order effect, which occurs in models of similarity judgments, that behaves in a slightly different way to the order effects we see in survey designs. Recall, measured order effects in survey designs result because the operator,

$$\Delta = P_A P_B P_A - P_B P_A P_B \qquad (114)$$

is non-zero. However, although $\Delta \neq 0$ it is *traceless*, which means that the average of Δ over the whole Hilbert space is zero. In other words, order effects in survey designs are generally balanced, for every cognitive state which gives rise to a

particular asymmetry, there is another state which gives the opposite result. In contrast, the order effects that produce asymmetries in similarity judgments do not have this property, and it can be that case that $Sim(A, B) \geq Sim(B, A)$ for all states in the cognitive space.

9. Summary

Order effects are an important feature of the study of human decision making. But order effects are just one of many different kinds of phenomena addressed by quantum models of cognition. Others include the conjunction fallacy, violations of total probability, over and under extension effects in conceptual combinations, etc. We hope this tutorial has convinced you that quantum models represent a useful and potentially powerful way to model these phenomena. The existence of order or asymmetry effects strongly suggests that theories beyond classical probability theory are needed to model this type of decision making. However simply exchanging non-contextual for contextual probability theories leaves us with too many feee parameters, and no underlying sense about *why* these effects occur. Quantum models, on the other hand, are both more constrained than the general case of a contextual probability theory, and also offer the possibility of achieving an understanding of why such effects occur in terms of incompatibility of mental representations. So far, it has been the success of quantum models at predicting observed constraints on various phenomena that has been most persuasive. In the future, we hope that these models will also impress by the story they can tell about the reasons behind the occurrence of other puzzling violations of classical probability theory in human decision making.

As well as having a particular focus on order effects, we hope this tutorial has also been useful to researchers looking to construct quantum models for other purposes. We have strived to set out in as clear a way as possible how and why quantum models are constructed to address particular questions. While it is impossible to cover every possible scenario one might encounter when trying to set up a model, hopefully this tutorial will have given the reader the confidence to experiment with constructing quantum models of their own.

Appendix A. Review of Linear Algebra

In this appendix we give a brief review of some relevant concepts from linear algebra. Our aim is not to cover all possible topics, but just to recap some key ideas. In particular we will show how some familiar ideas may be expressed in a basis independent way, which means they can be computed directly in terms of the elements of Hilbert space.

17

A Hilbert space $\mathcal{H}$ is essentially a complex vector space which is a complete metric space with respect to the inner product,

$$\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$$
$$(|u\rangle, |v\rangle) \to \langle u|v\rangle \tag{A.1}$$

with the usual properties (see Isham (1995); Reed and Simon (1980) for a fuller account.) The elements of $\mathcal{H}$ are denoted $|u\rangle$ and are known as 'kets'.

An operator on Hilbert space is a map from the space to itself,

$$A : \mathcal{H} \to \mathcal{H}$$
$$|u\rangle \to A|u\rangle \tag{A.2}$$

An important operator is the *identity operator* denoted $\mathbb{1}$. It is defined by,

$$\mathbb{1}A = A\mathbb{1} = A, \quad \forall A : \mathcal{H} \to \mathcal{H}. \tag{A.3}$$

The eigenvectors, $\{|a_i\rangle\}$ and eigenvalues a_i of A are defined by,

$$A|a_i\rangle = a_i|a_i\rangle \tag{A.4}$$

The object $\langle u|$ is known as a 'bra'. It is not technically an element of $\mathcal{H}$, but can be thought of either as an element of the dual space, or as a functional that maps elements of $\mathcal{H}$ to the complex numbers,

$$\langle u| : \mathcal{H} \to \mathbb{C}$$
$$|v\rangle \to \langle v|u\rangle \tag{A.5}$$

Hermitian conjugation of an operator is a useful concept in linear algebra. The Hermitian conjugate of an operator $A^\dagger$ is defined by,

$$\langle A|u\rangle, |v\rangle\rangle = \langle |u\rangle, A^\dagger|v\rangle\rangle, \quad \forall |u\rangle, |v\rangle \in \mathcal{H}. \tag{A.6}$$

In a particular basis, where A is represented by some matrix,

$$A^\dagger = (A^T)^* \tag{A.7}$$

where A^T is the transpose of A.

Hermitian conjugation has the following properties,

1. $(A^\dagger)^\dagger = A$
2. $(A + B)^\dagger = A^\dagger + B^\dagger$
3. $(AB)^\dagger = B^\dagger A^\dagger$
4. $(cA)^\dagger = c^* A^\dagger$ for $c \in \mathbb{C}$

It is useful to define,

$$\langle u| = |u\rangle^\dagger \tag{A.8}$$

From Eq.(A.4) we can see that,

$$A|a_i\rangle = a_i|a_i\rangle$$
$$\to (\langle a_i|A)^\dagger = (\langle a_i|a_i\rangle)^\dagger \tag{A.9}$$
$$\to A^\dagger|a_i\rangle = a_i^*|a_i\rangle$$

Thus the eigenvectors of $A^\dagger$ are the same as those of A, but the eigenvalues are the complex conjugates.

A useful class of operators in quantum theory are those which are *Hermitian* or more generally *Self-Adjoint* [15]. An operator is Hermitian if,

$$A^\dagger = A \tag{A.10}$$

We can now state a number of useful theorems about Hermitian operators.

Theorem 1. *The eigenvalues of an Hermitian operator are real.*

Proof: Let $|a_i\rangle$ be any non-zero eigenvector of A, with eigenvalue a_i.

$$A|a_i\rangle = a_i|a_i\rangle \tag{A.11}$$

Then,

$$\langle A|a_i\rangle, |a_i\rangle\rangle = \langle a_i|a_i\rangle, |a_i\rangle\rangle = a_i^* \langle |a_i\rangle, |a_i\rangle\rangle = a_i^* \langle a_i|a_i\rangle \tag{A.12}$$

But because A is Hermitian,

$$\langle A|a_i\rangle, |a_i\rangle\rangle = \langle |a_i\rangle, A^\dagger|a_i\rangle\rangle = \langle |a_i\rangle, A|a_i\rangle\rangle$$
$$= \langle |a_i\rangle, a_i|a_i\rangle\rangle = a_i \langle a_i|a_i\rangle \tag{A.13}$$

Thus,

$$0 = (a_i - a_i^*)\langle a_i|a_i\rangle \tag{A.14}$$

and since $|a_i\rangle$ is non-zero, we must have $a_i = a_i^*$. □

Note that the converse is also true, i.e. an operator is Hermitian if its eigenvalues are real.

Theorem 2. *The eigenvectors corresponding to two different eigenvalues are orthogonal.*

Proof: Let

$$A|a_1\rangle = a_1|a_1\rangle, A|a_2\rangle = a_2|a_2\rangle \tag{A.15}$$

with $a_1 \neq a_2$

Then,

$$\langle |a_2\rangle, A|a_1\rangle\rangle = \langle a_2, a_1|a_1\rangle\rangle = a_1 \langle a_2|a_1\rangle \tag{A.16}$$

But because A is Hermitian the left hand side is,

$$\langle A|a_2\rangle, |a_1\rangle\rangle = \langle a_2|a_2\rangle, |a_1\rangle\rangle = a_2^* \langle a_2|a_1\rangle = a_2 \langle a_2|a_1\rangle \tag{A.17}$$

Thus,

$$(a_1 - a_2)\langle a_2|a_i\rangle = 0 \tag{A.18}$$

and since $a_1 \neq a_2$, we must have $\langle a_2|a_1\rangle = 0$. □

Another important class of operators are those which are *Unitary*. An operator U is unitary if,

$$U^\dagger U = UU^\dagger = \mathbb{1} \tag{A.19}$$

---

[15] The distinction is only important for Hilbert spaces with an infinite number of dimensions. In these spaces it is not necessarily true that an operator and its adjoint have the same domain. Thus self-adjoint means that the domains of an operator and its adjoint match, and that the operators are equal on this shared domain (Reed and Simon, 1980).

18

i.e. $U^\dagger = U^{-1}$. The exponential of (i times) an Hermitian operator is a Unitary operator,

$$\left(e^{iH}\right)^\dagger = e^{(iH)^\dagger} = e^{-iH^\dagger} = e^{-iH} = \left(e^{iH}\right)^{-1} \tag{A.20}$$

(We will explain how to define the exponential of an operator below.)

A final important class of operators are those which are *Projection* operators or projectors. An Hermitian operator P is a projector if,

$$P^2 = P \tag{A.21}$$

We are now ready to state probably the most important theorem in quantum theory.

Theorem 3. *The Spectral Theorem for Hermitian (Self-Adjoint) Operators.*

The set of all eigenvectors of a Hermitian (self-adjoint) operator A defined on a Hilbert space $\mathcal{H}$ of dimension D forms an orthonormal basis for $\mathcal{H}$. Thus, any vector $|\psi\rangle \in \mathcal{H}$ can be expanded as,

$$|\psi\rangle = \sum_{i=1}^{D} \langle a_i|\psi\rangle |a_i\rangle . \tag{A.22}$$

The operator A itself can be written in the important form,

$$A = \sum_{i=1}^{D} a_i |a_i\rangle \langle a_i| \tag{A.23}$$

That is, any Hermitian (self-adjoint) operator may be written as a sum of projection operators onto subspaces spanned by the eigenvalues of A, each projection operator being multiplied by the associated eigenvalue.

Proof: The spectral theorem is a classical result in linear algebra. See any good book on linear algebra if you are interested.

Note: This analysis assumes the eigenvalues are not degenerate, i.e. that each eigenvector has a unique eigenvalue. All the theorems in this appendix can be trivially extended to the case where the spectrum of A is degenerate, see any good textbook on linear algebra or quantum theory Isham (1995); Reed and Simon (1980).

This theorem is important because Hermitian operators and their associated eigenspaces play a key role in quantum theory because of the association of Hermitian operators with *observables*. The spectral theorem says that any observable quantity defines a basis for the knowledge space, so that the space can be partitioned into subspaces each of which have a well defined value of the observable.

An important example of a Hermitian operator is the identity operator, $\mathbb{1}$. This can be thought of as an operator where all the eigenvalues are 1. Since it is Hermitian the spectral theorem applies, and so we can write,

$$\mathbb{1} = \sum_{i=1}^{D} |\phi_i\rangle \langle \phi_i| \tag{A.24}$$

where the $\{|\phi_i\rangle\}$ are any orthonormal basis for the space. The trick of inserting copies of the identity is often used in computations.

We can use this result to show that any state in the Hilbert space can be expanded in terms of the eigenstates of any observable. Consider an observable A, from the above we know,

$$\mathbb{1} = \sum_{i=1}^{D} |a_i\rangle \langle a_i| \tag{A.25}$$

Now consider this operator acting on any state vector,

$$|\psi\rangle = \mathbb{1} |\psi\rangle = \sum_{i=1}^{D} |a_i\rangle \langle a_i|\psi\rangle = \sum_{i=1}^{D} \psi_i^A |a_i\rangle \tag{A.26}$$

The square norms of the ψ_i^A can be interpreted as the probabilities that a measurement of the observable A would give the result a_i.

Since this decomposition of $|\psi\rangle$ can be performed in any basis, we can write any state as a superposition of the eigenstates of any observable in our theory.

One important consequence of the spectral theorem is that it lets us compute/define functions of operators. For example suppose we want to define A^2 where A is some Hermitian operator. Well we can write,

$$A^2 = \sum_{i=1}^{D} a_i |a_i\rangle \langle a_i| \sum_{j=1}^{D} a_j |a_j\rangle \langle a_j| = \sum_{i,j=1}^{D} a_i a_j \langle a_i|a_j\rangle |a_i\rangle \langle a_j|$$
$$= \sum_{i,j=1}^{D} a_i a_j \delta_{ij} |a_i\rangle \langle a_j| = \sum_{i=1}^{D} a_i^2 |a_i\rangle \langle a_i| \tag{A.27}$$

which is very simple. In general,

$$A^n = \sum_{i=1}^{D} a_i^n |a_i\rangle \langle a_i| \tag{A.28}$$

This suggests the following, suppose we take any function $f(\cdot)$, which can be defined in terms of a power series, i.e.

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + ... \tag{A.29}$$

then we can define,

$$f(A) = \alpha_0 + \alpha_1 A + \alpha_2 A^2 + ...$$
$$= \sum_{i=1}^{D} \left\{ \alpha_0 + \alpha_1 a_i + \alpha_2 a_i^2 + ... \right\} |a_i\rangle \langle a_i|$$
$$= \sum_{i=1}^{D} f(a_i) |a_i\rangle \langle a_i| \tag{A.30}$$

With sufficient care this can then be extended to essentially all functions. So we can take square roots, exponentials, trig functions of operators etc with abandon.

One final operation we need to define is trace of an operator.

$$\text{Tr}(A) = \sum_i \langle \phi_i| A |\phi_i\rangle \tag{A.31}$$

where the $\{|\phi_i\rangle\}$ form an orthonormal basis for $\mathcal{H}$. The trace of an operator is independent of the basis in which it is computed. The trace operation has the following useful properties,

1. $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$, etc.

2. $\mathrm{Tr}(A\,|\psi\rangle\langle\psi|) = \langle\psi|\,A\,|\psi\rangle$

In a particular basis, where A is represented by a matrix, the Trace of A is simply the sum of the diagonal elements of the matrix. One important consequence is that the trace of a projection operator is simply the dimension of the subspace onto which it projects.

Appendix B. Number of degrees of freedom associated with states, operators, etc.

Let us briefly explain how to compute the number of parameters associated with some objects in quantum theory. This is useful information for understanding the number of measurements needed to fix a particular state, or indeed how many parameters a quantum theory has in total, to compare with a classical theory.

A pure state $|\psi\rangle$ is a complex vector in a D dimensional space. As a column vector in a particular basis it would have D complex entries, which means $2D$ parameters to fix. However we also have two constraints. The first is that the state be normalised, the second is that the overall phase of a state is unobservable, so we can set it to 0 (Isham, 1995). This means a pure state has a total of $2D - 2$ free parameters.

An observable is an Hermitian operator, which in a particular basis can be written as a $D \times D$ matrix with complex entries, which would require $2D^2$ parameters to specify. However we have the constraint that an observable must be Hermitian, which in terms of matrix elements means $A_{ij} = A_{ji}^*$. This reduces the number of parameters to be specified to D^2.

Finally, a density operator is an Hermitian operator, together with the condition that it must be trace normalised ($\mathrm{Tr}(\rho) = 1$). This is one extra constraint, so reduces the number of parameters by one. The number of free parameters in a density operator is therefore $D^2 - 1$.

References

Aerts, D (2009). *Quantum structure in cognition.* Journal of Mathematical Psychology, 53(5), 314-348.

Aerts, D., Gabora, L., & Sozzo, S. (2013). *Concepts and Their Dynamics: A Quantum-Theoretic Modeling of Human Thought.* Topics in Cognitive Science, 5(4), 737-772.

Atmanspacher, H, & Filk, T (2010). *A proposed test of temporal non-locality in bistable perception.* Journal of Mathematical Psychology, 54, 314-321.

Atmanspacher, H, & Römer, H (2012). *Order effects in sequential measurements of non-commuting psychological observables.* Journal of Mathematical Psychology, 56(4), 274-280.

beim Graben P, Atmanspacher H (2006). *Complementarity in classical dynamical systems.* Found Phys 36(2):291-306.

Bell, J.S. (2004). *Speakable and Unspeakable in Quantum Mechanics.* Cambridge University Press: Cambridge, UK.

Bengtsson, I. (2006). *Three ways to look at mutual unbiases bases.* Unpublished, but available at arXiv.org quant-ph/0610216.

Blutner, B & beim Graben, P (2014). *Descriptive and Foundational Aspects of Quantum Cognition.* Available at ArXiv.org 1410.3961.

Blutner R, Pothos EM, & Bruza P (2013) *A quantum probability perspective on borderline vagueness.* Top Cogn Sci 5(4):711-736.

Bruza, P, Kitto, K, Ramm, B. J. & Sitbon, L. (2015) *A probabilistic framework for analysing the compositionality of conceptual combinations.* Journal of Mathematical Psychology, 67, 26-38.

Busch, P, Grabowski, M & Lahti, PJ (1995). *Operational Quantum Physics.* (Springer Berlin Heidelberg).

Busemeyer, JR & Bruza, PD (2014). *Quantum Models of Cognition and Decision.* (Cambridge University Press).

Caves CM, Fuchs, CA & Schack, R (2002). *Quantum Probabilities as Bayesian Probabilities.* Phys. Rev. A 65, 022305.

Chater, N, Tenenbaum, JB, & Yuille, A (2006). *Probabilistic models of cognition: Conceptual foundations.* Trends in cognitive sciences, 10(7), 287-291.

Dirac, PAM (1982). *Principles of Quantum Mechanics.* (Oxford University Press).

Durt, T, Englert, BG, Bengtsson, I& Życzkowski, K (2010). *On mutually unbiased bases*, Int. J. Quantum Information, 8, 535-640.

Gigerenzer, G, Hertwig, R & Pachur, T (Eds.) (2011). *Heuristics: the foundations of adaptive behaviour.* (Oxford University Press).

Hagan S, Hameroff SR, Tuszynski JA (2002) *Quantum computation in brain microtubules: Decoherence and biological feasibility.* Phys Rev E 65(6):061901.

Halliwell, JJ (2005). *How the Quantum Universe Became Classical.* Contemporary Physics **46**, 93. Also available at arXiv.org quant-ph/0501119.

Hameroff SR, Penrose R (1996) *Conscious events as orchestrated spacetime selections.* J Conscious Stud 3(1):36-53.

Isham, C (1995). *Lectures on Quantum Theory.* (Imperial College Press.)

Khrennikov A (2010). *Ubiquitous Quantum Structure: From Psychology to Finance* (Springer, Berlin).

Kvam, PD, Pleskac, TJ, YU, S & Busemeyer, JR (2015). *Interference effects of choice on confidence: Quantum characteristics of evidence accumulation.* PNAS, 112, 10645.

McKemmish LK, Reimers JR, McKenzie RH, Mark AE, Hush NS (2009) *Penrose-Hameroff orchestrated objective-reduction proposal for human consciousness is not biologically feasible.* Phys Rev E 80(2):021912.

Mogilianasky, AL, Zamir, S, & Zwirn, H (2009). *Type indeterminacy: A model of the KT (Kahneman-Tversky)-man.* Journal of Mathematical Psychology, 53(5), 349-361.

Moore DW (2002) *Measuring new types of question order effects.* Public Opin Q 66(1): 80-91.

Neilsen, M & Chuang, I (2000). *Quantum Computation and Quantum Information.* (Cambridge University Press.)

Plenio, M. *Quantum Mechanics.* Lecture notes available at www3.imperial.ac.uk/pls/portallive/docs/1/613904.PDF

Peres A (1998). *Quantum Theory: Concepts and Methods.* (Kluwer Academic, Dordrecht, The Netherlands).

Pothos, EM, Barque-Duran, A, Yearsley, JM, Trueblood, JS, Busemeyer, J, & Hampton, JA. (2015). *Progress and current challenges with the quantum similarity model.* Frontiers in Psychology, **6**, 205.

Pothos, E.M. & Busemeyer, J.R. (2009). *A quantum probability explanation for violations of 'rational' decision theory.* Proc. R. Soc. B, 276, 2171.

Pothos EM, Busemeyer JR (2013) *Can quantum probability provide a new direction for cognitive modeling?* Behav Brain Sci 36(3):255-274.

Pothos, EM, Busemeyer, JR & Trueblood, JS (2013). *A quantum geometric model of similarity.* Psych. Rev. 120(3), 679.

Reed, M & Simon, B (1980). *Methods of Modern Mathematical Physics (Vol 1).* (Academic Press, London).

Schwinger, J (1960). *Unitary Operator Bases.* (Harvard University).

Tegmark M (2000) *Importance of quantum decoherence in brain processes.* Phys Rev E 61(4):4194-4206.

Trueblood, J. S. & Busemeyer, J. R. (2011). *A comparison of the belief-adjustment model and the quantum inference model as explanations of order effects in human inference.* Cognitive Science, 35, 1518-1552.

Tversky, A. (1977). *Features of Similarity.* Psychological Review, 84, 327-352.

Tversky, A. & Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases.* Science, 185, 1124-1131.

Wang, Z & Busemeyer, JR. (2013). *A quantum question order model supported by empirical tests of an a priori and precise prediction.* Topics in Cognitive Science, 5(4), 689-710.

Wang Z, Busemeyer JR, Atmanspacher H, Pothos EM (2013) *The potential of using quantum theory to build models of cognition.* Top Cogn Sci 5(4):672-

688.

Wang, Z, Solloway, T, Shiffrin, RM & Busemeyer, JR (2014). *Context effects produced by question orders reveal the quantum nature of human judgments.* PNAS, 111, 9431.

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). *Sometimes it does hurt to ask: the constructive role of articulating impressions.* Cognition, 133(1), 48.

Yearsley JM, (in preparation). *Tutorial on Advanced Tools and Concepts for Quantum Cognitive Theories.*

Yearsley, JM & Pothos, EM (2014). *Challenging the classical notion of time in cognition: a quantum perspective.* Proceedings of the Royal Society B 381, 1781, 20133056.

Yearsley, JM & Pothos, EM (in preparation). *Zeno's paradox in human decision making.* In this paper we ask some participants the same question up to 13 times! (With other things happening in between.) In this set up it matters hugely whether the measurements are *exactly*, or only *approximately* perfect.

Yukalov, VI, & Sornette, D (2011). *Decision theory with prospect interference and entanglement.* Theory and Decision, 70(3), 283-328.

Zurek, WH (1991). *Decoherence and the Transition from Quantum to Classical.* Physics Today **44**, 10, 36. Also available in an updated form online at arXiv.org quant-ph/0306072.