Introduction to Quantum Probability

for Social and Behavioral Scientists

Jerome R. Busemeyer

Indiana University

June 1, 2008

Send correspondence to:

Jerome R. Busemeyer
Department of Psychological and Brain Sciences
Indiana University
1101 E. 10th St.
Bloomington Indiana, 47405
jbusemey@indiana.edu

There are two related purposes of this chapter. One is to generate interest in a new and fascinating approach to understanding behavioral measures based on quantum probability principles. The second is to introduce and provide a tutorial of the basic ideas in a manner that is interesting and easy for social and behavioral scientists to understand.

It is important to point out from the beginning that in this chapter, quantum probability theory is viewed simply as an alternative mathematical approach for generating probability models. Quantum probability may be viewed as a generalization of classic probability. No assumptions about the biological substrates are made. Instead this is an exploration into new conceptual tools for constructing social and behavioral science theories.

Why should one even consider this idea? The answer is simply this (cf., Khrennikov, 2007). Humans as well as groups and societies are extremely complex systems that have a tremendously large number of unobservable states, and we are severely limited in our ability to measure all of these states. Also human and social systems are highly sensitive to context and they are easily disturbed and disrupted by our measurements. Finally, the measurements that we obtain from the human and social systems are very noisy and filled with uncertainty. It turns out that classical logic, classic probability, and classic information processing force highly restrictive assumptions on the representation of these complex systems. Quantum information processing theory provides principles that are more general and powerful for representing and analyzing complex systems of this type.

Although the field is still in a nascent stage, applications of quantum probability theory have already begun to appear in areas including information retrieval, language, concepts, decision making, economics, and game theory (see Bruza, Lawless, van Rijsbergen, & Sofge, 2007; Bruza, Lawless, van Rijsbergen, & Sofge, 2008; also see the Special Issue on Quantum Cognition and Decision to appear in Journal of Mathematical Psychology in 2008).

The chapter is organized as follows. First we describe a hypothetical yet typical type of behavioral experiment to provide a concrete setting for introducing the basic concepts. Second, we introduce the basic principles of quantum logic and quantum probability theory. Third we discuss basic quantum concepts including compatible and incompatible measurements, superposition, measurement and collapse of state vectors.

A simple behavioral experiment.

Suppose we have a collection of stimuli (e.g., criminal cases) and two measures: a random variable X with possible values x_i , i = 1,..,n, (e.g. 7 degrees of guilt); and a random variable Y with possible values y_j , j = 1,…,m (e.g., 7 levels of punishment) under study. A criminal case is randomly selected with replacement from a large set of investigations and presented to the person. Then one of two different conditions is randomly selected for each trial:

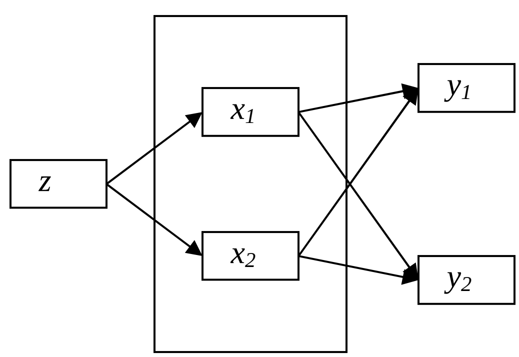
Condition Y: Measure Y alone (e.g. rate level of punishment alone).

Condition XY: Measure X then Y (e.g. rate guilt followed by punishment).

Over a long series of trials (say 100 trials per person to be concrete) each criminal case can be paired with each condition several times. We sort these 100 trials into conditions and pool the results within each condition to estimate the relative frequencies of the answers for each condition. (For simplicity, assume that we are working with a stationary process after an initial practice session that occurs before the 100 experimental trials).

The idea of the experiment is illustrated in Figure 1 below where each measure has only two responses, yes or no. Each trial begins with a presentation of a criminal case. This case places the participant in a state indicated by the little box with the letter z. From this initial state, the individual has to answer questions about guilt and punishment. The large box indicates the first of the two possible measurements about the case. This question appears in a large box because on some trials there is only the second question in which case the question in the large box does not apply. The final stage represents the second (or only) question. The paths indicted by the arrows indicate all the possible answers for two binary valued questions.

Figure 1: Illustration of various possible measurement outcomes for condition XY.



Classic probability theory.

Events. Classic probability theory assigns probabilities to classic events.[1] An event (such as the event $x = X > 4$ or the event $y = Y < 3$ or the event $z = X + Y = 3$) is defined algebraically as a set belonging to a field of sets. There is a null event represented

[1] For simplicity we restrict our attention to experiments that produce a finite number of outcomes.

by the empty set ∅ and a universal event *U* that contains all other events. New events can be formed from other events in three ways. One way is the negation operation, denoted $\sim x$, which is defined as the set complement. A second way is the conjunction operation $x \wedge y$ which is defined by intersection of two sets. A third way is the disjunction operation $x \vee y$ defined as the union of two sets. The events obey the rules of Boolean algebra:

1. Commutative: $x \vee y = y \vee x$
2. Associative: $x \vee (y \vee z) = (x \vee y) \vee z$
3. Complementation: $x \vee (\sim y \wedge y) = x$
4. Absorption: $x \vee (x \wedge y) = x$
5. Distributive: $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

The last axiom, called the distributive axiom, is crucial for distinguishing classic probability theory from quantum probability theory.

Classic Probabilities. The standard theory of probability used throughout the social and behavioral sciences is based on the Kolmogorov axioms:

1. $1 \geq \Pr(x) \geq 0$, $\Pr(\varnothing) = 0$, $\Pr(U) = 1$.
2. If $x \wedge y = \varnothing$ then $\Pr(x \vee y) = \Pr(x) + \Pr(y)$.

When more than one measurement is involved, the conditional probability of *y* given *x* is defined by the ratio:

$$\Pr(y|x) = \Pr(y \wedge x)/\Pr(x),$$

which implies the formula for joint probabilities

$$\Pr(y \wedge x) = \Pr(x) \cdot \Pr(y|x).$$

Classic Probability Distributions.

The simple experiment above is analyzed as follows. Consider first condition XY. We observe $n \cdot m$ distinct mutually exclusive and exhaustive distinct outcomes, such as $x_i y_j$ which occurs when the pair x_i and y_j are observed. Other events can be formed by union such as the event $x_i = x_i y_1 \vee x_i y_2 \vee \ldots \vee x_i y_m$ and the event $y_j = x_1 y_j \vee x_2 y_j \vee \ldots \vee x_n y_j$. New sets can also be defined by the intersection operation for sets such as the event $x_i \wedge y_j = x_i y_j$. These sets obey the rules of Boolean algebra, and in particular, the distributive rule states that $y_j = y_j \wedge U = y_j \wedge (x_1 \vee x_2, \ldots, \vee x_n) = (y_j \wedge x_1) \vee (y_j \wedge x_2), \ldots, \vee (y_j \wedge x_n)$. For binary valued measures (n=m=2), all of the nonzero events are shown in Table 1.

Table 1: Events generated by Boolean Algebra operators.

Events	y_1	y_2	$(y_1 \vee y_2)$
x_1	$x_1 \wedge y_1$	$x_1 \wedge y_2$	$x_1 \wedge (y_1 \vee y_2)$
x_2	$x_2 \wedge y_1$	$x_2 \wedge y_2$	$x_2 \wedge (y_1 \vee y_2)$
$(x_1 \vee x_2)$	$y_1 \wedge (x_1 \vee x_2)$	$y_2 \wedge (x_1 \vee x_2)$	$U = (x_1 \vee x_2) \wedge (y_1 \vee y_2)$

Note: $y_1 \wedge (x_1 \vee x_2) = (x_1 \wedge y_1) \vee (x_2 \wedge y_1)$.

The Boolean rules are used in conjunction with the Kolmogorov rules to derive the *law of total probability* :

$$\Pr(y_j) = \Pr(y_j \wedge U) = \Pr((y_j \wedge (x_1 \vee x_2 \vee \ldots \vee x_n))$$

$$= \Pr((y_j \wedge x_1) \vee (y_j \wedge x_2) \vee \ldots \vee (y_j \wedge x_n))$$

$$= \sum_i \Pr(x_i \wedge y_j) .$$

$$= \sum_i \Pr(x_i) \cdot \Pr(y_j | x_i) .$$

Thus the marginal probability distribution for *Y* is determined from the joint probabilities, and this is also true for *X*. Finally, *Bayes rule* follows from the conditional probability rule, joint probability rule, and the law of total probability:

$$\Pr(y_j|x_i) = \frac{\Pr(y_j \wedge x_i)}{\Pr(x_i)} = \frac{\Pr(y_j) \cdot \Pr(x_i|y_j)}{\sum_k \Pr(y_k) \cdot \Pr(x_i|y_k)} .$$

In our experiment, recall that under one condition we measure *X* then *Y*, but under another condition we only measure variable *Y*. According to classic probability, there is nothing to prevent us from postulating joint probabilities such as $\Pr(x_i \wedge y_j)$ for condition Y, which only involves a single measurement. Indeed, the Boolean axioms require the existence of all the events generated by that algebra. Only y_j is observed, but this observed event is assumed to be broken down into the counterfactual events, $(y_j \wedge x_1) \vee (y_j \wedge x_2) \vee \ldots \vee (y_j \wedge x_n)$. In particular, during condition *Y*, the event $x_i \wedge y_j$ can be considered the counterfactual event that you would have responded at degree of guilt x_i to *X* if you were asked (but you were not), and responding level of punishment y_j when asked about *Y*. Thus all of the joint probabilities $\Pr(x_i \wedge y_j |\mathrm{Y})$ are assumed to exist even when we only measure *Y*. So in the case where only *Y* is measured, we postulate that the marginal probability distribution, $\Pr(y_j)$, is determined from the joint probabilities such as $\Pr(x_i \wedge y_j)$ according to the law of total probability. This is actually a big assumption that is routinely taken for granted in the social and behavioral sciences.

This critical assumption can be understood more simply using Figure 1. Note that under condition Y, the large box containing X is not observed. However, according to classic probability theory, the probability of starting from *z* and eventually reaching y_1 is equal to the sum of the probabilities from the two mutually exclusive and exhaustive paths: the joint probability of transiting from *z* to x_1 and then transiting from x_1 to y_1 plus the joint probability of transiting from z to x_2 and then transiting from x_2 to y_1. How else could one travel from z to y_1 without passing through one of states for x?

If we assume the joint probabilities are the same across conditions, then according to the law of total probability we should find $\Pr(y_j \mid XY) = \Pr(y_j \mid Y)$. Empirically, however, we often find that $\Pr(y_j \mid XY) \neq \Pr(y_j \mid Y)$, and the difference is called an *interference effect* (Khrennikov, 2007). Unfortunately, when these effects occur, as they often do in the social and behavioral sciences, classic probability theory does not provide any way to explain these effects. One is simply forced to postulate a different joint distribution for each experimental condition. This is where quantum probability theory can make a contribution.

Quantum probability Theory.

Events. Quantum theory assigns probabilities to quantum events (see Hughes, 1989, for an elementary presentation). A quantum event (such as L_x representing $X > 4$ or L_y representing $Y < 3$ or the event $z = X + Y = 3$) is defined geometrically as a subspace (e.g. a line or plane or hyperplane, ect.) within a Hilbert space H (i.e., a complex vector space).[2] There is a null event represented by the zero point in the vector space, and the universal event is H itself. New events can be formed in three ways. One way is the negation operation, denoted $L_x^{\perp}$, which is defined as the maximal subspace that is orthogonal to L_x. A second way is the meet operation $x \wedge y$ which is defined by intersection of two subspaces $L_x \wedge L_y$. A third way is the join operation $x \vee y$ defined as the *span* of two subspaces L_x , L_y. Span is quite different than union, and this is where quantum logic differs from classic logic. Quantum logic obeys all of the rules of Boolean

[2] For simplicity, we will only consider finite dimensional Hilbert spaces. Quantum probability theory includes infinite dimensional spaces, but the basic ideas remain the same for finite and infinite spaces.

logic except for the distributive axiom, i.e., it is not necessarily true that $L_z \wedge (L_x \vee L_y) = (L_z \wedge L_x) \vee (L_z \wedge L_y)$.

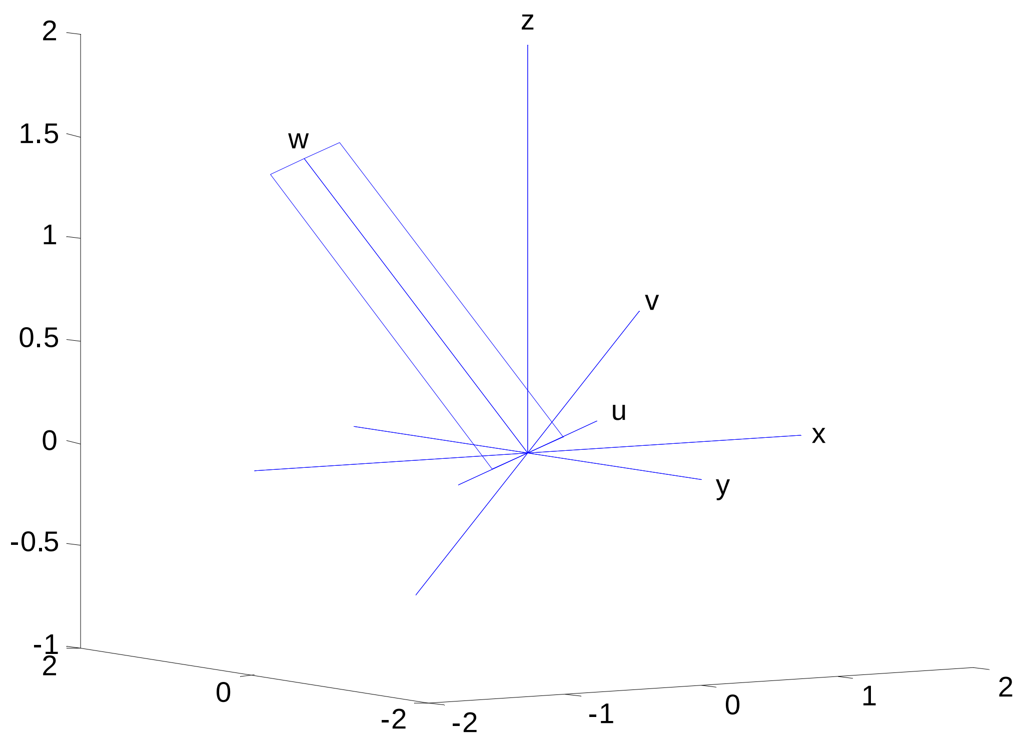
Figure 2 illustrates an example of a violation of the distributive axiom. Suppose H is a 3-dimensional space. This space can be defined in terms of an orthogonal basis formed by the three vectors symbolized $|x\rangle$, $|y\rangle$, and $|z\rangle$ corresponding to the three lines L_x, L_y , L_z in Figure 2.[3] Alternatively, this space can be defined in terms of an orthogonal basis defined by the three vectors $|u\rangle$, $|v\rangle$, and $|w\rangle$ corresponding to the lines L_u , L_v , L_w in Figure 2.[4] Consider the event $(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z)$. Now the event $(L_x \vee L_y \vee L_z)$ spans H and $(L_u \vee L_w)$ is a plane contained within H and so $(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z) = L_u \vee L_w$. According to the distributive axiom, we should have $(L_u \vee L_w) \wedge ((L_x \vee L_y) \vee L_z) = (L_u \vee L_w) \wedge (L_x \vee L_y) \vee (L_u \vee L_w) \wedge L_z$. The first part gives $(L_u \vee L_w) \wedge (L_x \vee L_y) = L_u$ because these two planes intersect along the line L_u. The second part gives $(L_u \vee L_w) \wedge L_z = 0$ because the intersection between the line and the plane is exactly at zero. In sum, we find that

$$(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z) = L_u \vee L_w \neq (L_u \vee L_w) \wedge (L_x \vee L_y) \vee (L_u \vee L_w) \wedge L_z = L_u \vee 0 = L_u.$$

This inequality is a violation by quantum logic of the distributive axiom of Boolean logic.

[3] Dirac notation is used here. The ket $|v\rangle$ corresponds to a column vector, the bra $\langle z|$ corresponds to a row vector, the bra- ket $\langle x|y\rangle$ is an inner product, and $\langle x|\mathrm{P}|y\rangle$ is a bra-matrix-ket product.
[4] $|u\rangle = |x\rangle/\sqrt{2} + |y\rangle/\sqrt{2}$; $|v\rangle = |x\rangle/2 + |y\rangle/2 + |z\rangle/\sqrt{2}$; $|w\rangle = -|x\rangle/2 + |y\rangle/2 + |z\rangle/\sqrt{2}$

Figure 2: Violation of distributive axiom



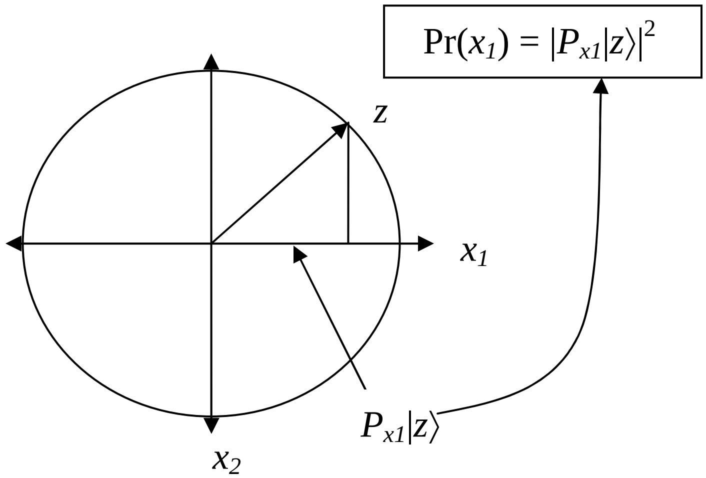
Probabilities. Quantum probabilities are computed using projective rules that involve three steps. First, the probabilities for all events are determined from a *unit* length *state* vector $|z\rangle \in H$, with $| |z\rangle | = 1$. This state vector depends on the preparation and context (person, stimulus, experimental condition). More is said about this state vector later, but for the time being, assume it is known. Second, to each event L_x there is a corresponding projection operator P_x that projects a state vector $|z\rangle$ in H onto L_x.[5] Finally, probability of an event L_x is equal to the squared length of this projection:

$$\Pr(x) = |(P_x|z\rangle|^2 = (P_x|z\rangle)^\dagger(P_x|z\rangle) = \langle z|P_x{}^\dagger P_x|z\rangle = \langle z|P_x \cdot P_x|z\rangle = \langle z|P_x\,|z\rangle.$$

[5] Projection operators are Hermitian and idempotent: $P = P^\dagger = PP$.

Figure 3 illustrates the idea of projective probability: In this figure, the squared length of the projection of $|z\rangle$ onto L_{x1} is the probability of the event L_{x1} given the state $|z\rangle$.

Figure 3: Illustration of projective probability.



Probability distributions for a single variable.

Consider, for the moment, the measurement of a single variable, say the degree of guilt, X, which can produce one of n distinct outcomes or values, x_i $i = 1,...,n$. For this section we assume that each outcome x_i cannot be decomposed or refined into other distinguishable parts. For the time being, we are ignoring the second measure Y. Later we will relax this assumption.

For each distinct outcome x_i we assign a corresponding line or ray, L_{xi}. Corresponding to this subspace is a unit length vector, called a basis state and symbolized as $|x_i\rangle$, that generates this ray by multiplication of a scalar, $a \cdot |x_i\rangle$. Thus the basis states are assumed to be orthonormal: they have inner products $\langle x_i|x_j\rangle = 0$ for all pairs of states, and lengths $|\langle x_i|x_i\rangle| = 1$ for each state. We can interpret the basis state $|x_i\rangle$ as follows: if the person is put into the initial state $|z\rangle = |x_i\rangle$, then you are certain to observe the outcome x_i.

The projector, P_{xi}, projects any point $|z\rangle$ in the Hilbert space onto the subspace L_{xi}, and it is constructed from the outer product $P_{xi} = |x_i\rangle\langle x_i|$. The resulting projection equals $(|x_i\rangle\langle x_i|) \cdot|z\rangle = |x_i\rangle\langle x_i|z\rangle = \langle x_i|z\rangle\cdot|x_i\rangle$ where $\langle x_i|z\rangle$ is the inner product. The inner product, $\langle x_i|z\rangle$ is interpreted as the *probability amplitude* of transiting to state $|x_i\rangle$ from state $|z\rangle$. (In general, this can be a complex number.) The *probability* of any event L_{xi} equals the squared projection, $|P_{xi}\ |z\rangle|^2 = |\ |x_i\rangle\langle x_i|z\rangle\ |^2 = |\ |x_i\rangle\ |^2 \cdot |\ \langle x_i|z\rangle\ |^2 = 1\cdot|\langle x_i|z\rangle|^2 = |\langle x_i|z\rangle|^2$. In other words, the probability of transiting to state $|x_i\rangle$ from state $|z\rangle$ equals the squared magnitude of the probability amplitude, $|\langle x_i|z\rangle|^2$.

The probability of the meet of two events, $x\wedge y$, is equal to the squared length of the projection of the intersection. For example, if $x = x_i\vee x_j$ and $y = x_i\vee x_k$ then $x\wedge y = x_i$ and $\Pr(x\wedge y\ ) = \Pr(x_i) = |\langle x_i|z\rangle|^2$. For $x_i \neq x_j$, the joint event is zero, $L_{xi}\wedge L_{xj} = 0$, and the projection onto zero is zero, so the joint probability is $\Pr(x_i\wedge x_j) = |0|^2 = 0$. The join of two events, say $x_i\vee x_j$, is the span of the two basis vectors, $\{|x_i\rangle, |x_j\rangle\}$, and the projector for this subspace is $P_{xi\vee xj} = P_{xi} + P_{xj} =|x_i\rangle\langle x_i| + |x_j\rangle\langle x_j|$. The probability for the event $x_i\vee x_j$ is simply the sum of the separate probabilities,

$$|P_{xi\vee xj}\ |z\rangle|^2 = |\ (|x_i\rangle\langle x_i| + |x_j\rangle\langle x_j|) \cdot|z\rangle\ |^2 = |\ |x_i\rangle \cdot\langle x_i|z\rangle + |x_j\rangle \cdot\langle x_j|z\rangle\ |^2 = |\langle x_i|z\rangle|^2 + |\langle x_j|z\rangle|^2,$$

where the final step follows from the orthogonality property. Finally, for any $|z\rangle$ we have $P_H \cdot|z\rangle = |z\rangle$ and so $|P_H \cdot|z\rangle|^2 = |\ |z\rangle\ |^2 = |\langle z|z\rangle|^2 = 1$. This also implies that

$$P_H = \sum_i P_{xi} = \sum_i |x_i\rangle\langle x_i| = \mathbf{I},$$

where $\mathbf{I}$ is the identity operator $\mathbf{I}\cdot|z\rangle = |z\rangle$. From these properties we see that quantum probabilities obey rules analogous to the Kolmogorov rules:

1. $1 \geq |P_x|z\rangle|^2 \geq 0$, $\Pr(0) = 0$, $\Pr(H) = 1$.

2. If $L_x \wedge L_y = 0$ then $\Pr(L_x \vee L_y) = \Pr(L_x) + P(L_y)$.

The state vector.

It is time to return to the problem of defining the state vector $|z\rangle$ prior to the measurement. This vector can be expressed in terms of the basis states as follows:

$$|z\rangle = \mathbf{I}\cdot|z\rangle = \left(\sum_i |x_i\rangle\langle x_i| \right) \cdot |z\rangle = \sum_i |x_i\rangle\langle x_i|z\rangle = \sum_i \langle x_i|z\rangle \cdot |x_i\rangle.$$

Thus the initial state vector is a *superposition* of the basis states. The inner product $\langle x_i|z\rangle$ is the coefficient of the state vector corresponding to the $|x_i\rangle$ basis state. To be concrete, one can define $|x_i\rangle$ as a column vector with zeros everywhere except that a one is placed in row i. Then the initial state is a column vector $|z\rangle$ containing coefficient $\langle x_i|z\rangle$ in row i.

The probability of obtaining x_i equals the squared amplitude, $|\langle x_i|z\rangle|^2$. Thus we form the initial state by choosing coefficients that have squared amplitudes equal to the probability of the outcome: choose $\langle x_i|z\rangle$ so that $\Pr(x_i) = |\langle x_i|z\rangle|^2$. In sum, when only one measurement is made, quantum probability theory is not much different than Kolmogorov probability theory.

Effect of measurement.

After one measurement, say X, is taken, and an arbitrary event x is observed, then this measurement changes the state from the initial state $|z\rangle$ to a new state $|x\rangle$ which is the normalized projection on the subspace L_x. In fact, one way to prepare an initial state is to take a measurement, which places the initial state equal to a state consistent with the obtained event. This is called the state reduction or state collapse assumption of quantum theory. Prior to the measurement, the person was in a superposed state, $|z\rangle$, but after measurement the person is in a new state $|x\rangle$. In other words, measurement changes the person.

Social and behavioral scientists generally adopt a classical view of measurement, which assumes that measurement simply records a pre-existing reality. In other words, properties exist in the brain at the moment just prior to a measurement, and the measurement simply reveals this preexisting property. Consider condition Y of our experiment, during which only the punishment level is measured. Even though guilt is not measured in this condition, it is still assumed that the criminal case evokes some specific degree of belief in guilt for the person. We just don't bother to measure its specific value. Thus both properties exist even though we only measure one of them.

The problem with the classical interpretation of measurement can be seen most clearly by reconsidering the example shown in Figure 1 with binary outcomes. If we present a case, then we suppose that it evokes a degree of belief in guilt and a level of punishment. Under condition Y, we only measure the level of punishment. If we measure level y_1, then event $y_1 = y_1 \wedge (x_1 \vee x_2)$ has occurred (here we assume that values x_1, x_2 are mutually exclusive and exhaustive). According to the distributive axiom, this event means that either the person is in the low guilty state and intends to punish at the low level (i.e., $x_1 \wedge y_1$) prior to our measurement (i.e., the brain experienced the upper path in Figure 1), or the person is in the high guilty state and intends to punish at the low level (i.e., $x_2 \wedge y_1$) just prior to our measurement (i.e., the brain experienced the lower path in Figure 1). Condition XY simply resolves the uncertainty about which of these two realities existed at the moment before the measurement.

The classic idea of measurement is rejected in quantum theory (see, e.g. Peres, 1998, p. 14). According to the latter, measurements *create* permanent records that we all can agree upon. To see how this creative process arises in quantum theory, suppose the distributive axiom fails. Referring again to Figure 1, if we measure punishment state y_1, then event $y_1 \wedge (x_1 \vee x_2)$ has occurred. But this does not imply the existence of any specific degree of belief in guilt. We cannot assume that either $(x_1 \wedge y_1)$ or $(x_2 \wedge y_1)$ and not both existed just prior to measurement (i.e., we cannot assume that either the upper path, or the lower path is traveled, see Feynman, Leighton, & Sands, page 9). If we measure *X* first in condition XY, then this measurement will create a state with a specific belief in guilt before measuring the punishment.

In some ways, quantum systems are more deterministic than classical random error systems. Suppose we measure *X* twice in succession, and suppose the first measure produced an event *x*. According to a quantum system, when we measure *X* a second time in succession, we would certainly observe the event *x* again because

$$\Pr(x|x) = |\langle P_x|x\rangle|^2 = |\ |x\rangle\ |^2 = 1.$$

Thus the event remains unchanged until a different type of measurement is taken. If a new type of measurement is taken after the first measurement, then the state changes again, and the outcome becomes probabilistic.

According to a random error system, the observed values are produced by a true score plus some error perturbation that appears randomly on each trial. In that case, the probability of observing a particular value should change following each and every measurement, regardless of whether or not the same measurement is taken twice in succession.

It is interesting to note that social and behavioral scientists are aware of the quantum principle. When they design experiments to obtain repeated measurements for a particular stimulus, they systematically avoid asking participants to judge the same stimulus back to back. Instead, they insert filler items (other measurements) between presentations (to avoid the deterministic result), and these filler items disturb the system to generate probabilistic choice behavior for spaced repetitions of the target items.

Probabilitydistributions for two or more measurements.

After first measuring *X* and observing the event *x*, the state changes to $|x\rangle = P_x|z\rangle/|P_x|z\rangle|$, where P_x is the projector onto the subspace L_x. Note that the squared length of the new state remains equal to one, $|\langle x|x\rangle|^2 = 1$, because of the normalizing factor in the denominator. This is important to maintain a probability distribution over outcomes of *Y* for the next measurement after measuring *X*. The probabilities for the next measurement are based on this new state. If we first measure *X* and observe the event *x*, then the probability of observing *y* when *Y* is measured next equals

$$\Pr(y|x) = |P_y|x\rangle|^2.$$

This updating process continues for each new measurement.

Quantum probability is more general than Kolmogorov probability when more than one measurement is involved. Quantum logic does not have to obey the distributive axiom of Boolean logic when more than one measurement is involved. If more than one measurement is made, then according to quantum theory, the analysis of the experimental situation depends on how one represents the relationship between the measurements *X* and *Y*. There are two possibilities: the measures may be compatible or incompatible.

Compatible Measurements.

Now we consider the problem of two measurements, and we first consider the case in which the two measures are compatible. Intuitively, compatibility means that *X* and *Y* can be measured or accessed or experienced simultaneously or sequentially without interfering with each other. Psychologically speaking, the two measures can be processed in parallel. If the measures are compatible, then we form the basis vectors for the two measurements from all the possible combination of distinct outcomes for *X* and *Y* of the form $x_i y_j$. The complete Hilbert space is defined by $n \cdot m$ orthonormal basis vectors, $|x_i y_j\rangle$, $i = 1,\dots,n$ and $j = 1,\dots,m$, spanning a $n \cdot m$ dimensional space. For example, in condition XY, the vector $|x_i y_j\rangle$ corresponds to observing x_i from *X* and y_j from *Y*. The orthogonal property implies for example that $\langle x_i y_j | x_j y_k\rangle = 0$ and the normal property implies $\langle x_i y_j | x_i y_j\rangle = 1$. This is called the tensor product space for two measures.

Notice that the event x_i is no longer a distinct outcome, and instead it is a course grain outcome that can be decomposed into more refined parts: $L_{xi} = |x_i y_1\rangle \vee |x_i y_2\rangle \vee \dots \vee |x_i y_m\rangle$. Furthermore, the meet $x_i \wedge y_j$ produces the subspace $L_{xi} \wedge L_{yj} = |x_i y_j\rangle$. This implies that the distribution rule holds for this representation:

$$L_{xi} = |x_i y_1\rangle \vee |x_i y_2\rangle \vee \dots \vee |x_i y_m\rangle = (L_{xi} \wedge L_{y1}) \vee (L_{xi} \wedge L_{y2}) \vee \dots \vee (L_{xi} \wedge L_{ym}).$$

Thus Table 1 provides an appropriate description of all the relevant events for binary outcomes. In other words, the assumption of compatible measures requires the existence of all joint events, and the individual outcomes can be obtained from the joint events.

The projection operator for the event $L_{xi,yj}$ is equal to $P_{xi,yj} = |x_i y_j\rangle\langle x_i y_j|$. The projection operator for the event L_{xi} is equal to $P_{xi} = \sum_j P_{xi,yj} = \sum_j |x_i y_j\rangle\langle x_i y_j|$. The projection operator for the event L_{yj} is equal to $P_{yj} = \sum_i P_{xi,yj} = \sum_i |x_i y_j\rangle\langle x_i y_j|$. The orthogonality properties imply

$$|x_i y_j\rangle\langle x_i\, y_j| = (\sum_j |x_i y_j\rangle\langle x_i y_j|) \cdot (\sum_i |x_i y_j\rangle\langle x_i y_j|) = P_{xi} \cdot P_{yj} = P_{yj} \cdot P_{xi}\,.$$

The first equality implies that the projection for the joint event $L_{xi,\,yj}$ can be viewed as a series of two successive measurements or vice-versa. The second equality shows that the projectors for X commute with the projectors for Y; the order of projection does not matter, and both orders project onto the same final subspace. The difference, $[P_{xi} \cdot P_{yj} - P_{yj} \cdot P_{xi}]$, is called the commutator, and it is always zero for compatible measures.

Now let us consider a series of two measurements. Using the reduction principle, if X is measured first and x_i is observed, then the new state after measurement is $|x_i\rangle = P_{xi}|z\rangle/|P_{xi}|z\rangle|$; similarly if Y is measured first and we observe y_j, then the new state after measurement is $|y_j\rangle = P_{yj}|z\rangle/|P_{yj}|z\rangle|$. Consider again the probability of the event $L_{xi,\,yj}$ when viewed as a series of projections.

$$P_{xi,\,yj}|z\rangle = P_{xi} \cdot (P_{yj}\,|z\rangle) = P_{xi} \cdot |y_j\rangle \cdot |\,P_{yj}|z\rangle\,|$$

$$P_{yi,\,xj}|z\rangle = P_{yj} \cdot (P_{xi}\,|z\rangle) = P_{yj} \cdot |x_i\rangle \cdot |\,P_{xi}|z\rangle\,|$$

and

$$\Pr(x_i \wedge y_j) = |\,P_{xi,\,yj}|z\rangle\,|^2 = |\,P_{xi} \cdot |y_j\rangle\,|^2 \cdot |\,P_{yj}|z\rangle\,|^2 = \Pr(x_i|y_j) \cdot \Pr(y_j).$$

$$\Pr(y_j \wedge x_i) = |\,P_{yj,\,xi}|z\rangle\,|^2 = |\,P_{yj} \cdot |x_i\rangle\,|^2 \cdot |\,P_{xi}|z\rangle\,|^2 = \Pr(y_j|x_i) \cdot \Pr(x_i).$$

From the last two equations we obtain the conditional probability rule:

$$\Pr(x_i|y_j) = |\,(P_{xi} \cdot |y_j\rangle\,|^2 = |P_{xi,\,yj}|z\rangle|^2 \,/\, |\,P_{yj}|z\rangle\,|^2\ ,$$

$$\Pr(y_j|x_i) = |\,(P_{yj} \cdot |x_i\rangle\,|^2 = |P_{yj,\,xi}|z\rangle|^2 \,/\, |\,P_{xi}|z\rangle\,|^2\ .$$

Note that in general, $|\,P_x|z\rangle\,|^2 \neq |\,P_y|z\rangle\,|^2$, and so $\Pr(y_j|x_i) \neq \Pr(x_i|y_j)$.

The projection onto L_{xi} is $P_{xi} \cdot |z\rangle = \sum_j |x_i y_j\rangle\langle x_i y_j|z\rangle$ and the probability of this event equals

$$\Pr(x_i) = \sum_j |\langle x_i y_j|z\rangle|^2 = \sum_j |P_{xi} \cdot |y_j\rangle|^2 \cdot |P_{yj}|z\rangle|^2 .$$

The above expression is the quantum version of the law of total probability. From the above facts, we can derive a quantum analogue of Bayes rule:

$$\Pr(y_j|x_i) = |P_{yj} \cdot |x_i\rangle|^2 = \frac{\left|P_{x_i}|y_j\rangle\right|^2 \cdot \left|P_{y_j}|z\rangle\right|^2}{\sum_k \left|P_{x_i}|y_k\rangle\right|^2 \cdot \left|P_{y_k}|z\rangle\right|^2} .$$

Let us re-examine the initial state vector $|z\rangle$ for the case of two compatible measurements. As before, this state vector can be described in terms of the basis vectors:

$$|z\rangle = \mathbf{I} \cdot |z\rangle = (\sum_i \sum_j |x_i y_j\rangle\langle x_i y_j|) \cdot |z\rangle = \sum_i \sum_j \langle x_i y_j|z\rangle \cdot |x_i y_j\rangle.$$

Once again, we see that the initial state is a superposition of the basis states. The inner product $\langle x_i y_j|z\rangle$ is the coefficient of the state vector corresponding to the $|x_i y_j\rangle$ basis state. The probability of obtaining the joint event $x_i y_j$ equals the squared amplitude of the corresponding coefficient, $|\langle x_i y_j|z\rangle|^2$. Thus we form the initial state by choosing coefficients that have squared amplitudes equal to the probability of the joint outcome: choose $\langle x_i y_j|z\rangle$ so that $\Pr(x_i y_j) = \Pr(x_i \wedge y_j) = |\langle x_i y_j|z\rangle|^2$.

In sum, all of these results exactly correspond to the classic probability rules. In short, quantum probability theory reduces to classic probability theory for compatible measures. If all measures were compatible, then quantum probability would produce exactly the same results as classical probability.[6]

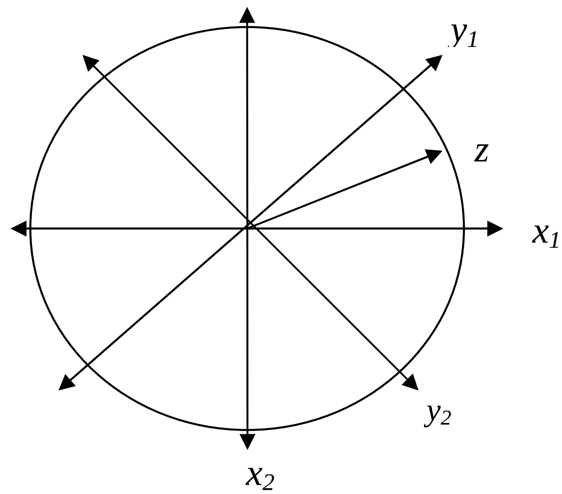
[6] This is not quite true. We are disregarding change caused dynamic laws, and we are only focusing on change caused by measurement at this point.

Incompatible measures.

Incompatibility means that *X* and *Y* cannot be measured or accessed or experienced simultaneously. Psychologically speaking, the two measures must be processed serially, and measurement of one variable interferes with the other. This implies that *X* produces *n* distinct outcomes x_i $i = 1,\ldots,n$ that cannot be decomposed into more refined parts, because we can't simultaneously measure *Y*. Also, *Y* produces *n* distinct outcomes y_i $i = 1,\ldots,n$ that cannot be decomposed into more refined parts, because we can't simultaneously measure *X*. In this case, we assume that the outcomes from the measure *X* produce one orthonormal set of basis states, $|x_i\rangle$, $i = 1,\ldots,n$; and the outcomes of *Y* produce another orthonormal set of basis states $|y_j\rangle$, $j = 1,\ldots,n$. To account for the fact that one measure influences the other, it is assumed that one set of basis states is a linear transformation of the other. Thus we now have two different bases for the *same* *n*-dimensional Hilbert space.

This idea is illustrated in Figure 4 shown below. In this figure, we assume that the outcomes are binary. The outcomes of the first measure (regarding the guilt) are represented by the basis vectors $|x_1\rangle$ and $|x_2\rangle$, and the outcomes of the second measure (regarding the punishment) are represented by the basis vectors $|y_1\rangle$ and $|y_2\rangle$. Note that the basis vectors for the *Y* measure are an orthogonal rotation of the basis vectors for the *X* measure (and vice-versa). One can either use the $|x_1\rangle$ and $|x_2\rangle$ basis to describe the state vector $|z\rangle$ or use the $|y_1\rangle$ and $|y_2\rangle$ basis to describe this same state but they cannot be used at the same time.

Figure 4: Illustration of rotated basis vectors for incompatible measurements



One cannot experience or measure both variables simultaneously because if one measures *X*, then one needs to project the state $|z\rangle$ on to the *X* basis rather than the *Y* basis. If one measures *X* and finds the value x_1 then the outcome for the next measurement of *Y* must be uncertain ($\Pr(y_j) = |\langle y_j|x_1\rangle|^2$). Similarly, if one measures *Y*, then the *Y* bases must be used. Also if *Y* is measured first and the value y_1 is observed, then the outcome for the next measurement on *X* must be uncertain ($\Pr(x_i) = |\langle x_i|y_1\rangle|^2$). It is impossible to be certain about both values simultaneously! Therefore, it is impossible to completely measure all the values of the system. This is essentially the idea behind the famous Heisenberg uncertainty principle (Peres, 1995, Ch. 2).

The distributive axiom of Boolean logic is violated with incompatible measures. For example, considering Figure 4, note that $|x_i\rangle\wedge|y_j\rangle = 0$ for all *i* and *j*.

$$L_{y1} = L_{y1}\wedge(L_{x1}\vee L_{x2}) \neq (L_{y1}\wedge L_{x1})\vee(L_{y1}\wedge L_{x2}) = 0 \vee 0 = 0.$$

Because of incompatibility, the event $L_{y1}\wedge L_{x1}$ is impossible and so is the event $L_{y1}\wedge L_{x2}$, but the event L_{y1} is clearly possible. This is where quantum probability deviates from classic probability. Table 2 shows the events for incompatible measures, clearly showing violations of the distributive axiom.

Table 2: Events generated by incompatible measures.

Events	y_1	y_2	$y_1 \vee y_2$
x_1	$\varnothing$	$\varnothing$	$(y_1 \vee y_2) \wedge x_1$
x_2	$\varnothing$	$\varnothing$	$(y_1 \vee y_2) \wedge x_2$
$x_1 \vee x_2$	$(x_1 \vee x_2) \wedge y_1$	$(x_1 \vee x_2) \wedge y_2$	$(x_1 \vee x_2) \wedge (y_1 \vee y_2)$

Note: $L_{y1} = L_{y1} \wedge (L_{x1} \vee L_{x2}) \neq (L_{y1} \wedge L_{x1}) \vee (L_{y1} \wedge L_{x2}) = 0$

To get a deeper understanding of the violation of the distributive axiom, let us return to Figure 1 again. Suppose only Y is measured, and we observe y_1. How does the person go from the initial state $|z\rangle$ to this observed state $|y_1\rangle$? We *cannot* say 'The person traveled one of two paths: either the $|z\rangle \rightarrow |x_1\rangle \rightarrow |y_1\rangle$ path or the $|z\rangle \rightarrow |x_1\rangle \rightarrow |y_2\rangle$, but we are uncertain about which path was taken.' In other words, if the person intends to punish at the low level, then we cannot say he or she reached that decision either by concluding that the person was guilty at a low degree or by concluding that the person was guilty at a high degree. When we do not measure guilt, we cannot assume that the person is definitely in one of these two guilt states, and instead the person is indefinite (or superposed) between these two states. When we do not observe what happens, quantum theory allows for a more general type of uncertainty regarding state changes as compared to classical probability theory.

The fact that there are two different bases for the same Hilbert space implies that the same state vector can be described by two different bases:

$|z\rangle = \mathbf{I}\cdot|z\rangle = (\sum_i |x_i\rangle\langle x_i|\ )\cdot|z\rangle = \sum_i |x_i\rangle\langle x_i|z\rangle,$

$|z\rangle = \mathbf{I}\cdot|z\rangle = (\sum_i |y_i\rangle\langle y_i|\ )\cdot|z\rangle = \sum_i |y_i\rangle\langle y_i|z\rangle.$

If the X basis is used to describe the state vector $|z\rangle$, then the inner products, $\langle x_i|z\rangle$, form the coordinates for $|z\rangle$. We can represent the initial state vector in this basis by a column vector, ψ, with $\langle x_i|z\rangle$ in row i. The marginal probability distribution for X is $\Pr(x_i) = |\psi_i|^2 = |\langle x_i|z\rangle|^2$. But if the Y basis is used to describe the state vector $|z\rangle$, then the inner products, $\langle y_i|z\rangle$, form the coordinates for $|z\rangle$. We can represent the initial state vector in this basis by a column vector, ϕ, with $\langle y_i|z\rangle$ in row i. The marginal probability distribution for Y is $\Pr(y_j) = |\phi_i|^2 = |\langle y_j|z\rangle|^2$. No joint distribution exists, but both marginal distributions are derived from a common state vector $|z\rangle$. The equality of the two representations implies

$\sum_i |x_i\rangle\langle x_i|z\rangle = \sum_i |y_i\rangle\langle y_i|z\rangle,$

$\rightarrow \langle x_j|\sum_i |x_i\rangle\langle x_i|z\rangle = \langle x_j|\sum_i |y_i\rangle\langle y_i|z\rangle,$

$\rightarrow \sum_i \langle x_j|x_i\rangle\langle x_i|z\rangle = \sum_i \langle x_j|y_i\rangle\langle y_i|z\rangle,$

$\rightarrow \langle x_j|z\rangle = \sum_i \langle x_j|y_i\rangle\langle y_i|z\rangle,$

which is the linear transformation that maps coefficients of the state described by the Y basis into coefficients of the state described by the X basis. The inner product, $\langle x_j|y_i\rangle = \langle y_i|x_j\rangle^*$, is the probability amplitude of transiting to the $|x_j\rangle$ state from the $|y_i\rangle$ state. The squared amplitude, $|\langle x_j|y_i\rangle|^2$ equals the probability of observing x_j on the next measurement of X given that y_i was previously obtained from a measure of Y. A similar argument produces

$\sum_i \langle y_j|x_i\rangle\langle x_i|z\rangle = \langle y_j|z\rangle$, which is the linear transformation that maps coefficients of the state described by the *X* basis into coefficients of the state described by the *Y* basis. The inner product, $\langle y_j | x_i \rangle$, is the probability amplitude of transiting to the $|y_j\rangle$ state from the $|x_i\rangle$ state. The squared amplitude, $|\langle y_j | x_i \rangle|^2$ equals the probability of observing y_j on the next measurement of *Y* given that x_i was previously obtained from a measure of *X*.

In sum, one constructs the first marginal distribution from (a) the inner products such as $\langle x_i|z\rangle$ relating the initial state to the states for the first basis; and (b) the second marginal distribution is constructed from the inner products such as $\langle y_j | x_i \rangle$ relating the states from the first basis to the states of the second basis.

The inner products relating one basis to another must satisfy several important constraints. First, the fact that $|\langle x_j | y_i \rangle|^2 = |\langle y_i | x_j \rangle|^2$ implies that incompatible measurements must satisfy $\Pr(x_j | y_i) = \Pr(y_i | x_j)$, which is called the law of reciprocity (Peres, 1995, p. 34). Of course, classic probability does not need to satisfy this constraint. It is important to note that the law of reciprocity only holds for transitions between basis states. It does not hold for more general (course grained) events.

Second, consider the matrix of coefficients, U , with element $\langle y_i | x_j \rangle$ in row *i* and column *j* representing the transition to state $|y_j\rangle$ from state $|x_i\rangle$. Then the initial state described with respect to the *X* basis, ψ, is related to the initial state described with respect to the Y basis, ϕ, by the linear transformation $\phi = U \cdot \psi$. Similarly, the initial state described with respect to the Y basis, ϕ, is related to the initial state described with respect to the *X* basis, ψ, by the linear transformation $\psi = U^{\dagger} \cdot \phi$. Notice that $\phi = UU^{\dagger} \cdot \phi$ and $\psi = U^{\dagger} U \cdot \psi$. Thus this matrix must be unitary, $U^{\dagger} U = \mathrm{I} = UU^{\dagger}$, where I is the identity matrix. This unitary property guarantees that the transformation preserves the lengths of the vectors (to one) before and after transformation. The unitary property implies that the transition matrix T, with elements $|\langle y_i | x_j \rangle|^2$ in row i and column j, must be doubly stochastic. That is, both the rows and columns of T must sum to unity, which is called the *doubly stochastic law* (Peres, 1995, p. 33). The transition matrix for classic probability theory must be stochastic (only the columns must sum to one), but it does not need to be doubly stochastic.

Thus quantum probabilities for incompatible measures must obey two laws that are not required by classic probability: the *law of reciprocity* and the *doubly stochastic law*. Classic probability must obey the *law of total probability* which is not required by quantum probabilities for incompatible measures. These three properties can be used to empirically distinguish quantum versus classical models.

The projection operator for the event L_{xi} is $P_{xi} = |x_i\rangle\langle x_i|$ and the projector for the event L_{yj} is $P_{yj} = |y_j\rangle\langle y_j|$. It is interesting to compare the two projections produced by measuring Y first followed by X:

$$P_{xi} \cdot P_{yj} = |x_i\rangle\langle x_i|\cdot|y_j\rangle\langle y_j| = \langle x_i|y_j\rangle|x_i\rangle\langle y_j|$$

with the two produced by the measuring X first followed by Y:

$$P_{yj} \cdot P_{xi} = |y_j\rangle\langle y_j|\cdot|x_i\rangle\langle x_i| = \langle y_j|x_i\rangle|y_j\rangle\langle x_i|.$$

The difference, called the commutator, is

$$P_{xi} \cdot P_{yj} - P_{yj} \cdot P_{xi} = \langle x_i|y_j\rangle|x_i\rangle\langle y_j| - \langle y_j|x_i\rangle|y_j\rangle\langle x_i|,$$

which is nonzero for some i and j. This implies that different orders of measurement can produce different final projections and thus different probabilities. In other words, quantum probability provides a theory for explaining *order effects* on measurements, a problem which is replete throughout the social and behavioral sciences.

Let us now examine the event probabilities in the case of incompatible measures. Here we have to carefully analyze the different experimental conditions separately. First consider condition XY. In this case we have

$$\Pr(y_j \wedge x_i|\mathrm{XY}) = |\, P_{yj} P_{xi} \cdot |z\rangle\, |^2 = |\langle y_j|x_i\rangle|^2 \cdot |\langle x_i|z\rangle|^2 = \Pr(y_j|x_i,\, \mathrm{XY}) \cdot \Pr(x_i|\mathrm{XY}),$$

so that

$$\Pr(y_j|\mathrm{XY}) = \sum_i |\langle x_i|z\rangle|^2 \cdot |\langle y_j|x_i\rangle|^2$$

similar to that found with classic probability and with compatible measurements.

To get a more intuitive idea, refer again to Figure 1. The probability of responding x_1 to question X on the first measure equals the squared probability amplitude of transiting from the initial state $|z\rangle$ to the basis vector $|x_1\rangle$, which equals $\Pr(x_1|\mathrm{XY}) = |\langle x_1|z\rangle|^2$. Given that first measurement produces x_1, and the state now equals $|x_1\rangle$, the probability of responding $Y = y_1$ to second question is equal to the squared probability amplitude of transiting from $|x_1\rangle$ to $|y_1\rangle$, which equals $\Pr(y_1|x_1,\mathrm{XY}) = |\langle y_1|x_1\rangle|^2$. The probability of observing $X = x_1$ on the first test and then $Y = y_1$ on the second test equals $\Pr(x_1|\mathrm{XY}) \cdot \Pr(y_1|x_1,\mathrm{XY}) = |\langle x_1|z\rangle|^2 \cdot |\langle y_1|x_2\rangle|^2$. A similar analysis produces $\Pr(x_2|\mathrm{XY}) \cdot \Pr(y_1|x_2,\mathrm{XY}) = |\langle x_2|z\rangle|^2 \cdot |\langle y_1|x_2\rangle|^2$ for the probability of observing $X = x_2$ on the first test and then $Y = y_1$ on the second test. Thus the probability of observing $Y = y_1$ on for the XY condition equals $\Pr(y_1|\mathrm{XY}) = |\langle x_1|z\rangle|^2 \cdot |\langle y_1|x_1\rangle|^2 + |\langle x_2|z\rangle|^2 \cdot |\langle y_1|x_2\rangle|^2$.

Next consider first the probability of responding to question Y alone. The projection of the initial state onto the L_{yj} event is

$P_{yj} \cdot |z\rangle = |y_j\rangle\langle y_j| \cdot |z\rangle = |y_j\rangle\langle y_j|z\rangle$, and so

$$\Pr(y_j|\mathrm{Y}) = \langle z|y_j\rangle\langle y_j|y_j\rangle\langle y_j|z\rangle = |\langle y_j|z\rangle|^2 .$$

More intuitively, this is obtained from the squared amplitude of transiting from the initial state $|z\rangle$ to the basis vector $|y_j\rangle$ without measuring or knowing anything about the first question. Expansion of the identity operator produces the following interesting results:

$$\Pr(y_j|\mathrm{Y}) = |\langle y_j|z\rangle|^2 = |\langle y_j|\mathbf{I}|z\rangle|^2 = |\sum_i \langle y_j|(|x_i\rangle\langle x_i|)|z\rangle|^2 = |\sum_i \langle y_j|x_i\rangle\langle x_i|z\rangle|^2 .$$

Thus we find that $\Pr(y_j|\mathrm{Y}) \neq \Pr(y_j|\mathrm{XY})$ and this difference can explain interference effects. Let us analyze the interference effect in more detail for the special case shown in Figure 1 in which there are only two outcomes for each measure.

$$|\langle y_1|z\rangle|^2 = (\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle)(\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle)^*$$

$$= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + \langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle^*\langle x_2|z\rangle^* + \langle y_1|x_2\rangle\langle x_2|z\rangle\langle y_1|x_1\rangle^*\langle x_1|z\rangle^*$$

$$= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2$$

$$+ |\langle y_1|x_1\rangle|\cdot|\langle x_1|z\rangle|\cdot|\langle y_1|x_2\rangle|\cdot|\langle x_2|z\rangle|\cdot(e^{i\cdot\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle\langle x_2|z\rangle} + e^{-i\cdot\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle\langle x_2|z\rangle})$$

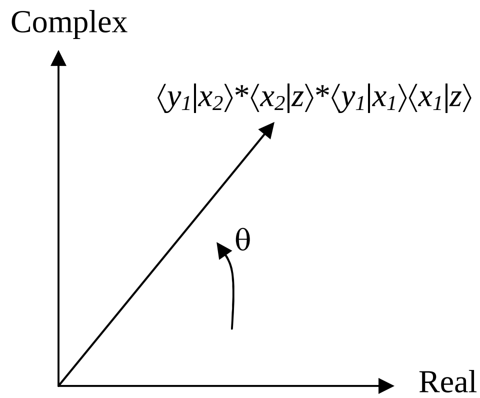
$$= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2$$

$$+ |\langle y_1|x_1\rangle|\cdot|\langle x_1|z\rangle|\cdot|\langle y_1|x_2\rangle|\cdot|\langle x_2|z\rangle|\cdot(Cos(\theta) + i\cdot Sin(\theta) + Cos(\theta) - i\cdot Sin(\theta))$$

$$= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + 2|\langle y_1|x_1\rangle|\cdot|\langle x_1|z\rangle|\cdot|\langle y_1|x_2\rangle|\cdot|\langle x_2|z\rangle|\cdot Cos(\theta),$$

where θ is the angle of the complex number $\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle^*\langle x_2|z\rangle^*$ in the complex plane (see Figure 5 below). If we restrict the probability amplitudes to real numbers, then we are restricted to the horizontal line in Figure 4, so that $\theta = 0$ or π, and $Cos(\theta) = \pm 1$.

Figure 5: The angle between probability amplitudes.



Note that the first two terms in the above expression for $\Pr(y_1|\mathrm{Y})$ exactly match those found when computing $\Pr(y_1|\mathrm{XY})$. If the cosine in the third term is zero, then $\Pr(y_1|\mathrm{Y}) - \Pr(y_1|\mathrm{XY}) = 0$ and there would be no interference. Thus the difference $\Pr(y_1|\mathrm{Y}) - \Pr(y_1|\mathrm{XY})$ is contribute solely by the cosine term, which is called the interference term. Here we see the uniquely quantum prediction of interference effects for incompatible measures.

Quantum probability provides a more coherent and elegant explanation interference effects than classic probability theory. The former uses a single interference coefficient (θ) to relate the two marginal distributions, $\Pr(y_1|\mathrm{Y})$ and $\Pr(y_1|\mathrm{XY})$. Whereas the classic probability theory postulates two separate joint probability distributions and derives the marginals for each condition from these separate joint distributions.

It is also worthwhile to compare the probabilities of the binary valued responses for condition XY with YX:

$$\Pr(x_1 \wedge y_1|\mathrm{XY}) = |\, P_{y1}P_{x1} \cdot |z\rangle\, |^2 = |\langle x_1|z\rangle|^2 \cdot |\langle y_1|x_1\rangle|^2$$

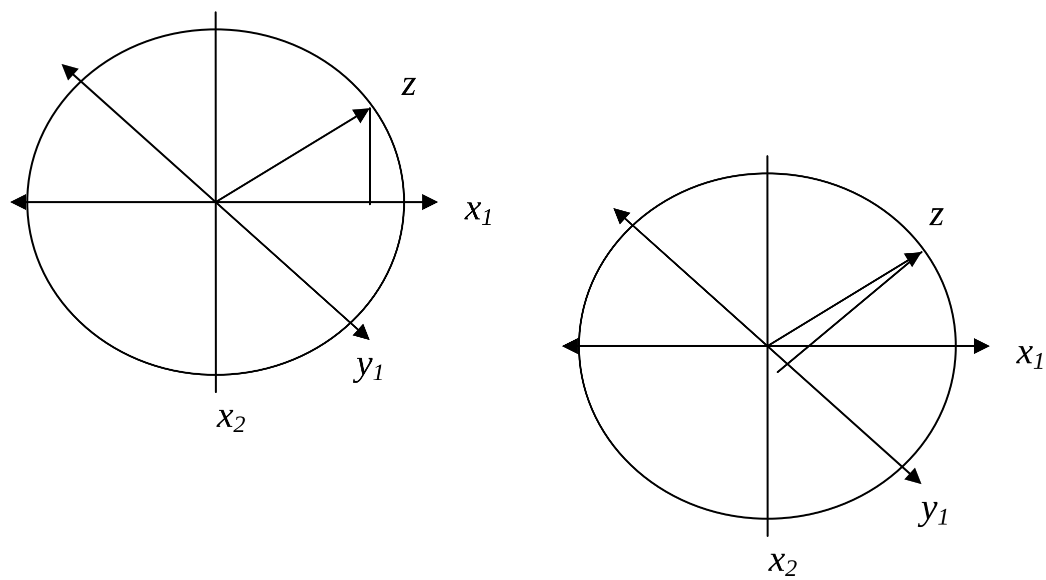
$$\Pr(y_1 \wedge x_1|\mathrm{YX}) = |\, P_{x1}P_{y1} \cdot |z\rangle\, |^2 = |\langle y_1|z\rangle|^2 \cdot |\langle x_1|y_1\rangle|^2.$$

Note that $|\langle y_1|x_1\rangle|^2 = |\langle x_1|y_1\rangle|^2$ and so

$$\Pr(x_1 \wedge y_1|XY) - \Pr(y_1 \wedge x_1|YX) = |\langle x_1|y_1\rangle|^2 \cdot (|\langle x_1|z\rangle|^2 - |\langle y_1|z\rangle|^2),$$

which differs from zero as long as $|\langle x_1|z\rangle|^2 \neq |\langle y_1|z\rangle|^2$. An illustration of these two different projections is illustrated in Figure 6 below. Once again, quantum theory provides a direct explanation for the relation between the distributions produced by the two conditions, whereas classic probability theory needs to assume an entirely new probability distribution for each condition.

Figure 6: Projections for the initial state on two different basis vectors.



Finally it is interesting to re-examine the conditional probabilities for incompatible measures.

$$\Pr(y_1|x_1,XY) = |\langle y_1|x_1\rangle|^2 = |\langle x_1|y_1\rangle|^2 = \Pr(x_1|y_1,YX).$$

This law of reciprocity places a very strong constraint on the quantum probability theory. This relation only holds, however, for complete measures that involve transitions from one basis state to another. It is no longer true for course measurements which are disjunctions of several basis vectors.

Why Complex Numbers?

Consider again Figure 1 which involves binary outcomes for each measure. If we are restricted to real valued probability amplitudes, then we obtain the following simplification of our basic theoretical result for incompatible measures:

$$\Pr(y_1|\mathrm{Y}) = |\langle y_1|x_1\rangle|^2\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle|^2\langle x_2|z\rangle|^2 \pm 2\cdot|\langle y_1|x_1\rangle||\langle x_1|z\rangle||\langle y_1|x_2\rangle||\langle x_2|z\rangle|.$$

The interference term is now simply determined by the sign and magnitude of $|\langle u|x\rangle||\langle x|z\rangle||\langle u|y\rangle||\langle y|z\rangle|$.

Complex probability amplitudes can be shown to be needed under the following conditions and results. Suppose we can perform variations on our basic experiment by changing some experimental factor, F. Suppose we find that changing the experimental factor, from level F_1 to level F_2, produces the same sign of the interference effect, but increases the magnitude of the interference:

$$|\Pr(y_1|\mathrm{Y},F_2) - \Pr(y_1|\mathrm{XY},F_2)| > |\Pr(y_1|\mathrm{Y},F_1) - \Pr(y_1|\mathrm{XY},F_1)|$$

Also suppose that this same manipulation does not change the joint probabilities so that

$$\Pr(x_1\wedge y_1|\mathrm{XY},F_1) = |\langle y_1|x_1\rangle|^2\langle x_1|z\rangle|^2 = \Pr(x_1\wedge y_1|\mathrm{XY},F_2),$$

$$\Pr(x_2\wedge y_1|\mathrm{XY},F_1) = |\langle y_1|x_2\rangle|^2\langle x_2|z\rangle|^2 = \Pr(x_2\wedge y_1|\mathrm{XY},F_2).$$

Then these results imply that changes in this factor F leave $|\langle y_1|x_1\rangle\langle x_1|z\rangle||\langle y_2|x_2\rangle\langle x_2|z\rangle|$ constant and vary $Cos(\theta)$ instead.

Consider the example from physics called the paradox of recombined beams (French & Taylor, 1978, p. 295-296, also refer to Figure 1). In this experiment, a plane polarized photon (z) is shot through a quarter wave plate to produce a circularly polarized photon. There are two possible channel outputs for the quarter wave plate, a left clockwise or right clockwise rotation (labeled x_1 = left or x_2 = right in Figure 1). A final detector determines whether the output from the quarter wave plat can be detected (symbolized y_1 in Figure 1) by a linear polarized detected rotated at angle ϕ with respect to the original state of the photon. In this situation, the critical factor F which is manipulated is the angle ϕ between the initial and final linear polarization.

The two channel outputs from the quarter wave plate form two ortho-normal bases, $|x_1\rangle$ and $|x_2\rangle$, for representing the state. The probability amplitude of transiting from the initial state to the final state equals

$$\langle y_1|z\rangle = \langle y_1|\mathbf{I}|z\rangle = \langle y_1|(|x_1\rangle\langle x_1| + |x_2\rangle\langle x_2|)|z\rangle = \langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle.$$

When the right channel is closed, then probability of passing through the left channel is $|\langle x_1|z\rangle|^2 = ½$, and the probability of detection is also $|\langle y_1|x_1\rangle|^2 = ½$. The same is true when the left channel is closed: then the probability of passing through right channel is $|\langle x_2|z\rangle|^2 = ½$ and the probability of detection is $|\langle y_1|x_2\rangle|^2 = ½$. When both channels are open, then the probability of detection is $Cos(\phi)^2$. Therefore, we have 5 equations and four unknowns ($\langle x_1|z\rangle$, $\langle y_1|x_1\rangle$, $\langle x_2|z\rangle$, $\langle y_1|x_2\rangle$):

$$|\langle x_1|z\rangle| = |\langle y_1|x_1\rangle| = |\langle x_2|z\rangle| = |\langle y_1|x_2\rangle| = \sqrt{½}\ ,$$

$$\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle = Cos(\phi).$$

The first set of equations does not depend on ϕ, but the last one does. Therefore, we are forced to find a solution using complex numbers. In this case, the solutions are $\langle x_1|z\rangle = 1/\sqrt{2} = \langle x_2|z\rangle$, $\langle y_1|x_1\rangle = e^{-i\phi}/\sqrt{2}$, $\langle y_1|x_2\rangle = e^{i\phi}/\sqrt{2}$.

What is the difference between probability mixture and superposition?

A superposition state is a linear combination of the basis states for a measurement. The initial state $|z\rangle$ is not restricted to just one of the basis states. According to quantum logic, if L_{x1} is an event corresponding to the observation of x_1 and

L_{x2} is another event corresponding to the observation *y*, then we can form a new disjunction event $L_{x1} \vee L_{x2}$ which is the set of all linear combinations

$$|z\rangle = a \cdot |x_1\rangle + b \cdot |x_2\rangle \text{ , where } |a|^2 + |b|^2 = 1.$$

In the above case, the initial state, $|z\rangle$, would be in a superposition state with respect to the basis states for measure A. In this case we observe the value x_1 with probability $|\langle x_1|z\rangle|^2 = |a|^2$ and we observe the value x_2 with probability $|\langle x_2|z\rangle|^2 = |b|^2$.

It is difficult to interpret the superposition state. The following interpretation is invalid and is only applicable to classic states: immediately before measurement, you are either in state $|x_1\rangle$ with probability $|a|^2$ or you are in state $|x_2\rangle$ with probability $|b|^2$. The latter describes a classic *mixed* state rather than a quantum *superposition* state. Quantum theory allows for both mixed states and superposition states, but these two types make distinctive probability predictions (an example is provided at the end of this section). There is no well agreed upon psychological interpretation of a superposition state and the interpretation of this concept has produced great controversy (Schroedinger's cat problem). But it is something like a *fuzzy* and *uncertain* representation of a state.

To clearly see the difference between a mixed state and a superposition state, consider the following example. Consider again Figure 1 with binary outcomes. Suppose the two bases are related to each other as follows:

$$|y_1\rangle = (|x_1\rangle + |x_2\rangle)/\sqrt{2},$$

$$|y_2\rangle = (|x_1\rangle - |x_2\rangle)/\sqrt{2},$$

$$|x_1\rangle = (|y_1\rangle + |y_2\rangle)/\sqrt{2},$$

$$|x_2\rangle = (|y_1\rangle - |y_2\rangle)/\sqrt{2}.$$

After a measurement of $Y = y_1$, we are in state $|y_1\rangle = (|x_1\rangle + |x_2\rangle)/\sqrt{2}$; from this state we find $\Pr(x_1) = \Pr(x_2) = .50$, and $\Pr(y_1) = 1$, $\Pr(y_2) = 0$. This can be distinguished from the following mixed state: there is a .50 probability that we are in basis state $|x_1\rangle$, and there is a .50 probability that we are in basis state $|x_2\rangle$. Note that if you are in state $|x_1\rangle$, then y_1 and y_2 are equally likely measurements for Y; and if you are in state $|x_2\rangle$, then y_1 and y_2 are again equally likely measurements for Y. Thus the mixed state produces $\Pr(x_1) = \Pr(x_2) = \Pr(y_1) = \Pr(y_2) = .50$, which differs dramatically from the probabilities produced by the superposition state. In sum, an equal mixture of $|x_1\rangle$ and $|x_2\rangle$ does not produce the same results as an equally weighted superposition of $|x_1\rangle$ and $|x_2\rangle$. However, to reveal this difference, it is necessary to obtain probabilities from two different incompatible measures, X and Y.

Concluding Comments

Quantum probability was discovered by physicists in the early 20th century solely for applications to physics. But Von Neumann axiomatized the theory and discovered that it implied a new logic, quantum logic, and a new probability, quantum probability. Just as the mathematics of differential equations spread from purely physical applications in Newtonian mechanics to applications throughout the social and behavioral sciences, it is very likely that the mathematics of quantum probability will also see new applications in the social and behavioral sciences. Such applications have already begun to appear in areas including information retrieval, language, concepts, decision making, economics, and game theory (see Bruza, Lawless, van Rijsbergen, & Sofge, 2007; Bruza, Lawless, van Rijsbergen, & Sofge, 2008; also see the Special Issue on Quantum Cognition and Decision to appear in Journal of Mathematical Psychology in 2008).

Quantum probability reduces to classical probability when all the measures are compatible. But quantum probability departs dramatically from classical probability when the measures are incompatible. In particular, quantum probabilities do not have to obey the law of total probability as required by classical probabilities. Thus one can view quantum probability as a generalization of classical probability with the inclusion of incompatible measures. However, there are several important restrictions on quantum probabilities for incompatible measures. In this case the quantum probabilities must obey the law of reciprocity and the doubly stochastic law, which classical probabilities do not have to obey.

There are several advantages for using a quantum probability approach over a classical probability approach. First, the quantum approach does not always require or need to assume a joint probability space to derive and relate marginal probabilities from different measures. Marginal probabilities from different measures can all be derived from a common state vector without postulating a common joint distribution. Second, quantum probability theory provides an explanation for order effects on measurements, which is a pervasive problem in the social and behavioral sciences. Third, quantum probability provides an explanation for the interference effect that one measure has on another measure, which is another pervasive problem of measurements in the social and behavioral sciences. Finally, quantum probabilities allow for deterministic as well as probabilistic behavior, which matches human behavior better than random error theories.

Quantum probability theory is a new and exciting field of mathematics with many interesting and potentially useful applications to the social and behavioral sciences. The intention of this chapter was to show the simplicity, coherence, and generality of quantum probability theory.

References

Bruza, P. D., Lawless, W., van Rijsbergen, C.J., Sofge, D., Editors. (2007) *Proceedings of the AAAI Spring Symposium on Quantum Interaction, March 27-29. Stanford University, 2007*. AAAI Press.

Bruza, P. D., Lawless, W., van Rijsbergen, C.J., Sofge, D., Editors. (2008) *Proceedings of the second conference on Quantum Interactions, March 26-28, 2008, Oxford University.*

Feynman, R. P., Leighton, R. B., & Sands, M. (1966) *The Feynman Lectures on Physics*: *Volume III.* Reading MA: Addison Wesley.

Hughes, R. I. G. (1989) *The structure and interpretation of Quantum mechanics.* Cambridge, MA: Harvard University Press.

Khrennikov, A. (2007) Can quantum information be processed by macroscopic systems? *Quantum Information Theory*, in press.

Nielsen, M. A. & Chuang, I. L. (2000) *Quantum computation and Quantum information*. Cambridge, UK: Cambridge University Press.

Peres, A. (1995) Quantum theory: concepts and methods. Kluwer Academic: Dordrecht.

Sakurai, J. J. (1994) Modern quantum mechanics. Pearson Education Inc.