Qualitatiive Mathematics for Social Sciences
Mathematical Models for Research
on Cultural Dynamics
Edited by L. Rudolph and J. Valsiner
Published in Routledge's Behavioral Science Monographs
Series, 2012

3

Introduction to Quantum Probability for Social and Behavioral Scientists

Jerome R. Busemeyer

This chapter has two related purposes: to generate interest in a new and fascinating approach to understanding behavioral measures based on quantum probability principles, and 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 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

Keywords: quantum probability, quantum information processing, quantum logic, quantum event, quantum measurement, superposition state, mixed state.

on representations 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, Busemeyer, & Gabora, 2009; Bruza, Lawless, van Rijsbergen, & Sofge, 2007, 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, $i = 1, \ldots, n$ (*e.g.*, 7 degrees of guilt); and a random variable Y with possible values y_j, $j = 1, \ldots, m$ (*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 Fig. 3.1, 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 possible answers for two *yes/no* questions.

Classic probability theory

Events. Classic probability theory assigns probabilities to *classic events*. Each event (such as the event $x = X \geq 4$ or the event $y = Y < 3$ or the event $z = X + Y = 3$) is represented algebraically as a set belonging to a *field of sets*. That is, there is a *null event* represented by the empty set $\varnothing$, and a *universal event* $\mathcal{U}$ that contains all other events.[1] Further, new events can be formed from other events in three ways. One way is the *negation* operation, denoted $\sim x$, 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 axioms of Boolean algebra, as follows.

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

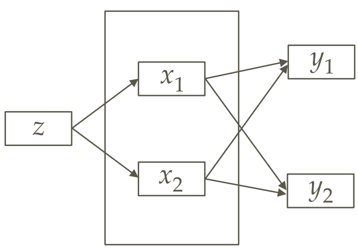

Figure 3.1: The possible measurement outcomes for Condition XY.

[1]For simplicity we restrict attention to experiments with only finitely many outcomes; then $\mathcal{U}$ can be assumed to be a finite set.

The distributive axiom B(5) is crucial for distinguishing classic probability theory from quantum probability theory.

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

K(1) Normalized: $0 \leq \Pr(x) \leq 1, \Pr(\varnothing) = 0, \Pr(\mathcal{U}) = 1$.

K(2) Additive: 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 $\Pr(y|x)$, defined by the ratio

$$\Pr(y|x) = \Pr(y \wedge x)/\Pr(x), \tag{1}$$

which implies the formula for *joint probabilities*

$$\Pr(y \wedge x) = \Pr(x)\Pr(y|x). \tag{2}$$

Classic probability distributions

Classically, our simple behavioral experiment is analyzed as follows. Consider first condition XY. We observe $n \times m$ 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 events $x_i = x_i y_1 \vee x_i y_2 \vee \ldots \vee x_i y_m$ and $y_j = x_1 y_j \vee x_2 y_j \vee \ldots \vee x_n y_j$. 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 axioms of Boolean algebra, and in particular, the distributive axiom B(5) states that

$$\begin{aligned} y_j = y_j \wedge \mathcal{U} = y_j \wedge (x_1 \vee x_2 \vee \ldots \vee x_n) \\ = (y_j \wedge x_1) \vee (y_j \wedge x_2) \vee \ldots \vee (y_j \wedge x_n). \end{aligned}$$

For binary valued measures ($n = m = 2$), all of the nonzero events are shown in Table 3.1.

The Boolean axioms B(1)–B(5) are used in conjunction with the Kolmogorov axioms K(1), K(2) to derive the *law of total probability*:

$$\begin{aligned} \Pr(y_j) = \Pr(y_j \wedge \mathcal{U}) = \Pr((y_j \wedge (x_1 \vee x_2 \vee \ldots \vee x_n)) \\ = \Pr((y_j \wedge x_1) \vee (y_j \wedge x_2) \vee \ldots \vee x_n)) \\ = \textstyle\sum_i \Pr(x_i \wedge y_j) = \sum_i \Pr(x_i)\Pr(y_j|x_i). \end{aligned} \tag{3}$$

Thus the marginal probability distribution for Y is determined from the joint probabilities, and this is also true for X. Finally, *Bayes's rule* follows from (1), (2), and (3):

$$\Pr(y_j|x_i) = \frac{\Pr(y_j \wedge x_i)}{\Pr(x_i)} = \frac{\Pr(y_j)\Pr(x_i|y_j)}{\sum_k \Pr(y_k)\Pr(x_i|y_k)}. \tag{4}$$

Recall that, in our experiment, under one condition we measure X then Y, but under another condition we measure only variable Y. According to classic probability, there is nothing to prevent us from postulating a joint probability like $\Pr(x_i \wedge y_j)$ 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 counterfactual events,

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

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 $\Pr(x_i \wedge y_j|Y)$ are assumed to exist even when we measure only 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 (3). This is actually a big assumption, although it is routinely taken for granted in the social and behavioral sciences.

This critical assumption can be understood more simply using Fig. 3.1. 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

Table 3.1: Events generated by Boolean Algebra operators.

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

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	$x_2 \wedge y_1$	$x_2 \wedge y_2$	$x_2 \wedge (y_1 \vee y_2)$
$x_1 \vee x_2$	$y_1 \wedge (x_1 \vee x_2)$	$y_2 \wedge (x_1 \vee x_2)$	$(x_1 \vee x_2) \wedge (y_1 \vee y_2) = \mathcal{U}$

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 (3) we should find $\Pr(y_j|XY) = \Pr(y_j|Y)$. Empirically, however, we often find that $\Pr(y_j|XY) \neq \Pr(y_j|Y)$; 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 them. 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, etc.) within a Hilbert space H (*i.e.*, a vector space with *complex numbers*[2] as scalars, and equipped with a *Hermitian inner product* used to measure length)[3]. The *null event* is represented by the zero subspace **0** (containing just the zero vector **0**) of the vector space H, and the *universal event* by 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: $L_{x \wedge y} = L_x \wedge L_y$. A third way is the *join* operation $x \vee y$ defined as the *span* of two subspaces L_x and L_y. Span is quite different than union, and this is where quantum logic differs from classic logic. It is due to this difference that, although quantum logic obeys axioms B(1)–B(4) of Boolean logic (and therefore all rules of Boolean logic that can be proved from just those axioms), it does not obey the distributive ax-

[2]Complex numbers cannot be avoided in quantum probability; see the section "Why complex numbers?" below, p. 64 *ff*.

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

iom B(5). That is, there are cases in quantum logic where the equation $L_z \wedge (L_x \vee L_y) = (L_z \wedge L_x) \vee (L_z \wedge L_y)$ fails to be true (of course it is true in some cases, *e.g.*, if $x = y = z$).

Fig. 3.2 illustrates an example of a violation of the distributive axiom. Suppose H is a 3-dimensional space.[4] This space can be defined in terms of an orthogonal basis formed by the three vectors labeled $|x\rangle$, $|y\rangle$, and $|z\rangle$ corresponding to the three standard coordinate axes (lines) L_x, L_y, L_z.[5] Alternatively, the same 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 other three (pairwise perpendicular) lines $L_u, L_v,$ and L_w in Fig. 3.2.[6] Consider the event $(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z)$. Since x, y, and z are a basis of H, the span of the three lines L_x, L_y, and L_z is all of H: thus the event $L_x \vee L_y \vee L_z$

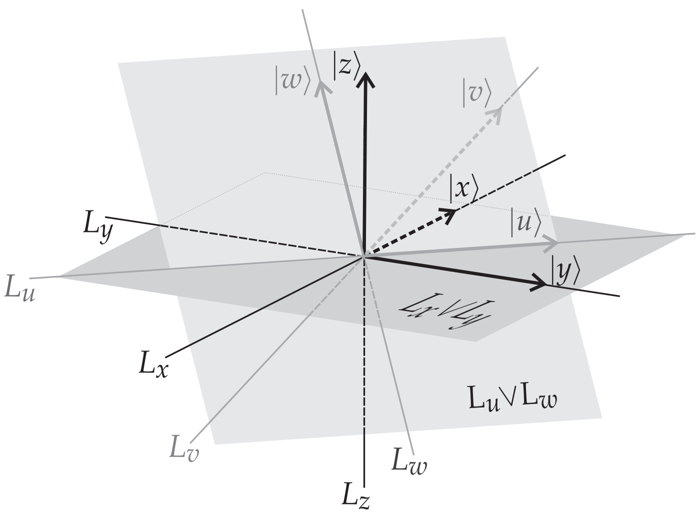

Figure 3.2: Violation of the distributive axiom.

[4]Although Fig. 3.2 is a depiction of ordinary real 3-space $\mathbb{R}^3$ in the usual way, it faithfully reflects the situation for complex 3-space $\mathbb{C}^3$.

[5]Dirac notation is used here. The *ket* $|v\rangle$ corresponds to a column vector, the *bra* $\langle v|$ corresponds to a row vector, the *bra-ket* $\langle x|y\rangle$ is an inner product, and $\langle x||P|y\rangle$ is a bra-matrix-ket product.

[6]Precisely, here $|u\rangle = |x\rangle/\sqrt{2} + |y\rangle/\sqrt{2}$, $|v\rangle = |x\rangle/2 + |y\rangle/2 + |z\rangle/\sqrt{2}$, and $|w\rangle = -|x\rangle/2 + |y\rangle/2 + |z\rangle/\sqrt{2}$.

is equal to the event H. The span of u and w is a plane within H; as an event that plane is $L_u \vee L_w$ (as indicated in Fig. 3.2). Similarly the span of x and y is a plane (the xy-plane), which as an event is $L_x \vee L_y$. From the definitions, we calculate

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

If the distributive axiom B(5) were applicable, we would then have

$$\begin{aligned}(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z) &= (L_u \vee L_w) \wedge ((L_x \vee L_y) \vee L_z) \\ &= (L_u \vee L_w) \wedge (L_x \vee L_y) \vee (L_u \vee L_w) \wedge L_z. \quad (5)\end{aligned}$$

Now, $(L_u \vee L_w) \wedge (L_x \vee L_y) = L_u$ because (as shown in Fig. 3.2) the intersection of the two planes is exactly the u-axis, *i.e.*, the event L_u; and $(L_u \vee L_w) \wedge L_z = \mathbf{0}$ because (again, as shown in Fig. 3.2) the intersection of the z-axis and the uw-plane is the single point $\mathbf{0}$. In sum, we find that

$$\begin{aligned}(L_u \vee L_w) \wedge (L_x \vee L_y \vee L_z) &= L_u \vee L_w \\ &\neq (L_u \vee L_w) \wedge (L_x \vee L_y) \vee (L_u \vee L_w) \wedge L_z = L_u \vee \mathbf{0} = L_u,\end{aligned}$$

contradicting (5). This example illustrates how quantum logic can violate the distributive axiom B(5) of Boolean logic.

Probabilities. Quantum probabilities are computed using projective rules that involve three steps. First, the probabilities for all events are determined from a *state vector* $|z\rangle \in H$ of unit length (*i.e.*, $\||z\rangle\| = 1$). 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 each state vector $|z\rangle \in H$ onto L_x.[7] Finally, probability of an event L_x is equal to the squared length of this projection:

$$\begin{aligned}\Pr(x) = \|\mathrm{P}_x|z\rangle\|^2 &= (\mathrm{P}_x|z\rangle)^\dagger(\mathrm{P}_x|z\rangle \\ &= \langle z|\mathrm{P}_x^\dagger\,\mathrm{P}_x|z\rangle = \langle z|\mathrm{P}_x\,\mathrm{P}_x|z\rangle = \langle z|\mathrm{P}_x|z\rangle.\end{aligned}$$

Fig. 3.3 illustrates the idea of projective probability. In this figure, the squared length of the projection of $|z\rangle$ onto L_{x_1} is the probability of the event L_{x_1} given the state $|z\rangle$.

[7] Projection operators are characterized as being *Hermitian* and *idempotent*. To say P is Hermitian means that $\mathrm{P} = \mathrm{P}^\dagger$; in matrix terms, for every i and j, the entry $p_{i,j}$ in row i, column j of P and the entry $p_{j,i}$ in row j, column i of P are complex conjugates of each other. To say P is idempotent means $\mathrm{P}^2 = \mathrm{P}$.

Quantum 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 ($i = 1, \ldots, n$). We will assume that no outcome x_i can be decomposed or refined into other distinguishable parts.

To each distinct outcome x_i we assign a corresponding line or ray L_{x_i} in our Hilbert space H. Corresponding to this subspace is a unit length vector, called a *basis state* and symbolized as $|x_i\rangle$, which generates this ray as the set of its scalar multiples $a\,|x_i\rangle$.[8] The basis states are assumed to be orthonormal: the inner product $\langle x_i|x_j\rangle$ is 0 for all pairs x_i, x_j of states with $i \neq j$, while for each state x_i the length $\|x_i\| = \sqrt{\langle x_i|x_i\rangle}$ equals 1. 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_{x_i} projects any point $|z\rangle$ in H into the subspace L_{x_i}. It is constructed from the *outer product* $|x_i\rangle\langle x_i|$; *i.e.*, for all z,

$$\mathrm{P}_{x_i}\,|z\rangle = (|x_i\rangle\langle x_i|)\,|z\rangle = |x_i\rangle\langle x_i|z\rangle = \langle x_i|z\rangle\,|x_i\rangle,$$

where $\langle x_i|z\rangle$ is the *inner product* (or "bra-ket"; *cf.* footnote 5). The inner product $\langle x_i|z\rangle$, in turn, can be interpreted as the probability amplitude[9] of transiting to state $|x_i\rangle$ from state $|z\rangle$. The *probability*

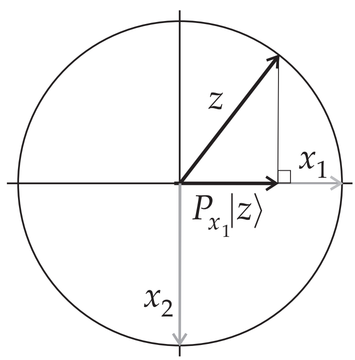

Figure 3.3: Projective probability: $\Pr(x_1) = \|\mathrm{P}_{x_1}|z\rangle\|^2$.

[8] Actually, both $|x_i\rangle$ and $-|x_i\rangle$ are basis states for x_i; the choice is immaterial.
[9] In general, this can be a complex number.

of any event L_{x_i} equals the squared projection,

$$\|\mathrm{P}_{x_i}|z\rangle\|^2 = \||x_i\rangle\langle x_i|z\rangle\|^2 = \||x_i\rangle\|^2 |\langle x_i|z\rangle|^2$$
$$= 1\cdot |\langle x_i|z\rangle|^2 = |\langle x_i|z\rangle|^2.$$

That is, the probability of transiting to state $|x_i\rangle$ from state $|z\rangle$ equals $|\langle x_i|z\rangle|^2$, the squared magnitude of the probability amplitude.

The probability of the meet $x \wedge y$ of two events x and 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

$$\Pr(x \wedge y) = \Pr(x_i) = |\langle x_i|z\rangle|^2.$$

We have $x_i \neq x_j$ for $i \neq j$, so the joint event $x_i \wedge x_j$ is zero, $L_{x_i} \wedge L_{x_j} = \mathbf{0}$, and the projection $\mathrm{P}_{\mathbf{0}}$ onto the zero subspace $\mathbf{0}$ is the zero operator $\mathbf{0}$; thus the joint probability $\Pr(x_i \wedge x_j)$ is $\|\mathbf{0}\|^2 = 0$.

The join of two events, say $x_i \vee x_j$, is the span $\{|x_i\rangle, |x_j\rangle\}$ of the two basis vectors. The projector for this subspace is

$$\mathrm{P}_{x_i \vee x_j} = \mathrm{P}_{x_i} + \mathrm{P}_{x_j} = |x_i\rangle\langle x_i| + |x_j\rangle\langle x_j|.$$

The probability of the event $x_i \vee x_j$ is thus simply the sum of the separate probabilities,

$$\|\mathrm{P}_{x_i \vee x_j}|z\rangle\|^2 = \|(|x_i\rangle\langle x_i| + |x_j\rangle\langle x_j|\,|z\rangle)\|^2$$
$$= \||x_i\rangle\,\langle x_i|z\rangle + |x_j\rangle\,\langle x_j|z\rangle\|^2$$
$$= |\langle x_i|z\rangle|^2 + |\langle x_i|z\rangle|^2,$$

where the final step follows from the orthogonality property.

Finally, for any $|z\rangle$ we have $\mathrm{P}_H\,|z\rangle = |z\rangle$ and so

$$\|\mathrm{P}_H\,|z\rangle\|^2 = \||z\rangle\|^2 = |\langle z|z\rangle|^2 = 1.$$

This also implies that

$$\mathrm{P}_H = \textstyle\sum_i \mathrm{P}_{x_i} = \sum_i |x_i\rangle\langle x_i| = \mathbf{I},$$

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

Q(1) $0 \leq \|\Pr(x)|z\rangle\|^2 \leq 1, \Pr(\mathbf{0}) = 0, \Pr(H) = 1.$
Q(2) if $L_x \wedge L_y = 0$ then $\Pr(L_x \vee L_y) = \Pr(L_x) + \Pr(L_y).$

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:

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

Thus the initial state vector is a *superposition* (*i.e.*, linear combination) of the basis states. The inner product $\langle x_i|z\rangle$ is the coefficient (or *component*) of the state vector that corresponds to the $|x_i\rangle$ basis state. To be concrete, one can define $|x_i\rangle$ as a column vector with a 0 in every row except for row i, where there is a 1. 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 short, 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, 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, after which the person is in a state consistent with the event so obtained. 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 measure only one.

The problem with the classical interpretation of measurement can be seen most clearly by reconsidering the example shown in Fig. 3.1 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 measure only 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 B(5), this event means that prior to our measurement either the person is in the low guilty state and intends to punish at the low level $x_1 \wedge y_1$ (*i.e.*, the brain experienced the upper path in Fig. 3.1), or the person is in the high guilty state and intends to punish at the low level $x_2 \wedge y_1$ (*i.e.*, the brain experienced the lower path in Fig. 3.1). 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, 1995, 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 B(5) fails. Referring again to Fig. 3.1, if we measure punishment state y_1, then event $y_1 \wedge (x_1 \vee x_2)$ has occurred, but from this we cannot infer 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, 1966, p. 9). On the contrary, 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 $\Pr(x|x) = |\langle \mathrm{P}_x | x \rangle|^2 = \| |x\rangle \|^2 = 1$. 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.

Quantum probability distributions for several variables

After we have first measured X and observed the event x, the state changes to $|x\rangle = \mathrm{P}_x|z\rangle / \|\mathrm{P}_x|z\rangle\|$, 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 $\Pr(y|x) = \|\mathrm{P}_y|x\rangle\|^2$. This updating process continues for each new measurement.

When more than one measurement is involved, quantum probability is more general than Kolmogorov probability, and quantum logic does not have to obey the distributive axiom B(5). In quantum theory, the analysis of an experimental situation in which more than one measurement is made, depends on how one represents the relationship between the measurements. There are two possibilities: the measures may be *compatible* or *incompatible*.

Compatible measurements. We consider specifically the problem of two measurements, first in 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 combinations of distinct outcomes for X and Y of the form $x_i y_j$. The complete Hilbert space is defined by $n \times m$ orthonormal basis vectors $|x_i y_j\rangle$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$, spanning a space H of dimension $n \times m$. 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 that $\langle x_i y_j | x_k y_\ell \rangle = 0$, and the normal property that $\langle x_i y_j | x_i y_j \rangle = 1$. This Hilbert space H is called the *tensor product space* for the two measures.

Notice that the event x_i is no longer a distinct outcome. Instead, it is a coarse-grained outcome that can be expressed, using the tensor decomposition, in terms of more refined parts:

$$L_{x_i} = |x_i y_1\rangle \vee |x_i y_2\rangle \vee \ldots \vee |x_i y_m\rangle.$$

Furthermore, the meet $x_i \wedge y_j$ produces the subspace $L_{x_i} \wedge L_{y_j} = |x_i y_j\rangle$. This implies that, for this tensor decomposition, the distributive axiom B(5) does hold:

$$\begin{aligned} L_{x_i} &= |x_i y_1\rangle \vee |x_i y_2\rangle \vee \ldots \vee |x_i y_m\rangle \\ &= (L_{x_i} \wedge L_{y_1}) \vee (L_{x_i} \wedge L_{y_2}) \vee \ldots \vee (L_{x_i} \wedge L_{y_m}). \end{aligned}$$

Thus Table 3.1 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 operators for the events L_{x_i,y_j}, L_{x_i}, and L_{y_j} are

$$\begin{aligned} \mathrm{P}_{x_i,y_j} &= |x_i y_j\rangle\langle x_i y_j|, \\ \mathrm{P}_{x_i} &= \textstyle\sum_j \mathrm{P}_{x_i,y_j} = \sum_j |x_i y_j\rangle\langle x_i y_j|, \\ \mathrm{P}_{y_j} &= \textstyle\sum_i \mathrm{P}_{x_i,y_j} = \sum_i |x_i y_j\rangle\langle x_i y_j|. \end{aligned}$$

The orthogonality properties then imply

$$\begin{aligned} |x_i y_j\rangle\langle x_i y_j| &= \left(\textstyle\sum_j |x_i y_j\rangle\langle x_i y_j|\right)\left(\textstyle\sum_i |x_i y_j\rangle\langle x_i y_j|\right) && (6) \\ &= \mathrm{P}_{x_i}\,\mathrm{P}_{y_j} && (7) \\ &= \mathrm{P}_{y_j}\,\mathrm{P}_{x_i}. && (8) \end{aligned}$$

(6) implies that the projection for the joint event L_{x_i,y_j} can be viewed as a series of two successive measurements, and vice versa. (7) and (8) show that the projectors for X *commute with* the projectors for Y: that is, the order of projection does not matter—both orders project onto the same final subspace. In general, given operators (in particular, projectors) A and B, the combination $AB - BA$ is called the

commutator of A and B. We have shown that the commutator of compatible measures is always zero.

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 $|x_i\rangle = \mathrm{P}_{x_i}|z\rangle / \|\mathrm{P}_{x_i}|z\rangle\|$; similarly if Y is measured first and we observe y_j, then the new state after measurement is $|y_j\rangle = \mathrm{P}_{y_j}|z\rangle / \|\mathrm{P}_{y_j}|z\rangle\|$. Consider again the probability of the event L_{x_i,y_j}, viewed as a series of projections.

$$
\begin{aligned}
\mathrm{P}_{x_i,y_j}|z\rangle &= \mathrm{P}_{x_i}\left(\mathrm{P}_{y_j}|z\rangle\right) = \mathrm{P}_{x_i}\,|y_j\rangle\,\|\mathrm{P}_{y_j}|z\rangle\| \\
\mathrm{P}_{y_j,x_i}|z\rangle &= \mathrm{P}_{y_j}\left(\mathrm{P}_{x_i}|z\rangle\right) = \mathrm{P}_{y_j}\,|x_i\rangle\,\|\mathrm{P}_{x_i}|z\rangle\| \\
\Pr(x_i \wedge y_j) &= \|\mathrm{P}_{x_i,y_j}|z\rangle\|^2 \\
&= \|\mathrm{P}_{x_i}\,|y_j\rangle\|^2\,\|\mathrm{P}_{y_j}|z\rangle\|^2 = \Pr(x_i|y_j)\Pr(y_j) \qquad (9) \\
\Pr(y_j \wedge x_i) &= \|\mathrm{P}_{y_j,x_i}|z\rangle\|^2 \\
&= \|\mathrm{P}_{y_j}\,|x_i\rangle\|^2\,\|\mathrm{P}_{x_i}|z\rangle\|^2 = \Pr(y_j|x_i)\Pr(x_i). \qquad (10)
\end{aligned}
$$

From (9) and (10) we obtain the conditional probability axioms for quantum probabilities:

$$
\begin{aligned}
\Pr(x_i|y_j) &= \|\mathrm{P}_{x_i}\,|y_j\rangle\|^2 = \|\mathrm{P}_{x_i,y_j}|z\rangle\|^2 / \|\mathrm{P}_{y_j}\,|z\rangle\|^2, \qquad (11) \\
\Pr(y_j|x_i) &= \|\mathrm{P}_{y_j}\,|x_i\rangle\|^2 = \|\mathrm{P}_{x_i,y_j}|z\rangle\|^2 / \|\mathrm{P}_{x_j}\,|z\rangle\|^2. \qquad (12)
\end{aligned}
$$

In general $\|\mathrm{P}_{x_i}|z\rangle\|^2 \neq \|\mathrm{P}_{y_j}|z\rangle\|^2$ and so also $\Pr(y_j|x_i) \neq \Pr(x_i|y_j)$.

The projection onto L_{x_i} is $\mathrm{P}_{x_i}\,|z\rangle = \sum_j |x_i y_j\rangle\langle x_i y_j|z\rangle$ and the probability of this event equals

$$\Pr(x_i) = \sum_j |\langle x_i y_j|z\rangle|^2 = \sum_j \|\mathrm{P}_{x_i}\,|y_j\rangle\|^2\,\|\mathrm{P}_{y_j}\,|z\rangle\|^2. \qquad (13)$$

(13) is the law of total probability for quantum probabilities. From (12) and (13), we can derive a quantum analogue of Bayes's rule (4):

$$\Pr(y_j|x_i) = \|\mathrm{P}_{y_j}\,|x_i\rangle\|^2 = \frac{\|\mathrm{P}_{x_i}|y_j\rangle\|^2\,\|\mathrm{P}_{y_j}|z\rangle\|^2}{\sum_k \|\mathrm{P}_{x_i}|y_k\rangle\|^2\,\|\mathrm{P}_{y_k}|z\rangle\|^2}.$$

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:

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

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 axioms. 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.[10]

Incompatible measurements. 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. Similarly, 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 that 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 (non-identity) linear transformation of the other. Thus we now have two *different* bases for the *same* n-dimensional Hilbert space. This idea is illustrated in Fig. 3.4. 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) by the basis vectors $|y_1\rangle$ and $|y_2\rangle$. Note that the basis vectors for the Y measure are a linear transformation—specifically, an orthogonal rotation—of the basis vectors for the X measure (and vice versa). One can use either the $|x_1\rangle$, $|x_2\rangle$ basis or the $|y_1\rangle$, $|y_2\rangle$ basis to describe the state vector $|z\rangle$, but one cannot

[10] This is not quite true. We are only focusing on change caused by measurement, and disregarding change caused by dynamic laws.

use both bases at the same time.

One cannot experience or measure both variables X and Y simultaneously. If one measures X, then one needs to project the state $|z\rangle$ onto the X basis, not 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: $\Pr(y_j) = |\langle y_j|x_1\rangle|^2$ ($j = 1, 2$). Similarly, if one measures Y, then the Y basis must be used, and if Y is measured first and the value y_1 is observed, then the outcome for the next measurement on X must be uncertain: $\Pr(x_i) = |\langle x_i|y_1\rangle|^2$ ($i = 1, 2$). It is impossible to be certain about both values simultaneously! Therefore, it is impossible to completely and correctly 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 B(5) of Boolean logic is violated by incompatible measures. For example, considering Fig. 3.4, note that $|x_i\rangle \wedge |y_j\rangle = 0$ for all i and j, and therefore we have

$$L_{y_1} = L_{y_1} \wedge (L_{x_1} \vee L_{x_2}) \neq (L_{y_1} \wedge L_{x_1}) \vee (L_{y_1} \wedge L_{x_2}) = 0 \vee 0 = 0,$$

a violation of distributivity. In this example, because of incompatibility the event each of the events $L_{y_1} \wedge L_{x_1}$ and $L_{y_2} \wedge L_{x_2}$ is impossible, yet clearly the event L_{y_1} is possible. This is where quantum probability deviates from classic probability. Table 3.2 shows the events for incompatible measures, including clear violations of the distributive axiom B-(5).

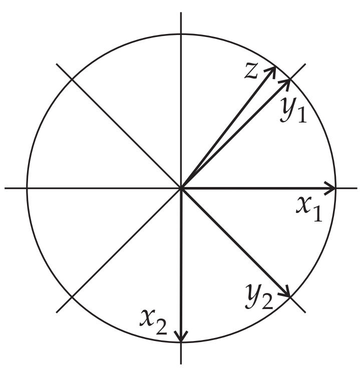

Figure 3.4: Illustration of rotated basis vectors for incompatible measurements.

To get a deeper understanding of the violation of the distributive axiom B-(5), let us return to Fig. 3.1 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 $|z\rangle \rightarrow |x_1\rangle \rightarrow |y_1\rangle$ or $|z\rangle \rightarrow |x_2\rangle \rightarrow |y_1\rangle$—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 having first concluded that the person was guilty at a low degree or that the person was guilty at a high degree. Because we do not measure guilt, we cannot assume that the person is definitely in one of these two guilt states; on the contrary, in the quantum probability model, the person is indefinite (or superposed) between these two states. When we do not observe what happens, quantum theory allows for a type of uncertainty regarding state changes that is more general than classical probability theory.

The fact that there are two different bases for the same Hilbert space implies that the same state vector has two different descriptions in terms of those bases:

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

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 $|z\rangle$. In this basis, we can represent the initial state vector by a column vector Ψ with $\langle x_i|z\rangle$ in row i, and 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 $|z\rangle$; in this basis, we can represent the initial state vector by a column vector Φ with $\langle y_i|z\rangle$ in row i, and the marginal probability distribution for Y is $\Pr(y_j) = \|\Phi_i\|^2 = |\langle y_j|z\rangle|^2$. No joint distribution

Table 3.2: Events generated by incompatible measures.

Note: $L_{y_1} = L_{y_1} \wedge (L_{x_1} \vee L_{x_2}) \neq (L_{y_1} \wedge L_{x_1}) \vee (L_{y_1} \wedge L_{x_2}) = \mathbf{0}$, etc.

Events	y_1	y_2	$(y_1 \vee y_2)$
x_1	$\mathbf{0}$	$\mathbf{0}$	$x_1 \wedge (y_1 \vee y_2)$
x_2	$\mathbf{0}$	$\mathbf{0}$	$x_2 \wedge (y_1 \vee y_2)$
$(x_1 \vee x_2)$	$y_1 \wedge (x_1 \vee x_2)$	$y_2 \wedge (x_1 \vee x_2)$	$(x_1 \vee x_2) \wedge (y_1 \vee y_2)$

exists, but both marginal distributions are derived from a common state vector $|z\rangle$. The equality of the two representations implies

$$\begin{aligned}
\textstyle\sum_i |x_i\rangle\langle x_i|z\rangle &= \textstyle\sum_i |y_i\rangle\langle y_i|z\rangle, \\
\implies \langle x_j| \textstyle\sum_i |x_i\rangle\langle x_i|z\rangle &= \langle x_j| \textstyle\sum_i |y_i\rangle\langle y_i|z\rangle, \\
\implies \textstyle\sum_i \langle x_j|x_i\rangle\langle x_i|z\rangle &= \textstyle\sum_i \langle x_j|y_i\rangle\langle y_i|z\rangle, \\
\implies \langle x_j|z\rangle &= \textstyle\sum_i \langle x_j|y_i\rangle\langle y_i|z\rangle,
\end{aligned}$$

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[11]; its square $|\langle x_j|y_i\rangle|^2$ equals the probability of observing x_j on the next measurement of X given that y_i was obtained from a previous measure of Y. A similar argument produces

$$\textstyle\sum_i \langle y_j|x_i\rangle\langle x_i|z\rangle = \langle y_j|z\rangle,$$

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; its square $|\langle y_j|x_i\rangle|^2$ equals the probability of observing y_j on the next measurement of Y given that x_i was obtained from a previous measure of X.

In sum, one constructs (a) the first marginal distribution from the inner products like $\langle x_i|z\rangle$ that relate the initial state to the states for the first basis, and (b) the second marginal distribution from the inner products like $\langle y_j|x_i\rangle$ that relate 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 (even) incompatible measurements must satisfy

$$\Pr(x_j|y_i) = \Pr(y_i|x_j), \tag{14}$$

[11] The notation ζ^* stands for the complex conjugate of the complex number ζ (recall that, in general, probability amplitudes like $\langle y_i|x_j\rangle$ can be complex numbers). That $\langle y_i|x_j\rangle^*$ always equals $\langle x_j|y_i\rangle$ is ensured by the assumption (p. 46) that the bra-ket inner product on the Hilbert space H is Hermitian.

the so-called *law of reciprocity* (Peres, 1995, p. 34). Of course, classic probability is not subject to this constraint. It is important to note that (14) need hold only for transitions between basis states, not for for more general (coarse-grained) events.

Second, consider the matrix U of coefficients with element $\langle y_i | x_j \rangle$, representing the transition to state $|y_j\rangle$ from state $|x_i\rangle$, in row i and column j. Then the column vector Ψ (which describes the initial state in terms of the X basis) is related to the column vector Φ (which describes the initial state in terms of the Y) by the linear transformation $\Phi = \mathrm{U}\,\Psi$; and similarly, Φ is related to Ψ by the linear transformation $\Psi = \mathrm{U}^\dagger\,\Phi$. Notice that therefore

$$\Phi = \mathrm{U}\,\mathrm{U}^\dagger\,\Phi, \quad \Psi = \mathrm{U}^\dagger \mathrm{U}\,\Psi$$

and this is true (with the same U) no matter what the initial state (represented in terms of the two bases by Φ and Ψ) happens to be. It follows that

$$\mathrm{U}\,\mathrm{U}^\dagger = \mathbf{I} = \mathrm{U}^\dagger \mathrm{U};$$

in other words, U is a *unitary* matrix. Unitarity of U guarantees that U preserves the lengths of the vectors before and after transformation, and implies that the *transition matrix* T, which has $|\langle y_i | x_j \rangle|^2$ in row i and column j, must be *doubly stochastic*: each row and each column of T must sum to unity. This is called the *doubly stochastic law* (Peres, 1995, p. 33). In classic probability theory, the transition matrix must be stochastic (each column sums to unity) but need not be doubly stochastic.

Thus, for incompatible measures, quantum probabilities must obey two laws that are not required by classic probability: the law of reciprocity (14) and the doubly stochastic law. On the other hand, classic probability must obey the law of total probability (3), which is not required by quantum probability for incompatible measures. These three properties can be used to distinguish quantum models from classical models empirically.

The projector for the event L_{x_i} is $\mathrm{P}_{x_i} = |x_i\rangle\langle x_i|$, and that for the event L_{y_j} is $\mathrm{P}_{y_j} = |y_j\rangle\langle y_j|$. It is interesting to compare the composition of projections produced by measuring Y first followed by X,

$$\mathrm{P}_{x_i}\,\mathrm{P}_{y_j} = |x_i\rangle\langle x_i|\,|y_j\rangle\langle y_j| = \langle x_i | y_j \rangle |x_i\rangle\langle y_j|,$$

with that produced by measuring X first followed by Y,

$$\mathrm{P}_{y_j}\,\mathrm{P}_{x_i} = |y_j\rangle\langle y_j|\,|x_i\rangle\langle x_i| = \langle y_j | x_i \rangle |y_j\rangle\langle x_i|.$$

In contrast to the case of compatible measures, where the commutator is always zero (*cf.* p. 55), here the assumption of incompatibility ensures that the commutator

$$\mathrm{P}_{x_i}\,\mathrm{P}_{y_j} - \mathrm{P}_{y_j}\,\mathrm{P}_{x_i} = \langle x_i|y_j\rangle |x_i\rangle\langle y_j| - \langle y_j|x_i\rangle |y_j\rangle\langle x_i|$$

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 pervasive phenomenon throughout the social and behavioral sciences.

Let us now examine the event probabilities in the case of incompatible measures. Here we have to give separate careful analyses of the different possible experimental conditions.

First consider condition XY. In this case we have

$$\begin{aligned}\Pr(y_j \wedge x_i|XY) &= \|\mathrm{P}_{y_j}\mathrm{P}_{x_i}\,|z\rangle\|^2 = |\langle y_j|x_i\rangle|^2\,|\langle x_i|z\rangle|^2 \\ &= \Pr(y_j|x_i, XY)\Pr(x_i|XY),\end{aligned}$$

so that

$$\Pr(y_j|XY) = \textstyle\sum_i |\langle x_i|z\rangle|^2\,|\langle y_j|x_i\rangle|^2, \tag{15}$$

similar to the situations for compatible measurements in both classic and quantum probability.

To get a more intuitive idea, refer again to Fig. 3.1. The probability $\Pr(x_1|XY)$ of responding x_1 to question X on the first measure is equal to $|\langle x_1|z\rangle|^2$, the squared probability amplitude of transiting from the initial state $|z\rangle$ to the basis vector $|x_1\rangle$. Given that the first measurement produces x_1, and the state now equals $|x_1\rangle$, the probability $\Pr(y_1|x_1, XY)$ of responding $Y = y_1$ to the second question is equal to $|\langle y_1|x_1\rangle|^2$, the squared probability amplitude of transiting from $|x_1\rangle$ to $|y_1\rangle$. The probability of observing $X = x_1$ on the first test followed by $Y = y_1$ on the second test equals

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

A similar analysis produces

$$\Pr(x_2|XY)\Pr(y_1|x_2, XY) = |\langle x_2|z\rangle|^2\,|\langle y_1|x_2\rangle|^2$$

for the probability of observing $X = x_2$ on the first test followed by $Y = y_1$ on the second test. Thus the probability of observing $Y = y_1$ on the second test, given the XY condition, equals

$$\Pr(y_1|XY) = |\langle x_1|z\rangle|^2\,|\langle y_1|x_1\rangle|^2 + |\langle x_2|z\rangle|^2\,|\langle y_1|x_2\rangle|^2.$$

Next consider the probability of responding to question Y alone, not preceded by question X. The projection of the initial state onto the L_{y_j} event is $\mathrm{P}_{y_j}|z\rangle = |y_j\rangle\langle y_j|\,|z\rangle = |y_j\rangle\langle y_j|z\rangle$, and so $\Pr(y_j|Y) = \langle z|y_j\rangle\langle y_j|y_j\rangle\langle y_j|z\rangle = |\langle y_j|z\rangle|^2$. 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 result:

$$\Pr(y_j|Y) = |\langle y_j|z\rangle|^2 = |\langle y_j|\mathbf{I}|z\rangle|^2 = \\ |\textstyle\sum_i \langle y_j|(|x_i\rangle\langle x_i|)|z\rangle|^2 = |\sum_i \langle y_j|x_i\rangle\langle x_i|z\rangle|^2. \quad (16)$$

Comparing (16) with (15), we see that $\Pr(y_j|Y)$ and $\Pr(y_j|XY)$ need not be equal. This difference can explain interference effects. Let us analyze the interference effect in more detail for the special case shown in Fig. 3.1, with only two outcomes for each measure.

$$\begin{aligned}
|\langle y_1|z\rangle|^2 &= (\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle) \\
&\qquad (\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle)^* \\
&= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + \\
&\quad \langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle^*\langle x_2|z\rangle^* + \\
&\quad \langle y_1|x_2\rangle\langle x_2|z\rangle\langle y_1|x_1\rangle^*\langle x_1|z\rangle^* \\
&= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + \\
&\quad |\langle y_1|x_1\rangle|\,|\langle x_1|z\rangle|\,|\langle y_1|x_2\rangle|\,|\langle x_2|z\rangle| \\
&\quad (e^{i(\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle\langle x_2|z\rangle)} + e^{-i(\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle\langle x_2|z\rangle)}) \\
&= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + \\
&\quad |\langle y_1|x_1\rangle|\,|\langle x_1|z\rangle|\,|\langle y_1|x_2\rangle|\,|\langle x_2|z\rangle| \\
&\quad (\cos(\theta) + i\,\sin(\theta) + \cos(\theta) - i\,\sin(\theta)) \\
&= |\langle y_1|x_1\rangle\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle\langle x_2|z\rangle|^2 + \\
&\quad 2|\langle y_1|x_1\rangle|\,|\langle x_1|z\rangle|\,|\langle y_1|x_2\rangle|\,|\langle x_2|z\rangle|\,\cos(\theta),
\end{aligned}$$

where θ is the angle in the complex plane of the complex number $\langle y_1|x_1\rangle\langle x_1|z\rangle\langle y_1|x_2\rangle^*\langle x_2|z\rangle^*$ (see Fig. 3.5). If we restrict the probability amplitudes to real numbers, then we are restricted to the horizontal line in Fig. 3.5, so that $\theta = 0$ or $\theta = \pi$ and $\cos(\theta) = \pm 1$.

Note that the first two terms in the above expression for $\Pr(y_1|Y)$ exactly match those found when computing $\Pr(y_1|XY)$. If the cosine in the third term is zero, then $\Pr(y_1|Y) - \Pr(y_1|XY) = 0$ and there would be no interference. Thus the difference $\Pr(y_1|Y) - \Pr(y_1|XY)$ is contributed 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, $\Pr(y_1|Y)$ and $\Pr(y_1|XY)$, whereas the latter postulates two separate joint probability distributions and then derives the marginals for each condition from these separate joint distributions.

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

$$\Pr(x_1 \wedge y_1|XY) = |\mathrm{P}_{y_1}\mathrm{P}_{x_1}|z\rangle|^2 = |\langle x_1|z\rangle|^2\,|\langle y_1|x_1\rangle|^2,$$
$$\Pr(y_1 \wedge x_1|YX) = \|\mathrm{P}_{x_1}\mathrm{P}_{y_1}|z\rangle\|^2 = |\langle y_1|z\rangle|^2\,|\langle x_1|y_1\rangle|^2.$$

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

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

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 appears in Fig. 3.6. 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.

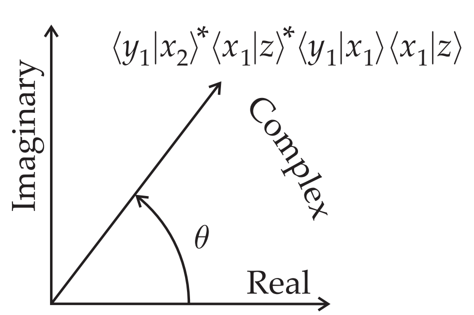

Figure 3.5: The angle between probability amplitudes.

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

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

This law of reciprocity places a very strong constraint on 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 coarse measurements that are disjunctions of several basis vectors.

Why Complex Numbers?

Consider again Fig. 3.1 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:

$$\Pr(y_1|Y) = |\langle y_1|x_1\rangle|^2|\langle x_1|z\rangle|^2 + |\langle y_1|x_2\rangle|^2|\langle x_2|z\rangle|^2 \\ \pm 2|\langle y_1|x_1\rangle|\,|\langle x_1|z\rangle|\,|\langle y_1|x_2\rangle|\,|\langle x_2|z\rangle|.$$

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, and that we find that

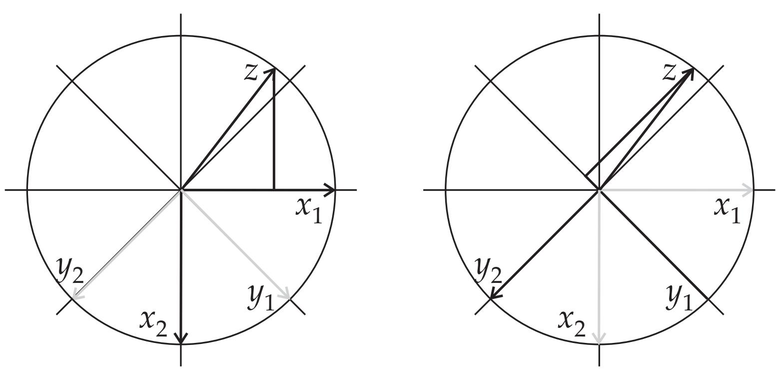

Figure 3.6: Projections of the initial state on basis vectors from two different orthonormal bases.

changing the experimental factor from level F_1 to level F_2 produces an interference effect of the same sign ($+$ or $-$), but increases the magnitude of the interference:

$$|\Pr(y_1|Y, F_2) - \Pr(y_1|XY, F_2)| > |\Pr(y_1|Y, F_1) - \Pr(y_1|XY, F_1)|.$$

Suppose also that this same manipulation does not change the joint probabilities, so that

$$\begin{aligned}\Pr(x_1 \wedge y_1|XY, F_1) &= |\langle y_1|x_1\rangle|^2 |\langle x_1|z\rangle|^2 \\ &= \Pr(x_1 \wedge y_1|XY, F_2), \qquad (17)\\ \Pr(x_2 \wedge y_1|XY, F_1) &= |\langle y_1|x_2\rangle|^2 |\langle x_2|z\rangle|^2 \\ &= \Pr(x_2 \wedge y_1|XY, F_2). \qquad (18)\end{aligned}$$

Together, (17) and (18) 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, pp. 295–296; *cf.* also Fig. 3.1). 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 and x_2 = right in Fig. 3.1). A final detector determines whether the output from the quarter wave plate can be detected (symbolized y_1 in Fig. 3.1) by a linear polarized detector rotated at angle φ with respect to the original state of the photon. In this situation, the critical factor F that is manipulated is the angle φ between the initial and final linear polarization.

The two channel outputs from the quarter wave plate form an orthonormal basis of two vectors, $|x_1\rangle$ and $|x_2\rangle$, in terms of which the state is represented. The probability amplitude of transiting from the initial state to the final state equals

$$\begin{aligned}\langle y_1|z\rangle = \langle y_1|\mathbf{I}|z\rangle &= \langle y_1|\left(|x_1\rangle\langle x_1| + |x_2\rangle\langle x_2|\right)|z\rangle \\ &= \langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle.\end{aligned}$$

When the right channel is closed, the then probability of passing through the left channel is $|\langle x_1|z\rangle|^2 = 1/2$, and the probability of detection is also $|\langle y_1|x_1\rangle|^2 = 1/2$. The same is true when the left channel is closed: then the probability of passing through right channel is $|\langle x_2|z\rangle|^2 = 1/2$ and the probability of detection is $|\langle y_1|x_2\rangle|^2 = 1/2$.

Further, when both channels are open, the probability of detection is $\cos(\varphi)$. Therefore, we have five equations in the four unknowns $\langle x_1|z\rangle$, $\langle y_1|x_1\rangle$, $\langle x_2|z\rangle$, $\langle y_1|x_2\rangle$:

$$|\langle x_1|z\rangle| = |\langle y_1|x_1\rangle| = |\langle x_2|z\rangle| = |\langle y_1|x_2\rangle| = 1/\sqrt{2},$$
$$\langle y_1|x_1\rangle\langle x_1|z\rangle + \langle y_1|x_2\rangle\langle x_2|z\rangle = \cos(\varphi).$$

The first four equations do not depend on φ, but the last one does. This forces us to find a solution using complex numbers. In this case, the solutions are

$$\langle x_1|z\rangle = 1/\sqrt{2} = \langle x_2|z\rangle, \langle y_1|x_1\rangle = e^{-i\varphi}/\sqrt{2}, \langle y_2|x_2\rangle = e^{i\varphi}/\sqrt{2}.$$

What is the difference between superposition and mixture?

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_{x_1} is an event corresponding to the observation of x_1 and L_{x_2} is another event corresponding to the observation y, then we can form a new disjunction event $L_{x_1} \vee L_{x_2}$ which is the set of all linear combinations $|z\rangle = a\,|x_1\rangle + b\,|x_2\rangle$, where the coefficients a and b are complex numbers with $|a|^2 + |b|^2 = 1$. 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 the value x_2 with probability $|\langle x_2|z\rangle|^2 = |b|^2$.

It is difficult to interpret the superposition state. There is no well agreed upon psychological interpretation of superposition—indeed, the interpretation of this concept has produced great controversy (Schroedinger's cat problem). Intuitively, a superposition seems to be something like a *fuzzy* and *uncertain* representation of a state. It is tempting, but invalid, to interpret superposition is meaning that immediately before measurement, you are either in state $|x_1\rangle$ with probability $|a|^2$ or in state $|x_2\rangle$ with probability $|b|^2$. In fact, that is a description of a *mixed state* (either classical or quantum), not a quantum *superposition state* (quantum only). These two types of quantum states are distinguishable by their probability predictions, as the following example shows.

Again, we start from Fig. 3.1 with binary outcomes, this time letting the relationship between the two bases be given by

$$|y_1\rangle = (|x_1\rangle + |x_2\rangle)/\sqrt{2}, |y_2\rangle = (|x_1\rangle - |x_2\rangle)/\sqrt{2},$$
$$|x_1\rangle = (|y_1\rangle + |y_2\rangle)/\sqrt{2}, |x_2\rangle = (|y_1\rangle - |y_2\rangle)/\sqrt{2}$$

(as in Fig. 3.6). After a measurement of $Y = y_1$, we are in the superposition state $|y_1\rangle = (|x_1\rangle + |x_2\rangle)/\sqrt{2}$, and we have the probabilities

$$\Pr(x_1) = \Pr(x_2) = 1/2, \Pr(y_1) = 1, \Pr(y_2) = 0. \qquad (19)$$

On the other hand, consider the mixed state in which the basis states $|x_1\rangle$ and $|x_2\rangle$ each have probability $1/2$. Whichever of the two basis states you are, y_1 and y_2 are equally likely measurements for Y. so the mixed state produces the probabilities

$$\Pr(x_1) = \Pr(x_2) = \Pr(y_1) = \Pr(y_2) = 1/2,$$

differing dramatically from (19). In sum, an equal mixture of $|x_1\rangle$ and $|x_2\rangle$ produces different results from an equally weighted superposition of $|x_1\rangle$ and $|x_2\rangle$: but this difference is only revealed by obtaining probabilities from both X and an incompatible measure 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 et al., 2009, 2007, 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., Busemeyer, J. R., & Gabora, L. (2009). Introduction to the special issue on quantum cognition. *Journal of Mathematical Psychology*, *53*(5), 303–305. (Special Issue: Quantum Cognition.)

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

Bruza, P. D., Lawless, W., van Rijsbergen, C., & Sofge, D. (Eds.).

(2008). *Proceedings of the Second Conference on Quantum Interactions, March 26–28, Oxford University.* AAAI Press.

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

French, A., & Taylor, E. (1978). *An introduction to quantum physics*. New York: W. W. Norton.

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

Khrennikov, A. (2007). Can quantum information be processed by macroscopic systems? *Quantum Information Theory*, *6*(6), 401–429.

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