American Psychological Association

Psychological Review

0033-295X/18/$12.00

2018, Vol. 125, No. 4, 572–591
http://dx.doi.org/10.1037/rev0000106

Hilbert Space Multidimensional Theory

Jerome R. Busemeyer
Indiana University

Zheng Wang
The Ohio State University

A general theory of measurement context effects, called Hilbert space multidimensional (HSM) theory, is presented. A measurement context refers to a subset of psychological variables that an individual evaluates on a particular occasion. Different contexts are formed by evaluating different but possibly overlapping subsets of variables. Context effects occur when the judgments across contexts cannot be derived from a single joint probability distribution over the complete set of values of the observed variables. HSM theory provides a way to model these context effects by using quantum probability theory, which represents all the variables within a low dimensional vector space. HSM models produce parameter estimates that provide a simple and informative interpretation of the complex collection of judgments across contexts. Comparisons of HSM model fits with Bayesian network model fits are reported for a new large experiment, demonstrating the viability of this new model. We conclude that the theory is broadly applicable to measurement context effects found in the social and behavioral sciences.

Keywords: quantum cognition, Bayesian networks, social cognition, context effects, contingency table analysis

This article presents a general theory for predicting and understanding the effects that measurement contexts have on human judgments. A measurement context refers to a set of psychological variables or attributes that an individual is required to evaluate on a particular occasion (Dzhafarov & Kujala, 2016). Different contexts are formed by evaluating different but possibly overlapping subsets of variables. A measurement context effect occurs when the judgments about the variables or attributes are affected by the measurement context in which they appear (Bruza, 2016). This can happen when the interpretation or meaning of some attributes change across measurement contexts (see, e.g., Schwarz & Sudman, 2012). Context effects raise problems for multivariate analysis because traditional approaches rely on the use of a single joint probability space based on the observed variables, which turns out to be invalid when context effects are present. We propose an approach to analyzing different measurement contexts based on quantum probability theory, which was originally developed to account for variables that have contextual dependencies.

The article is organized as follows. First, we provide a more precise definition of *context effects* after considering an illustrative example. Second, we empirically review context effects reported previously in the literature. Third, we introduce the general principles of quantum probability theory as well as the justification for adopting these principles. Fourth, we present the steps for building a Hilbert space multidimensional model of measurement context effects. Fifth, we present an example application with a new large empirical data set. Last, we present a summary and describe possible extensions of the theory.

Jerome R. Busemeyer, Department of Psychological and Brain Sciences, Indiana University; Zheng Wang, School of Communication, The Ohio State University.

This research was supported by National Science Foundation Grants SES-1560554, SES-1560501, and AFOSR FA9550-15-1-0343. The authors thank Marius Usher, Emmanuel Pothos and an unknown reviewer for very helpful comments.

Correspondence concerning this article should be addressed to Jerome R. Busemeyer, Department of Psychological and Brain Sciences, Indiana University, 101 East 10th Street, Bloomington, IN 47405-7007. E-mail: jbusemey@indiana.edu

Defining Measurement Context Effects

Before we provide a rigorous definition for *measurement context effects,* we present an illustrative example. There are numerous studies of context effects like this example in the social science literature (see Harrison & McLaughlin, 1993; Pouta, 2004, for just a few empirical examples). Suppose that the relations among four psychological variables, labeled *A, H, I,* and *U* are being investigated. These variables could represent judgments about the *a*ttractiveness, *h*onesty, *i*ntelligence, and *u*nusualness of political candidates obtained from a large social media source (see, e.g., Steinberg, 2001); or patient symptoms concerning *a*nxiety, *h*yperactivity, *i*rritation, and *u*nruliness obtained from a large medical record source; or comments about whether a food product is *a*ppetizing, *h*ealthy, *i*nteresting, and *u*nfamiliar obtained from a large consumer choice source (see, e.g., Popper, Rosenstock, Schraidt, & Kroll, 2004). It may be difficult or impossible to obtain judgments from individuals on all four attributes simultaneously and suppose that only single attributes or pairs of attributes are judged at a time. For example, the single attribute A may be judged in isolation, or the pair AI or the pair AH may be judged together. Each single or pair of measurements provides a context for requesting judgments. How does the context (e.g., pairs being measured) affect the judgments that people make?

Various different kinds of context effects can occur with this kind of investigation. For an illustration of several of them in one simple (but artificial) example, consider Table 1.[1] In this case, each judgment is answered with a yes (Y) or no (N) answer. Judgments about a single variable in isolation form a 1 – *way* table with two frequencies for answers Y, N; a pair of attributes forms a 2 × 2 table containing relative frequencies for pairs of answers YY, YN, NY, NN. Table 1 presents 10 different contexts formed by two different 1 – *way* tables and eight different 2 × 2 tables. Each cell within a row is a relative frequency of answers, and the cells within a row sum to 1. For example, the 1 – *way* table labeled H is a context produced by measuring H alone, and the relative frequency of yes when H is asked alone equals .369. For another example, the pair of attributes AI forms the context for the 2 × 2 table produced by measuring A and I together with A first, and the relative frequency of yes to attribute A and then no to attribute I equals .175. Order may matter, and so the context AH with A presented first is different from the context HA with H presented first. For simplicity, we included only two of all four possible 1 – *way* tables, and eight of all 12 possible 2 – *way* tables. These 10 tables are sufficient to make our points.

Note that context effects occur when we compare the 1 – *way* tables to the marginal probabilities in the 2 – *way* tables. For example, the relative frequency of Y to H equals .369 when asked alone, but it equals (.345 + .125 = .470) when asked together with attribute A in the pair AH. The table also contains context effects produced by order of presentation. For example, the relative frequency of YY equals .345 to the pair AH when A is asked first, but it equals .286 for the pair HA when H is asked first. Other kinds of context effects are described later.

It is commonly assumed that the distributions in Table 1 can be derived from a single joint distribution across the four observed variables. In particular, categorical data models apply to only a single joint distribution (Agresti & Katera, 2011) of the four observed variables. In this case, a single 4 – *way* joint probability distribution is defined by four binary random variables (A, H, I, U) that generate 16 latent joint probabilities, $\pi(A = w \cap H = x \cap I = y \cap U = z)$, where, for example, A is a random variable with values $w = 1$ for yes and $w = -1$ for no, and similar definitions

Table 1
Two 1 − way *and Eight 2 × 2 Tables Produced by Yes–No Answers to Variables—Attributes A, H, I, and U*

Variable	Y	N	YY	YN	NY	NN
Single						
A	.446	.554				
H	.369	.631				
Pair						
AH			.345	.101	.125	.429
AI			.271	.175	.084	.469
AU			.115	.331	.269	.285
HI			.335	.035	.021	.610
HU			.296	.073	.088	.543
IU			.300	.055	.100	.545
HA			.286	.083	.143	.488
UI			.325	.059	.095	.521

Note. Y = yes; N = no. Pair YN, for example, refers to yes to the first attribute and no to the second. Each cell within a row is a relative frequency, and all the cells within a row sum to 1. The order of questions may matter, such that, for example, the HA table (H asked before A) may differ from the AH table (A asked before H).

hold for the other three random variables. The distributions produced by each context can be predicted from the 4 – *way* distribution by marginalization. For example, the relative frequency $p(YN|AI)$ is predicted by the theoretical marginal (pooled over the joint probabilities that are not involved)

$$\pi(A = 1 \cap I = -1) = \sum_x \sum_z \pi(A = 1 \cap H = x \cap I = -1 \cap U = z).$$

Note that this 4 – *way* joint distribution is completely general (nonparametric), because no independence or conditional independence or parametric distribution assumptions are imposed.

Measurement context effects can be defined more precisely by asking the following basic question: Does a single 4 – *way* joint probability distribution of the four binary variables exist that can reproduce Table 1? This question is essential for any Bayesian network model (see, e.g., Darwiche, 2009) based on the observed variables, because these models, when applied to Table 1, are all special cases of the single 4 – *way* distribution. If the answer is yes, then there are no context effects; if the answer is no, then some type of context effect has occurred.

We use the example shown in Table 1 to illustrate some of the constraints implied by the 4 – *way* joint distribution model. It turns out that there is no single 4 – *way* joint distribution that can reproduce Table 1. First, the 4 – *way* distribution requires the marginal distribution of a single random variable to be invariant across contexts. As we pointed out earlier, this requirement fails. For example, the marginal probability of yes to random variable H is not invariant: $p(Y|H) = .369$, which differs from $p(YY|AH) + p(NY|AH) = .470$. Table 1 contains other examples of violations of marginal invariance, depending on whether the attribute appeared first or second.

The latter fact brings up a second problem: The order that questions are asked changes the 2 – *way* distributions for some pairs. For example, the distribution for the context AH is not the same as the distribution for the context HA, and an order effect also occurs for the two contexts UI and IU. Order effects violate the commutative property required by the 4 – *way* joint probability model: in particular, $\pi(A = w \cap H = x) = \pi(H = x \cap A = w)$, and $\pi(I = y \cap U = z) = \pi(U = z \cap I = y)$.

It is interesting to notice that in this example, both marginal invariance and commutativity (no order effects) are satisfied by the four contexts AI, AU, HI, and HU. Suppose we restrict our question to only these four tables. Can a 4 – *way* joint distribution reproduce these four tables? Surprisingly, the answer is still negative. These four tables violate a consistency requirement of a single 4 – *way* joint distribution, called the Clauser, Horne, Shimony, and Holt (CHSH) inequality (for applications in psychology, see Bruza, Kitto, Ramm, & Sitbon, 2015; Dzhafarov & Kujala, 2012).[2] The CHSH is described in detail in Appendix A, and we need not go into details here, except to point out that the inequality implies the following restriction on the correlations

[1] The artificial data set allows us to present all the context effects with one clear and simple example. Later we present an application to a real data set, but it requires a more complex individual level of analysis.

[2] The Clauser, Horne, Shimony, and Holt inequality is closely related to the Bell inequality. The latter was derived for the Bohm paradigm using a pair of entangled spin $\frac{1}{2}$ photons, which was used to test the famous Einstein Podolsky Rosen (EPR) paradox.

(expectation of products) predicted for the tables of the 4 – *way* joint probability model:

$$\begin{gathered} -2 \leq CHSH \leq 2, \\ CHSH = E(A \cdot I) + E(H \cdot I) + E(H \cdot U) - E(A \cdot U), \end{gathered} \tag{1}$$

where $E(X \cdot Y)$ stands for the expected value of the product of two random variables. For example, the estimate for $E(A \cdot I)$ equals $[p(YY|AI) + p(NN|AI)] - [p(YN|AI) + p(NY|AI]$. (When variables are relabeled, there are several different ways to compute the CHSH, which can produce different answers, but they all must fall between -2 and $+2$.) Using the data shown in Table 1, the CHSH value equals 2.25, which exceeds the bound (≤ 2) required by the 4 – *way* joint probability model. In this case, it is the pattern of correlations in the 2×2 tables that violate the joint distribution. The CHSH is only one of a number of constraints that are required for a single joint distribution to reproduce a collection of contingency tables. Another type of correlation inequality, called the temporal Bell inequality, applies to 3 – *way* joint distributions (Leggett & Garg, 1985; Suppes & Zanotti, 1981).

The constraints just described on the 4 – *way* joint probability model are all necessary. But another important question is raised by these constraints: What is the sufficiency and what are the logical relations of these constraints? First, if we restrict our attention to the four contexts AI, AU, HI, and HU, then it can be shown that (1) if marginal invariance is satisfied and (2) the CHSH inequality holds in all permutations, then the 4 – *way* joint distribution can reproduce all four tables (Fine, 1982). Using the empirical tables constructed from the psychological variables, the result of the CHSH test is logically independent of tests of marginal selectivity or commutativity (violations of both can occur, violations of neither can occur, violations of only one can occur). Marginal selectivity can be violated without violating commutativity. For example, a 1 – *way* distribution might differ from the corresponding margin from a 2 – *way* table. However, noncommutativity (order effects) implies some violation of marginal selectivity, because the marginal probability of a variable changes depending on order. Dzhafarov and Kujala (2012) derived and provided a general summary of linear constraints that are necessary and sufficient for a single joint distribution for larger size joint distributions and larger numbers of contingency tables. Next we provide a statistical test that is related to the general constraints derived in Dzhafarov and Kujala.

We should point out that a Hilbert space multidimensional (HSM) model, constructed using the principles described later in this article, can perfectly reproduce all the findings in the artificial data in Table 1 using three less parameters than the 4 – *way* joint probability model (see Busemeyer & Wang, 2017, for details). However, one should not get the impression from this perfect fit that an HSM can fit any data. In fact, all HSM models must satisfy another inequality, called the Tsirelson bound, which states that $|CHSH| \leq \sqrt{2} \cdot 2$.

Table 1 is just an example using four binary variables that form a collection of 2×2 tables. However, the definition for the presence of measurement context effects can be stated much more generally. When judgments are collected from different contexts, they can often be summarized by collections of contingency tables or cross-tabulation tables. Suppose there are p different variables $(Y_1, \ldots, Y_p)$ that can be used to measure objects, or events, or people. Each variable Y_j can have n_j different values (it is not necessary that $n_j = 2$). It may not always be possible to measure all p variables at once, and perhaps, only subsets of variables $(Y_{k_1}, \ldots, Y_{k_s})$, $1 \leq s \leq p$, can be measured at once (it is not necessary that $s = 2$). Each subset of variables forms a context k of measurement. More than one context can be collected, which forms a collection of K data tables $(T_1, \ldots, T_k, \ldots, T_K)$, each collected under a different context k. Each table T_k is a joint relative frequency, or contingency, table based on a subset of variables. The general question is the following: Can the K tables formed from p variables be derived from a single p – *way* joint distribution of the p observed variables? If the answer is yes, then there are no context effects; if the answer is no, then there are context effects (this is a broader definition of context effects than that proposed by Dzhafarov & Kujala, 2016). As pointed out earlier, this question is essential for any Bayesian network model or categorical data model based on the observed variables, because these models are all require the validity of the single p – *way* joint distribution of the observed variables.

Suppose the data in Table 1 are based on a hypothetical sample of $N = 100$ independent observations for each 2×2 table (we can hold out the 1 – *way* tables for a later generalization test). Then it is unclear whether the violations of the 4 – *way* joint probability distribution, described earlier, are statistically significant. To address this issue, we use the same standard statistical test used in categorical data modeling for testing nested hypotheses (testing a general vs. a restricted special case; see Agresti & Katera, 2011): We compare the restricted 4 – *way* joint probability model to a general saturated model.[3] The saturated model simply assumes that we have eight independent 2 – *way* tables and that each table has four probabilities that sum to 1. The saturated model is completely general and unrestricted, and it perfectly reproduces the sample data. The 4 – *way* joint probability model has 15 free parameters, because the 16 joint probabilities are constrained to sum to 1. The saturated model has $8 \times 3 = 24$ parameters, because the probabilities sum to 1 within each table. The 4 – *way* joint probability model is nested within the saturated model, and the difference in number of parameters equals $df = 24 - 15 = 9$. Maximum likelihood methods can be used to estimate the parameters of each model, and $G^2 = -2 \times loglikelihood$ can be determined for each model. Then a log-likelihood ratio (i.e., G-square difference) test, which is the commonly used method to test departures of restricted model from the general model in categorical data modeling (Agresti & Katera, 2011), can be used to compare models. Using this method with $N = 100$ (hypothetical) observations per table produces a G-square difference equal to $G^2_{\text{diff}} = G^2_{\text{joint}} - G^2_{\text{sat}} = 18.04$. Under the null hypothesis, G^2_{diff} has a central chi-square distribution with $df = 9$. Using the chi-square distribution, the G^2_{diff} must exceed a criterion of 16.92 to be statistically significant (i.e., produce a $p < .05$). The $G^2_{\text{diff}} = 18.04$ exceeds the criterion 16.92, and the p value for this G^2_{diff} equals $p = .031$. Therefore, using this classical statistical test, the joint probability model is rejected and therefore is statistically different from the saturated model. Note that this is a nonparametric test that requires no conditional independence or parametric distribution assumptions.

[3] This is what Dzhafarov and Kujala (2013, 2016) called the context by default assumption.

This nonparametric method for testing a single $4 - way$ joint distribution model can be generalized and applied to $p - way$ joint distributions as long as there is a sufficient number of tables that allow the saturated model to have more parameters than does the joint distribution model. For example, if only the four 2×2 tables (AI, AU, HI, and HU) are included in the design, then the saturated model has only $4 \times 3 = 12$ parameters, which is fewer than in the $4 - way$ joint distribution model.[4] However, if four $1 - way$ tables, produced by measuring each attribute alone, are included in the design to form a collection of eight tables (A, H, I, U, AI, AU, HI, and HU), then the saturated model has 16 parameters, which leaves $df = 1$ for testing the joint probability model.

On the one hand, an advantage of this nonparametric statistical test of the joint distribution model is that it tests all the constraints imposed by the joint distribution model (including marginal invariance, absence of order effects, CHSH type inequalities, and others) with a single test. On the other hand, it does not isolate the particular property that is violated. We have developed more specific log-likelihood statistical tests that are designed to test a particular property (e.g., a test of order effects vs. a test of marginal invariance), but these additional tests are not described in detail here.

Before ending this section, we need to point out that the proposed test of a $p - way$ joint distribution to account for a collection of contingency tables formed by subsets of the p–variables does not rule out *all* Bayesian models. This is because a more general $p + q$ joint distribution can be postulated with additional q latent variables that are not necessarily observed. The proposed nonparametric method tests a $p - way$ joint distribution based on only the observed p–variables.

Empirical Review of Measurement Context Effects

During our presentation of Table 1, we mentioned several empirical studies of context effects on judgments, but these studies were not specifically designed to test the joint distribution hypothesis. Next we summarize a number of earlier experiments that were specifically designed to investigate the different types of measurement context effects that violate the joint distribution hypothesis.

Question order effects are usually investigated by using two contexts formed by the pair of tables (AB, BA) that vary the order that the attributes are evaluated. It has long been known that question order effects commonly occur with human judgments (Schuman & Presser, 1981). There are many examples of question order effects in the literature (Tourangeau, Rips, & Rasinski, 2000), and we mention only a few examples here. Wang and Busemeyer (2016a) investigated judgments about effectiveness of a public health service message for self versus another: Effectiveness was rated higher for self when self was evaluated first, but the difference disappeared when the other was evaluated first. More generally, Wang, Solloway, Shiffrin, and Busemeyer (2014) reviewed the widespread occurrence of order effects across 70 different national surveys. Different kinds of order effects occur in these national surveys. Moore (2002) identified four different types of question order effects defined by the comparison of the marginal distributions for the two attributes when they appear in the first position (the difference between A for the AB table and B for the BA table) versus the second position (the difference between A for the BA table and B for the AB table): Contrast effects occur when the second position increases the difference compared to the first; synthesis (consistency) effects occur when the second position decreases the difference compared to the first; additive effects occur when both marginals increase in the second position; subtractive effects occur when both marginals decrease in the second position.

Violations of marginal invariance are typically studied using two contexts: a single variable context A, in which an attribute A is measured alone, and another two-variable context BA, in which attributes B and then A are measured. These experiments are also called tests of the law of total probability, or tests for interference effects (Khrennikov & Haven, 2009). A violation of total probability is said to occur when the marginal probability for A obtained from the BA table is different from the table when A is measured alone. For example, Croson (1999) investigated marginal invariance using a prisoner dilemma game: In the A alone condition, participants simply decided to cooperate or defect; in the BA condition, participants first made a prediction about their opponent and then decided for themselves. The marginal probability to cooperate decreased in the BA context when compared to the probability of cooperation in the A alone context. Wang and Busemeyer (2016b) investigated marginal invariance using a category–decision task: In the A alone condition, participants simply decided to act by "attacking" or "withdrawing" against an agent, and in the BA condition, participants first categorized the agent as "good guy" or "bad guy" and then decided to act. The marginal probability of A ("attack") decreased in the BA context when compared to the probability to "attack" in the A alone condition. Kvam, Pleskac, Yu, and Busemeyer (2015) examined marginal invariance using a signal detection task in which participants had to make judgments about movement direction of a noisy target: In the A alone condition, participants simply judged their confidence that the target was moving left or right at time t_2; in the BA condition, participants first made a binary decision about the direction at time t_1 and later judged their confidence at time t_2. The marginal distribution of confidence shifted to less confidence in the BA (choice–confidence) condition compared to the A (confidence) alone condition. Other examples of marginal invariance violations were found in perceptual judgments with ambiguous figures (Conte et al., 2009).

A number of experiments have tested the pattern of correlations by using collections of 2×2 tables. Aerts, Gabora, and Sozzo (2013) and Bruza, Kitto, et al. (2015) tested the CHSH inequality required by a $4 - way$ joint distribution using four different 2×2 tables composed of pairs of four binary variables. Asano, Hashimoto, Khrennikov, Ohya, and Tanaka (2014) and Atmanspacher and Filk (2010) tested the temporal Bell inequality required by a $3 - way$ joint distribution using three different 2×2 tables composed of three binary variables. Although the observed correlations in these studies violated the required inequalities for a joint probability distribution, they coincided with violations of marginal invariance (Dzhafarov et al., 2016). However, clear evidence for violations of the CHSH inequality, after correcting for violations of marginal selectivity, was recently

[4] Nevertheless, $G^2_{\text{diff}} = 2.56$ after fitting the $4 - way$ model to Table 1, which reflects the violation of the Clauser, Horne, Shimony, and Holt inequality.

found by Cervantes and Dzhafarov (2018). Using a different type of design, Gronchi and Strambini (2016) also tested the CHSH inequality for a 4 – *way* joint distribution. Table 2 provides a summary of the empirical findings.

Quantum Models of Judgment and Decision

As mentioned earlier, the purpose of this article is to present a general theory of context effects. Our theory is based on quantum probability theory, and so it is useful to first discuss the psychological justification for taking this approach.

Classical probability theory evolved over several centuries, beginning in the 18th century with contributions by Pascal and Laplace. However, an axiomatic foundation for classical probability theory did not exist until Kolmogorov (1933/1950) provided one. Much of the theory was initially motivated by problems arising in physics, and later applications appeared in economics, engineering, insurance, statistics, and so forth. Classical probability theory is founded on the premise that events are represented as subsets of a larger set called the sample space. The adoption of subsets as the basis for describing events entails a logic—the logic of subsets—which is equivalent to Boolean logic (more generally, a sigma algebra of events). Boolean logic includes some strict laws, such as the closure property that if A, B are events in the same sample space, then $A \cap B$ is an event, and the axiom that events are commutative, $(A \cap B) = (B \cap A)$, and distributive, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. Social and behavioral scientists are generally trained to accept these axioms (explicitly or implicitly), and consequently most of us consider the theory as the only way to think about events and probabilities. How could there be other ways?

Looking back into history, scientists were faced with similar questions, such as with Euclidean geometry. How could there be any other geometry other than Euclidean? Nevertheless, we now have many applications of non-Euclidean geometry. Could this happen with probability theory too? Quantum mechanics was invented by a brilliant group of physicists in the 1920s in response to physical phenomena that seemed paradoxical from a classical physics perspective. This theory has revolutionized our world by giving us transistors, lasers, a foundation for chemistry, and many other accomplishments. It is interesting that, though not at first realizing it, these physicists invented an entirely new theory of probability. It was not clear that they invented a new probability theory until an axiomatic foundation was provided by Dirac (1930/1958) and Von Neumann (1932/1955). Quantum theory is founded on the premise that events are represented as subspaces of a vector space (called a Hilbert space, hence the name of our model). The adoption of subspaces as the basis for describing events entails a new logic—the logic of subspaces—which relaxes some of the axioms of Boolean logic. In particular, this logic does not entail having events always be commutative and distributive, and the closure property does not always hold.

It is often argued that although the microworld is quantum, the macroworld that we observe is classical, and so why would nature evolve a noncommutative human reasoning system? This confuses an important point. We are comparing classical versus quantum probability models of observed (epistemic) phenomena. We are not comparing classical versus quantum models of the unobserved physical (ontological) world. Even classical physical models of the world can produce event probabilities that are observed to be noncommutative. The latter can happen when only coarse epistemic measurements concerning the underlying ontic physical states are available (Graben & Atmanspacher, 2006). The reliance on epistemic measurements is particularly relevant to human judgments, which may be why they are so frequently found to be contextual. When context effects are present, classical theory requires a separate probability space for each context without any way to connect them together in a simple manner, whereas quantum theory provides an elegant way to connect contexts together into a single coherent probability model (Pothos, Yearsley, Shiffrin, & Busemeyer, 2017).

The principles from quantum theory actually resonate with deeply rooted psychological conceptions (Aerts, Broekaert, & Gabora, 2011). First, consider the enigmatic quantum principle of superposition—it captures the intuitive feelings of conflict, ambiguity, or uncertainty. A superposition state is maintained across potential choices until a decision must be reached, at which point the state collapses to a specific choice (Lambert-Mogiliansky, Zamir, & Zwirn, 2009). This behavior of changing from a superposition to a specific decision is similar to what Bohr called the wave–particle aspects of quantum mechanics. Next, consider the principle of complementarity—taking a measurement of a system constructs rather than records a property of the system, and the first question sets up a context that changes the answer to the next question; thus, answering a question disturbs the answers to subsequent questions and the order of questions is important (Wang, Busemeyer, Atmanspacher, & Pothos, 2013). In quantum physics, order-dependent measurements are said to be noncommutative, and quantum theory was especially designed for these types of noncommutative measures. Finally, consider the unique quantum concept of entanglement—the event $A \cap B$ may be observed, and another event $C \cap D$ maybe observed, but the event $A \cap B \cap C \cap D$ may not even exist, violating closure. Entangled states provide a basis for explaining violations of joint probability distributions produced by nonclassical patterns of correlations (Aerts et

Table 2
Experimental Tests for Measurement Context Effects

Marginal (B, AB)	Order (AB, BA)	Three correlations (AB, BC, AC)	Four correlations (AC, AD, BC, BD)
Croson (1999)	Moore (2002)	Asano et al. (2014)	Aerts et al. (2013)
Conte et al. (2009)	Tourangeau et al. (2000)	Atmanspacher & Filk (2010)	Bruza, Kitto, et al. (2015)
Wang & Busemeyer (2016b)	Wang et al. (2014)		Dzhafarov et al. (2016)
Kvam et al. (2015)	Wang & Busemeyer (2016a)		Cervantes & Dzhafarov (in press)

al., 2013; Bruza, Kitto, et al., 2015; Cervantes & Dzhafarov, 2018).

For these reasons, it turns out that quantum probability theory is not only useful for explaining physical phenomena but also provides useful new tools to model human behavior (Blutner & beim Graben, 2016; Bruza, Wang, & Busemeyer, 2015; Pothos & Busemeyer, 2012). Note that we are not necessarily proposing that the brain is some kind of quantum computer (see, e.g., Hameroff, 2013, for an example of this interpretation), and instead, we are only using the mathematical principles of quantum theory to account for human behavior. In fact, the formal computations of a quantum model may be implemented by some underlying type of neural network (Stewart & Eliasmith, 2013). Also note that we are not proposing a theory that necessarily competes with earlier social psychology explanations for context effects (e.g., Schwarz & Bless, 2007), and instead, we are providing a mathematical way to formalize these concepts.

Quantum probability theory has proven to be very useful for modeling the different kinds of context effects reviewed in the previous empirical review section. In particular, Wang et al. (2013) proposed a quantum model for question order effects that was later tested in a larger study using 70 national surveys by Wang et al. (2014). Pothos and Busemeyer (2009) proposed a quantum model for violations of marginal invariance that was later tested by comparing it with more traditional decision models in a larger experiment by Busemeyer, Wang, and Shiffrin (2015). Aerts et al. (2013) proposed a quantum entanglement model to account for their findings of violations of CHSH and marginal invariance found with their collection of four 2×2 tables.

The problem is that each of these previous modeling advances has developed specialized models for the particular design that was under investigation. What is needed is a more general theoretical framework from which it is possible to easily construct new specialized quantum models. Therefore, the purpose of this article was to present a more general theoretical framework that can be used for new model development. The general framework can then be used to construct new applications for new designs by following a standard program for model construction.[5]

Multidimensional Hilbert Space Theory

Basics of Quantum Probability Theory

Our theory is based on quantum probability theory, which is unfamiliar to most social and behavioral scientists. A good way to introduce quantum theory is to compare it with the more familiar classical probability theory.[6] Although both classical and quantum theories are applicable to infinite spaces, for simplicity, we limit this presentation to finite spaces.

Suppose we have p psychological variables (Y_i; $i = 1, \ldots, p$) and each variable, such as Y_i, produces one of a finite set of n_i values when measured. In classical theory, variable Y_i is called a random variable, and in quantum theory, Y_i is called an observable. The measurement outcome generated by measuring one of the p variables produces an event. For example, if variable Y_1 is measured and it produces the value y_i, then we observe the event $A = (Y_1 = y_i)$.

Classical theory begins with a universal set $\mathcal{H}$ containing all events, which is called the sample space, and quantum theory replaces this with a vector space $\mathcal{H}$ containing all events, which is called the Hilbert space. Classical theory defines an event $A = (Y_1 = y_i)$ as a *subset* of the sample space, whereas quantum theory defines an event $A = (Y_1 = y_i)$ as a *subspace* of the Hilbert space. Each subspace, such as A, corresponds to a projector, denoted P_A for subspace A, which projects vectors into the subspace. (A projector satisfies $P_A = P_A^\dagger = P_A^2$, where † represents the Hermitian transpose operator). The change from subsets to subspaces is where the logic of events differs between the two theories.

Classical theory assumes closure: If $A = (Y_1 = y_i) \in \Omega$ is an event and $B = (Y_2 = y_j) \in \Omega$ is another event, then $A \cap B \in \Omega$ is also an event in the sample space. By definition of intersection, the classical event $A \cap B$ is commutative $A \cap B = B \cap A$. In quantum theory, the events $A \in \mathcal{H}$, $\mathcal{B} \in \mathcal{H}$ may not be commutative, and if they are not, then the conjunction does not exist, and closure does not hold. Instead, quantum theory uses the more general concept of a sequence of incompatible events.

In quantum theory, a sequence of events, such as A and then B, denoted AB, is represented by the sequence of projectors $P_B P_A$. If the projectors commute, $P_A P_B = P_B P_A$, then the product of the two projectors is a projector corresponding to the subspace $A \cap B$, that is, $P_B P_A = P(A \cap B) = P_A P_B$, and the events A and B are said to be *compatible*. When the events are compatible, they share the same basis in the vector space for their representation. However, if the two projectors do not commute, $P_B P_A \neq P_A P_B$, then their product is a not a projector, and the events are *incompatible*. In this case, we need to evaluate the sequence using the product $P_B P_A$ (operating from right to left, first project on A with P_A and then project on B with P_B). When the events are incompatible, they require changing the basis of the vector space from one that represents event A to another that represents event B.

Classical theory defines a set function p that assigns probabilities to events, which is required to be an additive measure: $p(A) \geq 0$, $p(\Omega) =$ one. And if $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$. Quantum theory uses a unit length state vector, denoted $|\psi\rangle \in \mathcal{H}$, to assign probabilities to events as follows:

$$p(A) = \| P_A |\psi\rangle \|^2, \tag{2}$$

where the state vector $|\psi\rangle$ represents a person's beliefs about the events under investigation. Quantum probabilities also satisfy an additive measure: $p(A) \geq 0$, $p(\mathcal{H}) =$ one. And if $P_A P_B = 0$, then $p(A \vee B) = p(A) + p(B)$. In fact, Equation 2 is the unique way to assign probabilities to subspaces that form an additive measure for dimensions greater than 2 (Gleason, 1957).

According to classical theory, if an event A is an observed fact, then the conditional probability of event B is defined as

[5] Hilbert space multidimensional (HSM) theory is certainly not the most general development in quantum cognition. For example, Asano, Basieva, Khrennikov, Ohya, and Tanaka (2017) have used mixed states instead of pure states; Aerts (2009) has used Fock spaces instead of a Hilbert space, Denolf, Martínez-Martínez, Josephy, and Barque-Duran (2017) have used positive operator valued measurements instead of projectors; and Martínez-Martínez and Sánchez-Burillo (2016) have used Lindblad operators in conjunction with unitary operators. HSM is a theory that is not too simple but not too complex.

[6] See Busemeyer and Bruza (2012), Haven and Khrennikov (2013), Khrennikov (2010), and van Rijsbergen (2004) for introductions to quantum probability theory written for social and behavioral sciences.

$$p(B|A) = \frac{p(A \cap B)}{p(A)},$$

and so the joint probability of $A \cap B$ equals

$$p(A \cap B) = p(A) \cdot p(B|A).$$

The corresponding definition in quantum theory is

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

and so the probability of the sequence AB equals

$$p(AB) = p(A) \cdot p(B|A) = \|P_B P_A|\psi\rangle\|^2. \quad (3)$$

The commutative property of classical probability requires that $p(A) \cdot p(B\,|\,A) = p(B) \cdot p(A\,|\,B)$, but this commutative property does not hold for quantum theory, so that $p(A) \cdot p(B\,|\,A) \neq p(B) \cdot p(A\,|\,B)$ occurs when events are incompatible.

Extensions to sequences with more than two events follows the same principles for both classical and quantum theories. The probability of the joint event $(A \cap B) \cap C$ equals $p((A \cap B) \cap C)$ for classical theory, and the probability of the sequence $(AB)C$ equals

$$p((AB)C) = \|P_C(P_B P_A)|\psi\rangle\|^2 \quad (4)$$

for quantum theory.

Equations 2–4 provide the essential ideas that we need to compute the predictions from quantum probability theory. In the Building the Hilbert Space and Bulding the Projectors section, we provide more technical details about specifying the dimension of the Hilbert space and building the projectors to represent events. However, there are two technical issues that should be mentioned up front. First, in general, the Hilbert space is defined on the field of complex numbers. Some of the previous applications of quantum theory have relied on the full complex field (e.g., Kvam et al., 2015), whereas some others have been restricted to real numbers (e.g., Pothos, Busemeyer, & Trueblood, 2013). The theory proposed here allows for the full complex field, but the application presented in the Simple HSM Model section requires only the real field. Mathematically, working with the complex field is not any more complicated than is working with the real field—the equations remain exactly the same. However, the use of the complex field requires more free parameters than when working with the restricted real field. Second, in general, a projector associated with an event spans a subspace of dimension one or greater. Some of the previous applications of quantum theory have made use of multidimensional (greater than one) projectors for the individual values of a variable (e.g., Wang & Busemeyer, 2016a), whereas other applications have assumed that each value corresponds to a one-dimensional subspace (White, Pothos, & Busemeyer, 2014). Using one-dimensional subspaces (i.e., rays) provides the simplest possible representation, but this imposes severe restrictions on the model (described later in the Additional Tests of the HSM Model section), and often multidimensional subspaces are needed. The theory proposed here allows the use of multidimensional subspaces to represent the values of a variable.

Figure 1 provides a simple "toy" illustration of these basic principles restricted to a real vector space and using one-dimensional rays to represent distinct outcomes. Suppose a person is judging the quality of a piece of art from a personal perspective and then from the perspective of a friend. In this case, we have two variables: Y_1 is the self perspective with three values (yes, no, uncertain); Y_2 is the other perspective, also with three values (yes, no, uncertain). In previous research, we have shown that these two variables (self vs. other perspective) are incompatible (see Wang & Busemeyer, 2016a). As we described earlier, quantum theory represents incompatible events by changing the basis. Therefore we use the basis in the left panel to represent evaluations of self, and we use the basis shown in the right panel to represent evaluations of other. Each panel has three axes representing three responses to the question "Is this piece of art worth buying": *Y*es, *N*o, *U*ncertain. The vector *S* in the figure represents the state vector $|\psi\rangle$, which encodes the person's superposed state of belief about the artwork. The probability of the event *Y* from the personal perspective (i.e., I wish to buy artwork) is obtained by projecting the state vector *S* down onto the ray spanned by the basis vector *Y*, producing the projection denoted by *T*. The squared length of *T* provides the probability of choosing *Y* from the personal perspec-

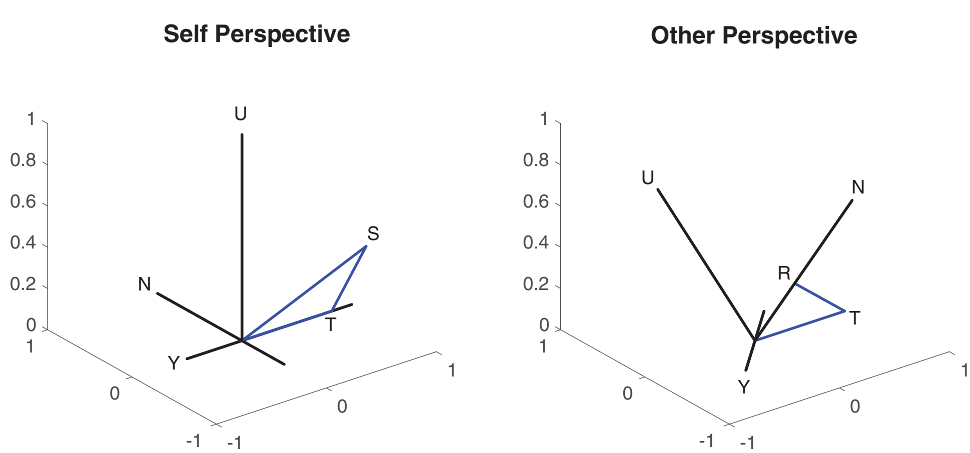

Figure 1. Three dimensional illustration of quantum probability computations for deciding yes (Y), no (N) or uncertain (U) from a self perspective (left panel) versus other person perspective (right panel). See the online article for the color version of this figure.

tive (which equals .67). The probability of the sequence of events, Y to the personal perspective and then N to the other perspective, is obtained by projecting T onto the ray spanned by the basis vector N from the other perspective, producing the second projection denoted R. The probability of this sequence is obtained by squaring the length of R (which equals .17). The probability of the reverse sequence, responding N to the friend's perspective and then Y to the personal perspective, turns out to equal .20, exhibiting a small order effect. Of course, we do not propose that people actually rotate vectors in their mind. Instead, these computations can be performed by some underlying neural network system (Stewart & Eliasmith, 2013).

All of the previously mentioned quantum models for context effects use these same rules (Equations 2–4) for computing probabilities. However, as we mentioned before, the previous models were specialized for a particular application. Two problems need to be solved to form a more general program for constructing models. One is to build the structure of the Hilbert space $\mathcal{H}$, and the second is to formulate the projectors for events. One of the main contributions of the present work is to describe a conceptual program of steps for solving these two problems. Next we describe the abstract principles, but later we illustrate these with a simple application to a new experiment.

Building the Hilbert Space

The structure of the Hilbert space is determined by the compatible variables. Consider three variables: Y_1 with n_1 values, Y_2 with n_2 values, and Y_3 with n_3 values.

Suppose the first two variables are compatible. In this case, we can define all $n_1 \cdot n_2$ conjunctions $(Y_1 = y_i \cap Y_2 = y_j)$, $i, j = 1, \ldots, n)$. Each of these conjunctive events corresponds to a subspace in the Hilbert space. Each subspace can be multidimensional, but the smallest subspace is a ray, which is a one-dimensional subspace spanned by a single vector. It is not necessary to use rays, but if they are used, then the minimum dimension to represent the two compatible variables is $N = n_1 \cdot n_2$. After determining the dimension N from the compatible variables, a common N-dimensional basis can be chosen to represent all the conjunctions. It is simplest to use the canonical basis to represent the coordinates of the chosen basis, in which case each basis vector can be simply represented by a $N \times 1$ column matrix $e_l = [0 \quad 0 \ldots 1 \ldots 0 \quad 0]^\dagger$, with one located in position l and zeros otherwise. Assuming the minimum dimension, the column matrix e_l represents the lth conjunction.

Now suppose Y_3 is incompatible with the conjunctions formed by Y_1, Y_2. Under this assumption, the dimension remains the same for all three Y_1, Y_2, Y_3 variables as it was for the two compatible Y_1, Y_2 variables. In other words, if an N dimensional space was used to represent Y_1, Y_2, then the same N dimensional space can be used for all three Y_1, Y_2, Y_3. The same N-dimensional canonical basis can be chosen to represent all of the conjunctions for the compatible events. However, to evaluate events involving the incompatible $(Y_3 = y_k)$ variable Y_3, the cognitive system needs to rotate to a new basis within the same N-dimensional space. This is where the quantum representation simplifies the representation of the variables. A Bayesian model would have to form a $n_1 \cdot n_2 \cdot n_3$-dimensional joint probability space for the three random variables Y_1, Y_2, Y_3. If Y_3 is incompatible with Y_1, Y_2, then the quantum model could represent all three variables in smaller $n_1 \cdot n_2$-dimensional space.

More generally, suppose we have at most q compatible variables. Then we can form conjunctions of all values for all q-compatible variables. The number of these conjunctions determines the minimum dimension N of the Hilbert space. If there are other incompatible variables, the system needs to rotate to a new basis within the N-dimensional space for each incompatible variable.

The dimension N needed to fit the data may be larger than the minimum. In other words, events can be represented by multidimensional subspaces. Therefore we start with the minimum dimension and add dimensions as needed. But dimensions are added only if they are required according to statistical model comparisons favoring higher over lower dimensional models (similar to multidimensional scaling or factor analysis programs).

Once a basis for the N-dimensional space is chosen, the coordinates of the state vector, $|\psi\rangle$, can be represented by an $N \times 1$ column matrix ψ. This state represents a person's superposition state of beliefs about the variables. Usually, this state of the judge is unknown to the researcher, but it can be estimated by best fits to the data. In general, each coordinate can be complex, containing a magnitude and a phase. Therefore, if the dimension equals N, then the state requires $2 \cdot N$ parameters. However, the state must satisfy the unit length constraint $\psi^\dagger \psi = 1$, which constrains one magnitude. Also, one phase can be arbitrarily fixed without any effect on the choice probability. In sum, only $2 \cdot (N - 1)$ parameters are estimated from the data.

Building Projectors

Consider once again the Hilbert space structure with variables Y_1 and Y_2 compatible with each other but incompatible with Y_3. Again, it is simplest to use the canonical basis to represent the compatible variables. Using this basis, the projector for the event $(Y_1 \geq y_i)$ is simply an $N \times N$ diagonal (indicator) matrix M_i with ones located in rows corresponding to the conjunctions satisfying $(Y_1 \geq y_i)$. The projector for the event $(Y_2 \geq y_j)$ is simply an $N \times N$ diagonal (indicator) matrix M_j with ones located in rows corresponding to the conjunctions satisfying $(Y_2 \geq y_j)$.

The projector for any events involving Y_3 require rotating to a new basis. An $N \times N$ unitary matrix, denoted U, is used to compute this rotation (a unitary matrix satisfies $U^\dagger \cdot U = I$). Each column of U represents an orthonormal basis vector, which is used to represent events involving Y_3. The projector for event $(Y_3 = y_k)$ is $P(Y_3 = y_k) = U \cdot M_k \cdot U^\dagger$, where M_k is a diagonal matrix with ones located in rows satisfying the event $(Y_3 = y_k)$ and zeros otherwise. The key problem is to construct the unitary matrix U. Usually, this matrix is initially unknown.

A general solution for low dimensional spaces is the following (see Appendix B or Busemeyer & Wang, 2017, for technical details). Any unitary matrix can be formed by an exponential transformation of a Hermitian matrix (a Hermitian matrix satisfies $H = H^\dagger$). In general, the Hermitian matrix has N diagonal entries that are real and $N \cdot (N - 1)/2$ off diagonal entries that can be complex. However, adding a constant to all the diagonal entries has no effect on the choice probabilities, and so one diagonal entry can be set to a fixed value. In sum, only $(N^2 - 1)$ parameters are estimated for a Hermitian matrix, so the general solution is to estimate the free parameters of the Hermitian matrix that best fit the data. This solution is feasible for low dimensional spaces. For high dimensional spaces, sparse matrices that are built from a small number of parameters can be used (see, e.g., Kvam et al., 2015; Wang & Busemeyer, 2016a).

The Hilbert Space Multidimensional Program

An HSM model is built using the following programmatic steps.[7] All these steps are illustrated in the next section using a concrete application to data from a new experiment on context effects (see Busemeyer & Wang, 2017, for technical details).

First, the researcher needs to determine which variables or attributes are commutative and which are not. If two events are compatible, then the events can be defined simultaneously and we can form all the conjunctions of two events, but if they are incompatible, they must be evaluated sequentially and we need to change from one basis to another basis to model the evaluation of the sequence. One way to determine compatibility is to observe whether a pair of variables produces order effects. Alternatively, one can statistically compare competing models with different hypothesized compatibility relations. Referring to Table 1, we see that attributes A and H do not commute, and so they are incompatible. We do not have a test for order effects for variables AI in Table 1, but we could test a model that assumes they are compatible.

Second, the dimension N of the Hilbert space is determined. Referring to the artificial example in Table 1, if there are at most two compatible variables and the variables are binary, then the minimum dimension equals four. Given the compatibility relations, an HSM modeling procedure can begin with the lowest possible dimension and increase the dimension only as required by model comparisons that favor a higher dimension.

Third, a basis is selected for representing the coordinates of the state vector $|\psi\rangle$. As described earlier, the simplest choice is to start with the canonical basis for the compatible variables. Once a basis is chosen, the coordinates of the state vector, represented by the $N \times 1$ column matrix ψ, can be estimated by best fits to the data. The state $|\psi\rangle$ is then represented by an $N \times 1$ matrix ψ with coordinate ψ_j representing the amplitude assigned to basis vector e_j.

Fourth, a unitary matrix is built to rotate from one basis to another for each pair of incompatible variables. If we construct the unitary matrix in a completely general manner, then $(N^2 - 1)$ parameters are estimated for each unitary matrix (see Appendix B for details; also see Busemeyer & Wang, 2017). For example, referring to the artificial example in Table 1, if we assume that variables AH are incompatible, then we construct a 2×2 unitary matrix to rotate from A to H.

Fifth, the quantum probability for a sequence of measurements is computed using Equations 2–4. Using the predicted probabilities, the model is used to compute the log-likelihood of the data. The parameters for the state and the Hermitian operators are estimated from the data using maximum likelihood, and the result is used to compute $G^2 = -2 \cdot loglikelihood$ statistics for model comparison.[8] The number of model parameters is determined by the number of parameters used to build the state vector plus the number of parameters used to estimate the unitary (rotation) matrices.

Sixth, the fit of the model returns parameters for the state that can be used to describe the probability distribution over a variable as if it were measured alone (free of context of other variables) and also the parameters of the unitary rotations that describe the relations between incompatible variables.

Seventh, an HSM model allows many opportunities for very strong generalization tests of the model. For example, if there are three variables and two of them are incompatible, then after estimating the model parameters from an HSM model for a collection of 2×2 tables, the same model and parameters can be used to make new predictions for new tables that were not included in the original design, such as smaller 1 – *way* tables or larger 3 – *way* tables. Even stronger a priori tests (e.g., estimating the model parameters with a collection of 2×2 tables and testing on a 4 – *way* joint distribution) can be designed for studies with a larger set of variables. In short, these models provide for many strong empirical tests.

A New Empirical Application

This section applies HSM modeling to a real experiment that was designed in a manner similar to that for the artificial example. A total of 184 participants made judgments on four attributes of antismoking public service announcements (PSAs). They were asked to judge how persuasive (P), believable (B), informative (I), and likable (L) they perceived various PSAs to be. The PSAs were in the form of a single static visual image with a title. Each person judged 16 different PSAs: One stimulus type included eight examples warning about smoking causing death (Death PSAs), and the other stimulus type included eight PSAs warning about smoking causing health harm (Harm PSAs). Each participant judged each PSA under 12 contexts: six combinations of two attributes with the attributes presented in two different orders. For example, one context was PI, where the participants answered the question of whether the PSA was persuasive and informative by choosing either YY, YN, NY, or NN (where, e.g., YN means Yes to persuasive and No to informative). Thus, each person provided responses to 16 (PSAs) $\times 12$ (contexts) $= 192$ questions, which were presented in a randomized order across participants. Altogether, this produced a total of 184 participants $\times$ 192 judgments per person $= 35{,}328$ observations.

The aggregate results are presented in Tables 3 and 4. The results are presented separately for each stimulus type and order pooled across participants. For example, when the Death PSA was presented, the relative frequency of Y to persuasive and then N to likable was .21, and the corresponding result for the Harm PSA was .18. Each 2×2 table for a pair of attributes and type of stimulus is based on $184 \cdot 8 = 1{,}472$ observations. However, (to simplify the presentation), this table of pooled results ignores

[7] The word *program* here refers to the set of procedures that we formulated to build Hilbert space multidimensional models. We are in the process of writing generalizable computer codes to implement the conceptual program described here, which will be published separately. At this point, we have created computer codes for collections of 1 – *way*, 2 – *way*, and 3 – *way* tables. The current codes are written in MATLAB and are available at http://mypage.iu.edu/~jbusemey/quantum/HilbertSpaceModelPrograms.htm

[8] Currently, we use a particle swarm method to estimate parameters to avoid local minimum.

important individual differences. All of the subsequent analyses were conducted at the individual level of analysis.[9]

Test of the Joint Probability Model

To determine whether there are context effects, we conducted a statistical chi-square test of a 4 – *way* joint probability model based on four binary random variables (*P, B, I,* and *L*) corresponding to those used in the experiment. Each individual produced a table in the same form as that for Table 3 but pooled across orders, so that question order was not a factor for this test. This reduced the number of tables to six per individual, and each individual 2 × 2 table had 16 observations (192 observations in total for both types of stimuli). The joint probability model states that the six rows of 2 × 2 tables for each stimulus type are produced by a joint distribution, $\pi(P = w \cap B = x \cap I = y \cap L = z)$, where $w = -1, 1$, $x = -1, 1$, $y = -1, 1$, and $z = -1, 1$, that has $16 - 1 = 15$ free parameters per stimulus type, or 30 parameters altogether. A completely unconstrained saturated model requires three parameters for each 2 × 2 table, producing a total of 18 parameters per stimulus type, or 36 parameters altogether. Using maximum likelihood estimation for each person, we computed the G^2_{sat} and G^2_{joint} for each person. A total of 44 participants produced a statistically significant G^2 difference based on six degrees using the critical cutoff equal to 12.59 for $p < .05$. A quantile–quantile plot of the observed G^2 differences, $G^2_{\text{diff}} = G^2_{\text{joint}} - G^2_{\text{sat}}$ versus the chi-square value, predicted by the chi-square distribution under the null hypothesis is shown in Figure 2. As can be seen in Figure 2, the observed G^2_{diff} exceeds the expected amount for large values of the predicted chi-square. We computed a lack of fit from the null chi-square distribution by comparing the observed versus expected frequencies using categories defined by cutoffs [0, 5, 10, 35]. The expected frequencies were [84, 77, 23], but the observed frequencies were [48, 75, 61]. We used a G^2 difference test of the difference between the expected and observed frequencies, and the G^2 difference was statistically significant, $G^2(2) = 78.84$, critical cutoff = 5.99, $p < .001$. We conclude that the 4 – *way* joint probability model systematically deviates from the observed results for a substantial number of individuals.

Table 3
Observed Relative Frequencies of Pairs of Answers for Death PSA

Order and attributes	YY	YN	NY	NN
Death PSA Order 1				
PI	.53	.17	.06	.23
PB	.61	.10	.07	.21
PL	.50	.21	.06	.23
IB	.54	.08	.12	.26
IL	.44	.18	.12	.25
BL	.50	.19	.08	.23
Death PSA Order 2				
IP	.52	.16	.08	.23
BP	.61	.08	.08	.24
LP	.50	.19	.07	.24
BI	.53	.07	.14	.26
LI	.44	.18	.13	.26
LB	.49	.18	.09	.23

Note. PSA = public service announcement; Y = yes; N = no; P = persuasive; I = informative; B = believable; L = likable. Pair YN, for example, refers to yes to the first attribute and no to the second.

Table 4
Observed Relative Frequencies of Pairs of Answers for Harm PSA

Order and attributes	YY	YN	NY	NN
Harm PSA Order 1				
PI	.44	.15	.04	.38
PB	.46	.11	.06	.38
PL	.38	.18	.08	.36
IB	.42	.08	.11	.40
IL	.34	.17	.11	.38
BL	.34	.20	.10	.36
Harm PSA Order 2				
IP	.43	.12	.06	.38
BP	.46	.09	.08	.37
LP	.38	.17	.09	.36
BI	.42	.08	.11	.39
LI	.31	.17	.14	.38
LB	.37	.17	.10	.36

Note. PSA = public service announcement; Y = yes; N = no; P = persuasive; I = informative; B = believable; L = likable. Pair YN, for example, refers to yes to the first attribute and no to the second.

Comparisons Between Bayesian Network and HSM Models

Any Bayesian network model, based on the four random variables *P, B, I,* and *L* is a special case of the 4 – *way* joint probability model, which implies that there is also some systematic deviation from any Bayesian network type of model. However, there may also be systematic deviations from an HSM model. Therefore, it is important to compare the fits of Bayesian network versus HSM models to the individual participant data. Maximum likelihood estimates and G^2 statistics were computed by fitting each model to the 192 observations separately for each of the 184 participants. Our goal was to select a model that was accurate yet relatively simple, and we needed to compare nonnested models; therefore we used the Bayesian information criterion (BIC) for model comparison. The BIC is defined as $G^2 + v \cdot ln(N)$, where $G^2_{\text{model}} = -2 \cdot loglikelihood$, v is the number of model parameters, and N is the number of observations, and so for this application $ln(192) = 5.2575$. The model with the lowest BIC is preferred.

Simple Bayesian Network Model

There is a large number of possible Bayesian network type of models that one can construct for this application. We chose the

[9] Rather than conducting individual level analyses, we could formulate a hierarchical Bayesian model that introduces assumptions about the distribution of individual differences and priors on these hyperparameters. At this early stage, we do not think this is a good place to start for comparing complex models such as the 4 – *way* joint probability model (184 participants with $15 \cdot 2 = 30$ parameters for each participant) because of lack of empirical support for specific parametric distributions of individual differences and lack of informative priors on the hyperparameters for these complex models. We did not want to confound our test of core models (classical versus quantum) with arbitrary assumptions about individual difference distributions and hyperpriors.

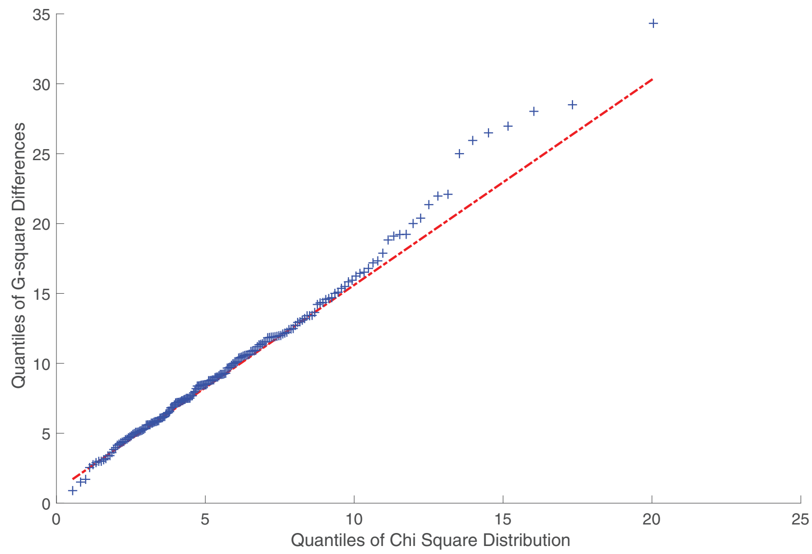

Figure 2. Quantile - quantile plot of the observed G-square versus expected chi square. Deviations fron the unit slope line indicate deviations from the null hypothesis that there are no context effects. See the online article for the color version of this figure.

following model because (a) it is simple and (b) it makes assumptions that match design of the stimuli and responses to the stimuli for this experiment. We note, however, that our conclusions are restricted to these particular models, and there may be other Bayesian network models that perform better than does the one here.

For the Bayesian network type of model, we assumed that the two attributes informative (I) and believable (B) are exogenous factors determined by the type of PSAs. Therefore, each type of stimulus produced a 2 – *way* joint distribution with four joint probabilities, $\pi(I = x \cap B = y \mid stimulus)$, $x = -1, 1$, $y = -1, 1$, and these were allowed to vary across the two types of stimuli. This produces $(4 - 1) \cdot 2 = 6$ parameters per stimulus type. Next we assumed that the response to attributes persuasive (P) and likable (L) depended on the stimulus attributes I and B, and this dependency was represented by the conditional probabilities $\pi(P = w \cap L = z \mid I = x \cap B = y)$, for $w = -1, 1$, and $z = -1, 1$. However, this model produces the same number ($15 \cdot 2 = 30$) of parameters as does the 4 – *way* joint probability model. To simplify the model, we assumed conditional independence, so that

$$\pi(P = w \cap L = z | I = x \cap B = y) = \pi(P = w | B = y) \cdot \pi(L = z | I = x).$$

We also assumed that the two conditionals, $\pi(P = w \mid B = y)$ and $\pi(L = z \mid I = x)$, did not depend on the stimulus type. Therefore, each of the two conditionals produces two parameters. Altogether, this model entails $(4 - 1) \cdot 2 + (2 \cdot 2) = 10$ parameters. We refer to this as the 10-parameter Bayesian model.

We also examined a constrained version of this Bayesian model that imposed a type of symmetry on the conditional probabilities:

$$\pi(P = +1 | B = -1) = 1 - \pi(P = +1 | B = +1)$$

$$\pi(L = +1 | I = -1) = 1 - \pi(L = +1 | I = +1).$$

This model requires only $(4–1) \cdot 2 + 2 = 8$ parameters. We refer to this as the eight-parameter Bayesian model. The purpose of examining this model was to provide a Bayesian model with the same number of parameters and similar constraints as has the quantum model described next.

Simple HSM Model

The simplest possible HSM model was applied to the real data from our experiment. First, we assumed that the attributes believable (B) and informative (I) are compatible. This assumption implies that the order of measurement for these two attributes does not matter and that the conjunction $(B = x \cap I = y)$ of their binary values can be defined. This is consistent with a lack of effect of the order effects of the two attributes in the aggregated data. Second, we assumed that persuasive (P) is a rotation of believable (B) and that likable (L) is a rotation of informative (I). In other words, B, P were assumed to be incompatible and so were I, L. This assumption was also consistent with order effects found at the aggregate level for these variables.

For the simplest case, we assumed that each conjunction for a compatible pair is represented by a single basis vector. Together, these compatibility assumptions imply that we required a –four-dimensional space, with four different bases: one basis defined by the four joint events $(B = x \cap I = y)$, a second basis defined by the four events $(B = x \cap L = z)$, a third basis defined by the four events $(P = w \cap I = y)$, and a fourth defined by four events $(P = w \cap L = z)$.

We chose to represent the state and projectors using the basis described by the B, I events $(B = x \cap I = y)$. Choosing this basis, the basis vector for the event $(B = +1 \cap I = -1)$, for example, can be represented by a 4×1 column matrix of coordinates

$[0 \ \ 1 \ \ 0 \ \ 0]^{\dagger}$. To reduce the number of model parameters to a minimum, we restricted the coordinates of the state to be a real valued 4×1 column matrix ψ with unit length $\psi^{\dagger}\psi = 1$:

$$\psi = \begin{bmatrix} \psi_{+1,+1} \\ \psi_{+1,-1} \\ \psi_{-1,+1} \\ \psi_{-1,-1} \end{bmatrix}.$$

The coordinate $\psi_{x,y}$ corresponds to the event $(B = x \cap I = y)$. To account for the effect of type of stimulus, we allowed the state vector to vary across the two types of stimuli, ψ_{Death} and $\psi_{\mathcal{H}arm}$.

The rotation matrices were based on the following simple type of rotation:

$$U = \begin{bmatrix} cos(\pi \cdot \theta) & -sin(\pi \cdot \theta) \\ sin(\pi \cdot \theta) & cos(\pi \cdot \theta) \end{bmatrix}. \quad (5)$$

Different parameters, θ_{PB} and θ_{LI}, were used to define the parameter θ in Equation 5 to produce rotation matrices U_{PB} for the PB incompatible pair and U_{LI} for the LI incompatible pair (π in Equation 5 refers to the mathematical constant for a circle). The transitions between basis states for incompatible variables B and P, as well as the transitions between basis states for incompatible variables I and U, should only depend the unitary transformation U, and the latter depends only on the variables and is independent of the stimulus.

To define the projectors for different events, we first define an indicator matrix for a yes response as $M_{+1} = diag[1 \ \ 0]$ and a no response as $M_{-1} = diag[0 \ \ 1]$. To form the matrices for the four-dimensional space, it is efficient to use the Kronecker (tensor) product operator, denoted $\otimes$ (see Appendix C for a review). Then the projectors for events were defined by

$$\begin{aligned} &P(B = x) = M_x \otimes I, \\ &P(I = y) = I \otimes M_y \\ &P(P = w) = \left(U_{PB} \cdot M_w \cdot U_{PB}^{\dagger}\right) \otimes I \\ &P(L = z) = I \otimes \left(U_{LI} \cdot M_z \cdot U_{LI}^{\dagger}\right). \end{aligned}$$

Finally, the probabilities within each $2 - way$ table were computed from the quantum rule for sequential events. For example, the probability of obtaining x on B and then z on L equals

$$p(B = x, L = z) = \| P(L = z) \cdot P(B = x) \cdot \psi \|^2.$$

We tested the prediction from the HSM model that the parameters of the unitary matrix are independent of the stimulus by comparing a model that allowed θ_{PB}, θ_{LI} to change across stimuli with a model that constrained these to be the same across stimuli. The constrained HSM model requires estimating a total of $(3 \cdot 2) + 2 = 8$ parameters, which we refer to as the eight-parameter HSM model. The model that allows θ_{PB}, θ_{LI} to change across stimuli adds two more parameters, and we refer to this as the 10-parameter quantum model.

In sum, the HSM model starts with the four-dimensional BI basis, which provides the coordinates that define the distribution ψ. The coordinates of ψ are then used to compute the $2 - way$ joint distribution for the BI table. The distributions for all of the other $2 - way$ tables are generated by rotating the basis of the four-dimensional space using the unitary matrices U_{LI} and U_{BP}. The new basis produced by rotation provides coordinates that are then used to compute the response probabilities for another table.

Results of Model Comparisons

Recall that the $4 - way$ joint probability model has 30 parameters, the Bayesian models have 10 or eight parameters, and the HSM models also have 10 or eight parameters.

First, consider comparisons within the Bayesian models. When comparing the 10-parameter Bayesian model to the joint probability model, we found that all 184 participants produced BICs favoring the 10-parameter Bayesian model. When comparing the 10- versus eight-parameter Bayesian models, we found that 102 participants produced BICs favoring the 10- over the eight-parameter model. Thus, the 10-parameter Bayesian model is the preferred model in this class.

Second, consider comparisons within the HSM models. When comparing the 10-parameter HSM model to the joint probability model, we found that all 184 participants produced BICs favoring the 10-parameter quantum model. When comparing the 10- versus eight-parameter HSM models, we found that 158 participants produced BICs favoring the eight- over the 10-parameter model. Thus, the eight-parameter HSM model is the preferred model in this class.

Third, consider comparing the preferred 10-parameter Bayesian model to the preferred eight-parameter HSM model. We found that 115 participants produced BICs that favored the eight-parameter HSM model over the 10-parameter Bayesian model. However, we also compared the two eight-parameter models because we can directly compare G^2 without any penalty for these models (they use the same number of parameters, so the penalty is the same for both). We found that 127 participants produced G^2 that favored the HSM model over the Bayesian model.

The predictions generated by the eight-parameter Markov model compared with the observed proportions, pooled across participants, are presented Figure 3. Similarly, Figure 4 shows the results for the HSM model. As can be seen in the figures, the constrained HSM model does a much better job of predicting the pooled results when compared to the Markov model. The most important errors

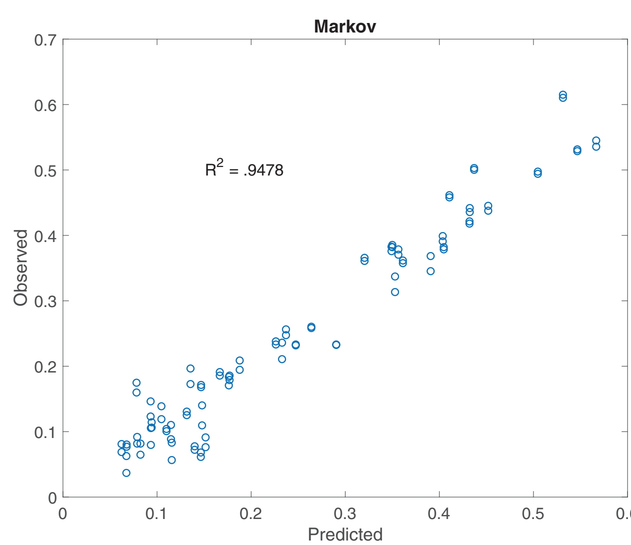

Figure 3. Observed proportions versus predictions by Markov model. See the online article for the color version of this figure.

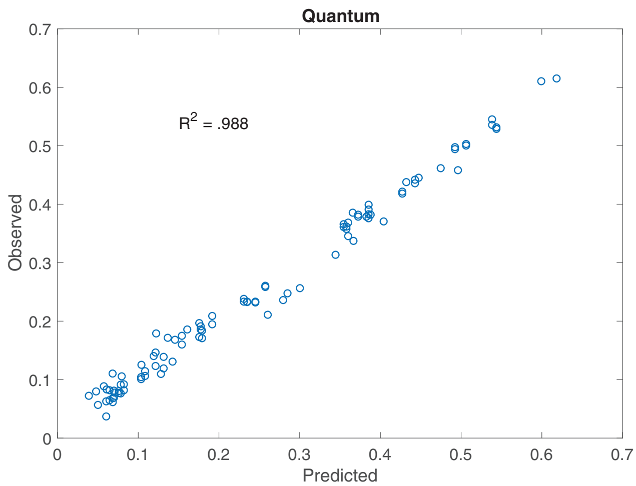

Figure 4. Observed proportions versus predictions by Quantum model. See the online article for the color version of this figure.

occur for the incompatible variables, where we constrained the model to use the same parameters across stimulus types.

Additional Tests of the HSM Model

The simplest possible subspaces, one-dimensional rays, were used in this application to represent each conjunction of two compatible events. For example, the conjunctive event ($B = +1 \cap I = -1$) was represented by a one-dimensional ray spanned by the vector represented by the 4 × 1 column matrix of coordinates $[0 \quad 1 \quad 0 \quad 0]^{\dagger}$. The use of one-dimensional rays to represent conjunctive events leads to the following strong predictions: The 2 × 2 matrix of transition probabilities $p(P = w | B = x)$ for each individual are equal to the 2 × 2 matrix of produced by squaring the magnitudes of the unitary matrix in Equation 5 using the parameter θ_{PB} for the incompatible pair P, B (see Appendix D for derivation). Furthermore, the 2 × 2 matrix of transition probabilities $p(B = x | P = w)$ is predicted to be identical to that for $p(P = w | B = x)$. This is called the law of reciprocity in quantum probability theory (Peres, 1998). Note that the Bayesian models are not required to satisfy this symmetry property of the HSM model. These predictions can be approximately checked by comparing the conditional probabilities computed from the aggregate observed relative frequencies. The predictions do not exactly follow the law of reciprocity anymore because of aggregation, but they remain close to this prediction. The predictions for the conditional probabilities based on the incompatible pair I, L are similar.

The results of this comparison, pooled across order and stimulus type, are shown in Table 5. These results are most challenging for the simple HSM model used in this application for the following two reasons. First, this simple model approximately satisfies the law of reciprocity, and second, the constrained HSM model was forced to use the same unitary matrices for both types of stimuli. As can be seen in Table 5, the observed conditionals match the general pattern predicted by the law of reciprocity fairly well. This is not always the empirical case—for example, Boyer-Kassem, Duchêne, and Guerci (2016) reported very large deviations from reciprocity. Some deviations from reciprocity are expected because they can be fabricated by aggregation across individuals. However, the observed deviations from reciprocity were a bit larger than the predicted deviations from reciprocity. This difference in deviations from reciprocity suggest that a higher dimensional model (e.g., a plane instead of a ray) may provide a more accurate representation each conjunction.

Interpretation of Parameters

The HSM model provides two sets of model parameters for each participant. The distributions of these parameters provide an interpretation of the data from the view of the HSM model. One set, which is based on the state ψ, describes the probabilities of responding "yes" to each variable when the variable is measured alone (free from context effects of other attributes). Figure 5 presents the relative frequency distribution of these response probabilities for each type of stimulus. For example, the bottom left panel shows the relative frequencies for "yes" responses to P attribute with the death appeal PSAs, and the right lower panel shows the results for the harm appeal PSAs. As can be seen in the figure, the probabilities are widely spread out among participants, but the probability of answering "yes" was generally higher for the death appeal PSAs. Similarly, we can compare the parameter distributions for the other three attributes between the two types of PSAs with different appeals (see Figure 5). In general, participants responded more positively toward death appeal PSAs on all the four attributes but clearly more so for the attributes of believable and persuasive.

The second set is based on the parameters θ_{PB}, θ_{LI} used for the rotation matrices for the two incompatible variables (recall that these are the same for the two types of stimulus). The squared magnitude of the coefficients within the unitary rotation matrices describe the probability of transiting from one basis to another, that is, transitioning from basis vectors for I to basis vectors for L and

Table 5
Observed and Predicted Probability of Column Values Conditioned on Row Values for the P,B and I,L Attributes

Variable	$B = 1$	$B = -1$
$P = 1$	.85/.89	.19/.14
$P = -1$	.15/.11	.81/.86
	$P = 1$	$P = -1$
$B = 1$	.88/.91	.24/.16
$B = -1$	.12/.09	.76/.84
	$L = 1$	$L = -1$
$I = 1$	.69/.74	.28/.27
$I = -1$	.31/.26	.72/.73
	$I = 1$	$I = -1$
$L = 1$	.75/.76	.36/.30
$L = -1$	.25/.24	.64/.70

Note. Observed probability of column values conditioned on row values appear above the slash, and predicted probability of column values conditioned on row values appear below the slash. P = persuasive; I = informative; B = believable; L = likable.

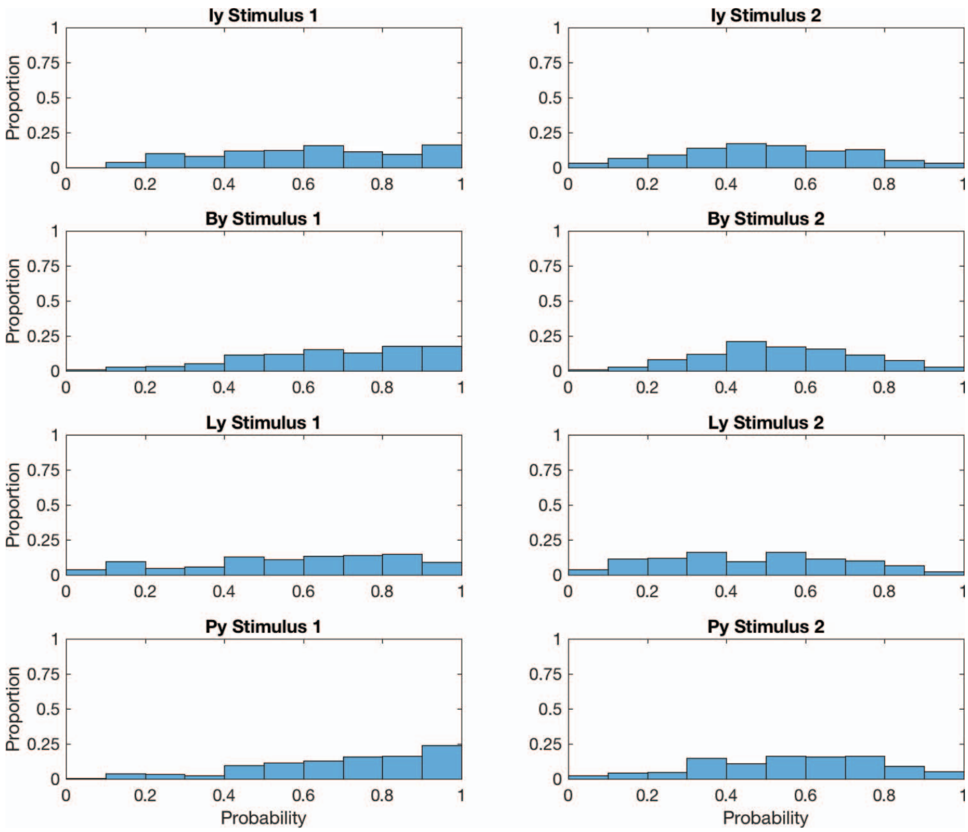

Figure 5. Relative frequency, pooled across participants, for each value of "yes probability" for the Death type of stimulus (left panel) and Harm type of stimulus (right panel). See the online article for the color version of this figure.

transitioning from basis vectors for B to basis vectors for P. Figure 6 presents the relative frequency of $cos(\theta \cdot \pi)^2$, which describes the probability of transiting from a "yes" to one variable to a "yes" to another variable that is incompatible with the former variable. The panel on the left presents the distribution for θ_{IL} and the distribution on the right is for θ_{PB}. As can be seen in Figure 6, the parameter for each pair of attributes is located at a high value on average, indicating that the two attributes are quite similar to each other. It is interesting, however, that the similarity between P and B tends to be higher across all participants than is the case for L and I; in addition, there are larger individual differences for the L and I transitions because the parameter distribution is more widely distributed compared to that for the P and B transitions (see Figure 6).

Summary, Related Theories, and Extensions

Summary of Contribution

In this article, we presented the general theory of measurement context effects based on quantum probability theory. HSM models provide a simple and low dimensional vector space representation of collections of contingency tables formed from measurement of subsets of p variables. HSM models are needed when responses to questions about a variable depend on the context formed by the other variables present in the subset and the order in which they are presented. HSM models provide tools for modeling context effects, and the model parameters provide two psychologically meaningful and useful interpretations of these effects. First, the state vector of an HSM model provides an estimation of the respondents' response tendencies to each of the p variables in a context-free manner, that is, as if a variable was measured in isolation. Second, the measurement operators describe the interrelations between the p measurements, independent of the response tendencies. Furthermore, once the variables being measured have been mapped into the Hilbert space by an HSM model, the parameters of the model can be used to make new predictions for new contexts not included in the original design. For example, if there are three variables and two of them are incompatible, then after estimating the model parameters from an HSM model for a collection $2 - way$ tables, the same model and parameters can be used to make new predictions for new tables that were not included in the original design, such as smaller $1 - way$ tables or larger $3 - way$ tables.

In the past, specific quantum models were built for particular applications. Here we have organized the principles used in the past into a new general program for constructing quantum models. To form this general program, we introduced two general principles: one for building the structure of the Hilbert space and one for building projectors. HSM models provide new contributions to the

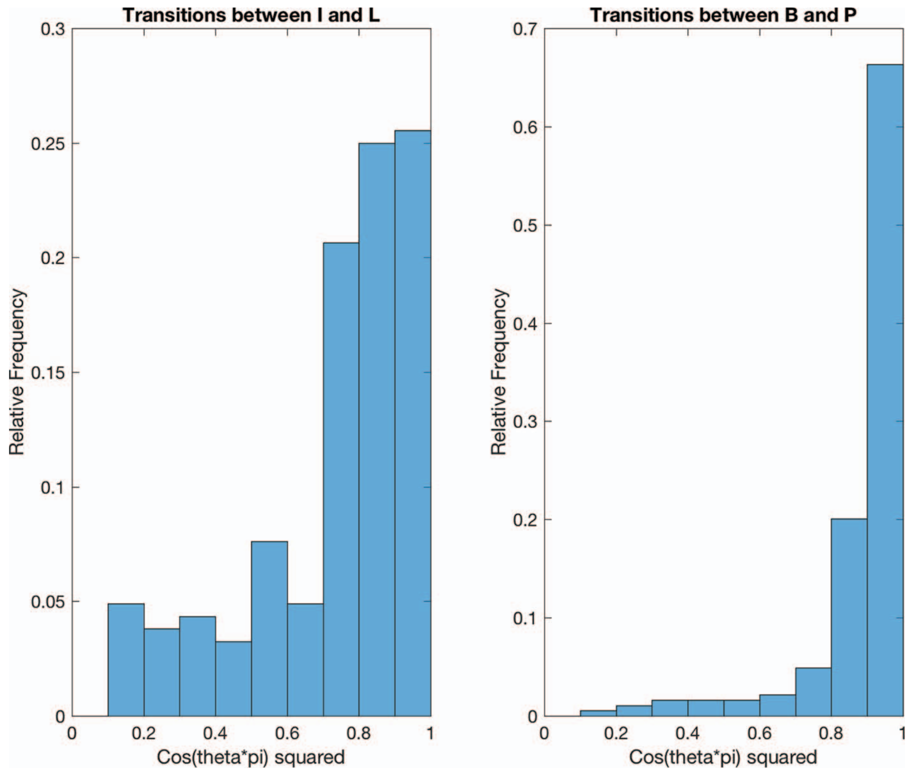

Figure 6. Relative frequencies, pooled across participants, for the transitions from I to L (left panel) and for the transitions from B to P (right panel). See the online article for the color version of this figure.

current set of probabilistic and statistical tools for contingency table analysis. Loglinear–categorical data models apply to only a single table containing all *p* variables, whereas the HSM models can be applied to multiple tables containing different subsets of the *p* variables. Bayesian network models can also be applied to collections of tables; however, they assume the existence of a complete *p – way* joint distribution of the observed variables, and it is often the case that no such *p – way* joint distribution exists. HSM models can be applied to collections of tables even when no *p – way* joint distribution exists to reproduce the collection.

In addition to presenting the general procedures for constructing HSM models, we presented an artificial data example and a real data example. The artificial example was designed to illustrate various kinds of violations of consistency requirements of the *p – way* joint distribution model. The real data example (a) presented the results of a new experiment investigating evaluations of health messages, (b) reported significant deviations from the 4 – *way* joint distribution, and (c) compared the fit of a simple HSM model to a simple Bayesian network model using Bayesian information criteria. We conclude from these analyses that HSM models are empirically viable for modeling collections of contingency tables.

Relation to Social Psychological Theories of Context Effects

Context effects on judgments have been extensively studied by social psychologists in the past (Schwarz & Sudman, 2012). These investigations have led to the development of influential conceptual theories to explain and predict context effects. In general, these theories postulate that although some answers to questions are simply based on retrieval (e.g., What is your political party affiliation?), many other answers have to be constructed by currently available information (e.g., Do you think a new policy will improve health care?). The current information can be affected by context generated from an earlier question (e.g., Is the policy economically sound?), which is carried over and used to answer the next question (Tourangeau et al., 2000). The context of an earlier question can (a) add or subtract information that is used to represent the target (e.g., the new policy) or (b) add or subtract information that is used to represent the standard of comparison (improve health care; Schwarz & Bless, 2007).

These concepts can be used to guide the mathematical formulation of an HSM model. According to quantum theory, answers to questions are constructed from a superposed (indefinite) belief state. In general, when answering a sequence of questions, the state is modified by each projection, so that an earlier question changes the state to form a context that carries over to influence the answers to the later questions. Changes in the target representation correspond to changes in the state ψ used to represent the person's state of beliefs, whereas changes in the standard can be represented by changes in the projectors (basis vectors) used represent the meaning of the answers to a question. For example, a priming event that adds positive expectations for the target policy can increase the amplitudes of the initial state ψ corresponding to a positive evaluation. An earlier question about the economic features of a policy could focus the projector used to answer the next question about health care on unique features other than economics, such as quality of care. The added advantage of formulating a quantum model, with the help of social psychological principles, is that the formal model provides new and more precise quantitative predictions (see, e.g., Trueblood & Busemeyer, 2010; Wang & Busemeyer, 2016a; White et al., 2014; Yearsely & Trueblood, 2017).

Extension to Other Research Settings

Besides those considered here, many other applications of HSM models are possible. For example, past research in consumer behavior has shown that measurements of preferences for different sets of consumer products are context-dependent (Huber, Payne, & Puto, 1982), and HSM models could be used to analyze these context effects. As another example, the HSM models can be useful for analyzing survey data from multiple sources, such as different family members or different cross-cultural groups (De Roover et al., 2012). Dynamic extensions of HSM models can be used to model changes in measurements across longitudinal or multiple stage surveys when different subsets of measurements are used across stages (McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). In sum, HSM models can be applied to complex data collected from a large number of different sources and contexts found in the social and behavioral sciences.

References

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

Aerts, D., Broekaert, J., & Gabora, L. (2011). A case for applying an abstracted quantum formalism to cognition. *New Ideas in Psychology, 29,* 136–146.

Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum-theoretic modeling of human thought. *Topics in Cognitive Science, 5,* 737–773.

Agresti, A., & Katera, M. (2011). *Categorical data analysis*. New York, NY: Springer.

Asano, M., Basieva, I., Khrennikov, A., Ohya, M., & Tanaka, Y. (2017). A quantum-like model of selection behavior. *Journal of Mathematical Psychology, 78,* 2–12.

Asano, M., Hashimoto, T., Khrennikov, A., Ohya, M., & Tanaka, Y. (2014). Violation of contextual generalization of the Leggett-Garg inequality for recognition of ambiguous figures. *Physica Scripta, 89*(T163), 014006.

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

Atmanspacher, H., & Filk, T. (2013). The relevance of Bell-type inequalities for mental systems. In H. Atmanspacher, E. Haven, K. Kitto, & D. Raine (Eds.), *QI 2013 Selected papers of the 7th International Conference on Quantum Interaction* (Vol. 8369, pp. 231–243). Berlin, Germany: Springer-Verlag.

Blutner, R., & beim Graben, P. (2016). Quantum cognition and bounded rationality. *Synthese, 193,* 3239–3291.

Boyer-Kassem, T., Duchêne, S., & Guerci, E. (2016). Testing quantum-like models of judgment for question order effect. *Mathematical Social Sciences, 80,* 33–46.

Bruza, P. D. (2016). Syntax and operational semantics of a probabilistic programming language with scopes. *Journal of Mathematical Psychology, 74,* 46–57.

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

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015). Quantum cognition: A new theoretical approach to psychology. *Trends in Cognitive Sciences, 19,* 383–393.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge, United Kingdom: Cambridge University Press.

Busemeyer, J. R., & Wang, Z. (2017). *Hilbert space multidimensional modeling*. Retrieved from https://arxiv.org/abs/1704.04623v1

Busemeyer, J. R., Wang, Z., & Shiffrin, R. S. (2015). Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. *Decision, 2,* 1–12.

Cervantes, V. H., & Dzhafarov, E. (2018). Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. *Decision*. Advance online publication. http://dx.doi.org/10.1037/dec0000095

Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems and Information Dynamics, 16*(1), 1–17.

Croson, R. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes, 80,* 118–133.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. New York, NY: Cambridge University Press.

Denolf, J., Martínez-Martínez, I., Josephy, H., & Barque-Duran, A. (2017). A quantum-like model for complementarity of preferences and beliefs in dilemma games. *Journal of Mathematical Psychology, 78,* 96–106.

De Roover, K., Ceulemans, E., Timmerman, M. E., Vansteelandt, K., Stouten, J., & Onghena, P. (2012). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological methods, 17,* 100–119.

Dirac, P. A. M. (1958). *The principles of quantum mechanics*. New York, NY: Oxford University Press. (Original work published 1930)

Dzhafarov, E. N., & Kujala, J. V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology, 56,* 54–63.

Dzhafarov, E. N., & Kujala, J. V. (2013). All-possible-couplings approach to measuring probabilistic context. *PloS ONE, 8*(5), e61712.

Dzhafarov, E. N., & Kujala, J. V. (2016). Context-content systems of random variables: The contextuality-by-default theory. *Journal of Mathematical Psychology, 74,* 11–33.

Dzhafarov, E. N., Zhang, R., & Kujala, J. (2016). Is there contextuality in behavioural and social systems? *Philosophical Transactions of the Royal Society A, 374,* 20150099.

Fine, A. (1982). Joint distributions, quantum correlations and commuting observables. *Journal of Mathematical Physics, 23,* 1306–1310.

Gleason, A. M. (1957). Measures on the closed subspaces of a Hilbert space. *Journal of Mathematical Mechanics, 6,* 885–893.

Graben, P. b., & Atmanspacher, H. (2006). Complementarity in classical dynamical systems. *Foundations of Physics, 36,* 291–306.

Gronchi, G., & Strambini, E. (2016). Quantum cognition and Bells inequality: A model for probabilistic judgment bias. *Journal of Mathematical Psychology, 78,* 65–75.

Hameroff, S. R. (2013). Quantum mechanical cognition requires quantum brain biology. *Behavioral and Brain Sciences, 36,* 287–288.

Harrison, D. A., & McLaughlin, M. E. (1993). Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. *Journal of Applied Psychology, 78,* 129–140.

Haven, E., & Khrennikov, A. (2013). *Quantum social science*. Cambridge, United Kingdom: Cambridge University Press.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9,* 90–98.

Khrennikov, A. Y. (2010). *Ubiquitous quantum structure: From psychology to finance*. Berlin, Germany: Springer-Verlag.

Khrennikov, A. Y., & Haven, E. (2009). Quantum mechanics and violations of the sure thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology, 53,* 378–388.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York, NY: Chelsea. (Original work published 1933)

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 10645–10650.

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

Leggett, A. J., & Garg, A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters, 54,* 857–860.

Martínez-Martínez, I., & Sánchez-Burillo, E. (2016). Quantum stochastic walks on networks for decision-making. *Scientific Reports, 6,* 23812.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological methods, 14,* 126–149.

Moore, D. W. (2002). Measuring new types of question-order effects. *Public Opinion Quarterly, 66,* 80–91.

Peres, A. (1998). *Quantum theory: Concepts and methods*. Dordrecht, the Netherlands: Kluwer Academic.

Popper, R., Rosenstock, W., Schraidt, M., & Kroll, B. (2004). The effect of attribute questions on overall liking ratings. *Food Quality and Preference, 15,* 853–858.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of "rational" decision making. *Proceedings of the Royal Society: B, Biological Sciences, 276,* 2171–2178.

Pothos, E. M., & Busemeyer, J. R. (2012). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences, 36,* 255–274.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review, 120,* 679–696.

Pothos, E. M., Yearsley, J., Shiffrin, R. M., & Busemeyer, J. R. (2017). The rational status of quantum cognition. *Journal of Experimental Psychology: General, 146,* 968–987.

Pouta, E. (2004). Attitude and belief questions as a source of context effect in a contingent valuation survey. *Journal of Economic Psychology, 25,* 229–242.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and content*. New York, NY: Academic Press.

Schwarz, N., & Bless, H. (2007). Mental construal processes: The inclusion/exclusion model. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (pp. 119–142). Philadelphia, PA: Psychology Press.

Schwarz, N., & Sudman, S. (2012). *Context effects in social and psychological research*. New York, NY: Springer Science & Business Media.

Steinberg, L. (2001). The consequences of pairing questions: Context effects in personality measurement. *Journal of Personality and Social Psychology, 81,* 332–342.

Stewart, T. C., & Eliasmith, C. (2013). Realistic neurons can compute the operations needed by quantum probability theory and other vector symbolic structures. *Behavioral and Brain Sciences, 36,* 307–308.

Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations probable? *Synthese, 48,* 191–199.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Trueblood, J. S., & Busemeyer, J. R. (2010). A quantum probability account for order effects on inference. *Cognitive Science, 35,* 1518–1552.

van Rijsbergen, C. (2004). *The geometry of information retrieval*. New York, NY: Cambridge University Press.

Von Neumann, J. (1955). *Mathematical foundations of quantum theory*. Princeton, NJ: Princeton University Press. (Original work published 1932)

Wang, Z., & Busemeyer, J. (2016a). Comparing quantum versus Markov random walk models of judgements measured by rating scales. *Philisophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 374,* 20150098.

Wang, Z., & Busemeyer, J. R. (2016b). Interference effects of categorization on decision making. *Cognition, 150,* 133–149.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5,* 672–688.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 9431–9436.

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition, 1,* 48–64.

Yearsely, J. M., & Trueblood, J. (2017). A quantum theory account of order effects and conjunction fallacies in political judgments. *Psychonomic Bulleting & Review*. Advance online publication. http://dx.doi.org/10.3758/s13423-017-1371-z

Appendix A

Derivation of the Clauser, Horne, Shimony, and Holt (CHSH) Inequality

This proof for the CHSH inequality was presented in Atmanspacher and Filk (2013). If we assume that a 4 – *way* joint distribution of four binary random variables A, H, I, and U can reproduce Table 1, then we assume that we can represent the distribution by the following table. First, we define the values w, x, y, and z of the four binary random variables A, H, I, and U, respectively, as -1 or 1. Then we define a new random variable $X = A \cdot I + H \cdot I + H \cdot U - A \cdot U$ (there are three other ways to permute this arrangement to get different inequalities that also satisfy the CHSH criterion). For example, if $A = 1$, $H = 1$, $I = 1$, and $U = 1$, then $X = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 - 1 \cdot 1 = 2$. The values assigned to X for each possible event are shown in Table A1.

The expected value of the random variable X can be written as follows:

$$E[X] = E(A \cdot I) + E(H \cdot I) + E(H \cdot U) - E(A \cdot U)$$
$$= \sum p(A = w \cap H = x \cap I = y \cap U = z)$$
$$\cdot (w \cdot y + x \cdot y + x \cdot z - w \cdot z),$$

where the sum extends across all 16 combinations of values of the four random variables. Note that the values of X ranges from-2 to $+2$, and the expectation is a convex combination of these values, and so the expectation of these values must lie between -2 and $+2$.

Table A1
Joint Probability Distribution Over 16 Combination of Events From 4 Binary Valued Variables, and the Values of 2 Different Random Variables Assigned to the 16 Events

A	H	I	U	X
−1	−1	−1	−1	2
−1	−1	−1	1	2
−1	−1	1	−1	−2
−1	−1	1	1	−2
−1	1	−1	−1	2
−1	1	−1	1	−2
−1	1	1	−1	2
−1	1	1	1	−2
1	−1	−1	−1	−2
1	−1	−1	1	2
1	−1	1	−1	−2
1	−1	1	1	2
1	1	−1	−1	−2
1	1	−1	1	−2
1	1	1	−1	2
1	1	1	1	2

Appendix B

General Method for Building a Unitary Matrix

Suppose H is an $N \times N$ Hermitian matrix. Then we can decompose H into its orthonormal eigenvector matrix V and its real eigenvalue diagonal matrix Λ as follows: $H = V \cdot \Lambda \cdot V^{\dagger}$. The matrix exponential of H is defined as

$$exp(H) = V \cdot exp(\Lambda) \cdot V^{\dagger},$$
$$exp(\Lambda) = diag\left[e^{\lambda_1} \ldots e^{\lambda_j} \ldots e^{\lambda_N}\right].$$

Any unitary matrix can be built from a matrix exponential of a Hermitian matrix as follows:

$$U = exp(-i \cdot H) = V \cdot exp(-i \cdot \Lambda) \cdot V^{\dagger}.$$

In general, the Hermitian matrix has N diagonal entries that are real and $N \cdot (N - 1)/2$ off diagonal entries that can be complex. However, adding a constant to all the diagonal entries has no effect on the choice probabilities, and so one diagonal entry can be set to a fixed value. In sum, only $(N^2 - 1)$ parameters are estimated for each Hermitian matrix.

(*Appendices continue*)

Appendix C

Kronecker Product

Suppose P is an $m \times n$ matrix and Q is an $r \times s$ matrix. Then the Kronecker product is an $(m \cdot r) \times (n \cdot s)$ matrix defined by

$$P \otimes Q = \begin{bmatrix} p_{11} \cdot Q & & \vdots & & p_{1n} \cdot Q \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & p_{ij} \cdot Q & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ p_{m1} \cdot Q & & \vdots & & p_{mn} \cdot Q \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 2 & 3 & 4 \\ 3 & 6 & -2 \\ 4 & -2 & 5 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 3 & 0 & 4 & 0 \\ 0 & 2 & 0 & 3 & 0 & 4 \\ 3 & 0 & 6 & 0 & -2 & 0 \\ 0 & 3 & 0 & 6 & 0 & -2 \\ 4 & 0 & -2 & 0 & 5 & 0 \\ 0 & 4 & 0 & -2 & 0 & 5 \end{bmatrix}.$$

The Kronecker product satisfies the following property (assuming the column dimension of P matches the row dimension of U, and likewise for Q and T):

$$(P \otimes Q) \cdot (U \otimes T) = (P \cdot U) \otimes (Q \cdot T).$$

Appendix D

Conditional Probabilities for the Model Used in Application

Here we derive the quantum predictions for the conditional probability $p(P = w \,|\, B = x)$ for the model used in the Additional Tests of the HSM Model section. Consider the case with $x = 1$, $w = -1$:

$$p(B = x) = \|(M_x \otimes I) \cdot \psi\|^2 = |\psi_{1,1}|^2 + |\psi_{1,-1}|^2.$$

$$\begin{aligned} p(B = x, P &= w) \\ &= \|((U_{PB} \cdot M_w \cdot U_{PB}^{\dagger}) \otimes I) \cdot (M_x \otimes I) \cdot \psi\|^2 \\ &= \|((U_{PB} \cdot M_w \cdot U_{PB}^{\dagger} \cdot M_x) \otimes I) \cdot \psi\|^2 \\ &= \left\| \left(\begin{bmatrix} u_{1,-1} \cdot u_{-1,1} & 0 \\ u_{-1,-1} \,.\, u_{-1,1} & 0 \end{bmatrix} \otimes I \right) \cdot \psi \right\|^2 \\ &= \left\| u_{-1,1} \cdot \begin{bmatrix} u_{1,-1} \\ u_{-1,-1} \end{bmatrix} \otimes \begin{bmatrix} \psi_{1,1} \\ \psi_{1,-1} \end{bmatrix} \right\|^2 \\ &= |u_{-1.1}|^2 \cdot ((|u_{1,-1}|^2 + |u_{-1,-1}|^2) \cdot |\psi_{1,1}|^2 + (|u_{1,-1}|^2 + |u_{-1,-1}|^2) \cdot |\psi_{1,-1}|^2) \\ &= |u_{-1,1}|^2 \cdot (|\psi_{1,1}|^2 + \cdot |\psi_{1,-1}|^2). \end{aligned}$$

(*Appendices continue*)

Therefore,

$$p(P = w \mid B = x) = \frac{p(B = x, P = w)}{p(B = x)} = |u_{-1,1}|^2$$

The fact that $U_{PB}^{\dagger}U_{PB} = I$ implies the equality $|u_{-1,1}|^2 = |u_{1,-1}|^2$. But using the same argument just explained, we obtain $p(B = x \mid P = w) = |u_{1,-1}|^2$, and therefore $p(B = x \mid P = w) = p(P = w \mid B = x)$. The same argument leads to $p(I = y \mid L = z) = p(L = z \mid I = y)$. This is called the law of reciprocity, but this law holds for only one-dimensional projectors like those used in the application.

Received May 5, 2017
Revision received February 3, 2018
Accepted February 18, 2018 ■