Hilbert Space Multi-dimensional Modeling

Jerome R Busemeyer Indiana University Zheng Wang Ohio State University

April 18, 2017

Abstract

This article presents general procedures for constructing, estimating, and testing Hilbert space multi-dimensional (HSM) models, which are based on quantum probability theory. HSM models can be applied to collections of K different contingency tables obtained from a set of p variables that are measured under different contexts. A context is defined by the measurement of a subset of the p variables that are used to form a table. HSM models provide a representation of the collection of K tables in a low dimensional vector space, even when no single joint probability distribution across the p variables exists. HSM models produce parameter estimates that provide a simple and informative interpretation of the complex collection of tables. Comparisons of HSM model fits with Bayes net model fits are reported for a new large experiment, demonstrating the viability of this new model. We conclude that the model is broadly applicable to social and behavioral science data sets.

When large data sets are collected from different contexts, often they can 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. It may not be possible to measure all p variables at once, and instead, only a subset of variables $(Y_{k_1}, \ldots, Y_{k_s})$, s < p, can be measured at once. Each subset 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.

For example, a research problem could involve three variables (Y_1, Y_2, Y_3) , but some tables might include only some subset of the three variables. One context might involve the measurement of a single variable Y_1 that has 5 values to form a 1-way frequency table T_1 composed of 5 frequencies. Another context could be used to form another 5×3 table T_2 , composed of joint frequencies for two variables (Y_1, Y_2) . A third context could form another 3×4 table T_3 containing variables (Y_2, Y_3) , and fourth could form a 5×4 table containing variables (Y_1, Y_3) .

A critical problem arises: How to integrate and synthesize these K different tables into a compressed, coherent, and interpretable representation? It is

common to apply categorical data analysis [5] to a single p-way table (e.g., a single $5\times3\times4$ table). However, the problem is different here because there are a collection of K tables of varying dimensions rather than a single p-dimensional table.

A common solution is to assume that the K tables are generated from a single latent p-way joint distribution, and then try to reproduce the frequencies in the K different tables by marginalizing across variables in the p-way table. Often Bayesian causal networks are used to reduce the number of latent probability parameters by imposing conditional independence assumptions [16, 6]. Unfortunately, however, in many cases, no such p-way joint distribution exists that can reproduce the observed tables! This occurs when the data tables violate consistency constraints required by classical (Kolmogorov) probability theory upon which Bayes nets are built; in this case, no Bayesian network representation composed of the p-variables can even be formed. For example, the tables may have inconsistent marginal probabilities for a variable Y_1 , or the probabilities assigned to two sequential measures (Y_1, Y_2) may not be commutative. In the following sections, we give concrete examples of the various types of possible joint probability violations.

Hilbert space multidimensional (hereafter, denoted HSM) modeling is based on quantum probability theory [27, 38, 45]. It provides a promising new solution to these problems faced by complex data by constructing a model that has (a) a single finite state vector that lies within a low dimensional vector space, and (b) by forming a set of measurement operators that represent the p measurements. In this way, we can achieve a compressed, coherent, and interpretable representation of the p variables that generate the complex collection of K tables, even when no standard p-way joint distribution exists. In a Hilbert space model, the state vector represents respondents' initial tendencies to select responses to each of the p measurements; the measurement operators describe the inter-relations between the p measurements (independent of the initial state of the respondents). ¹

HSM models are similar to traditional multidimensional scaling (MDS) models [57], but also different from them in important aspects. Like traditional MDS models, HSM models are based on similarity relations between entities located within a vector space. However, traditional MDS models define the similarity relations by inner products between vectors in a real vector space, whereas HSM models define similarity relations by projections onto subspaces of a complex vector space. Also, MDS models are designed to account for a single 2 - way similarity matrix, whereas HSM models can be applied to multiple similarity matrices (e.g., when the similarity relation is asymmetric, see Pothos et al. 42).

The article is organized as follows. First, we briefly justify our extension and application of quantum probability theory to study social and behavioral sciences. Second, we provide an artificial data example that illustrates how consistency requirements of a single p - way joint distribution can be violated.

¹Technically, a Hilbert space is a complete inner product vector space defined on a complex field. Our vector spaces are finite, and so they are always complete.

Third, we describe the general procedures for building HSM models. Fourth, we provide a concrete application to the artificial data. Fifth, we present an application of the principles to real data obtained from evaluations of public service announcements. Finally, we finish with a summary of the new contributions made by HSM models.

1 Why Apply Quantum Theory to Social and Behavioral Sciences?

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 [32] provided one. Much of the theory was initially motivated by problems arising in physics, and later applications appeared in economics, engineering, insurance, statistics, etc. 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 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 the 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 1920's 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. Interestingly, 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 [19] and Von Neumann [49]. 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 that events are always commutative and distributive, and the closure property does not always hold.

It turns out that quantum probability theory is not only useful for explaining physical phenomena, but it also provides useful new tools to model human behavior [41]. Notice we are not proposing that the brain is some kind of quantum computer, and instead, we are only using the mathematical principles of quantum theory to account for human behavior. We have found many cases where quantum probability theory provides a better account of human judgment and decisions than classical probability theory [10]. In particular, human judgments are not commutative, and order effects are pervasive [55, 47]; human decisions often violate the law of total probability that follows from the distributive axiom [40, 52, 33]; quantum theory provides a coherent account of many different types of probability judgment errors [14, 1] as well as violations of rational decision making [34, 58, 8]. Quantum theory provides a natural account of asymmetry in similarity judgments Pothos et al. [42]. We could go on with more examples, but a review of this rapidly growing literature is beyond the purpose of this article.

The principles from quantum theory actually resonate with deeply rooted psychological conceptions [3]. 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 [35]. 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 [54]. In quantum physics, order-dependent measurements are said to be non-commutative and quantum theory was especially designed for these types of non-commutative measures. Finally, consider the unique quantum concept of entanglement—the event $A \cap B$ may be observed, and another event $C \cap D$ may be observed, but the event $A \cap B \cap C \cap D$ may not even exist, violating closure. Quantum probabilities based on an entangled state provides a basis for explaining these types of non-classical systems [11, 2].²

2 An Artificial Example

In this section we present an artificial example that serves to illustrate several ways that a joint probability model can fail. Suppose that four variables are measured, labeled A,H,I, and U. For example, suppose queries are made from a large social media source on political candidates concerning Attractiveness,

²Reviewers often argue that although the micro world is quantum, the macro world that we observe is classical, and so why would nature evolve a non-commutative 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 physical models of an unobserved (ontological) macro world. Even classical physical models of the world can produce observed probabilities that are non-commutative. The latter can happen when only coarse epistemic measurements of the underlying ontic physical states are available [25].

Table 1: Eight different 2×2 tables produced by yes, no answers to pairs of attributes A,H,I,U. The label YN refers to yes to the first and no to the second attribute. Each cell within a row is a relative frequency, and all the cells within a row sum to one. The order of questions may matter so that the HA table (H asked before A) may differ from the AH table (A asked before H).

Pair	YY	YN	ΝŶ	NN
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

Honesty, Intelligence, and Unusualness. As another example, suppose queries are made from a large medical record source on patient symptoms concerning Anxiety, Hyperactivity, Irritation, and Unruliness. As a third example, suppose queries are made from a large consumer choice source about food products concerning whether the product is Appetizing, Healthy, Interesting, and Unfamiliar. It is difficult or impossible to obtain ratings from individuals on all four attributes simultaneously. Suppose that only pairs of attributes are queried at a time, for example, the pair A,I and the pair A,H. Each pair provides a context for answering the questions.

For simplicity, suppose each query is answered with a yes (Y) or no (N) answer. Thus a pair of yes-no answers to a pair of attributes forms one 2×2 table with relative frequencies for pairs of answers YY,YN,NY,NN. Suppose 8 contexts are used to form 8 different 2×2 tables as shown in Table 1. For example, the pair of attributes AI form the context for the second 2×2 table. Each cell within a row is a relative frequency for a 2×2 table, and the cells within a row sum to one. For example, the relative frequency of yes to attribute A and no to attribute I equals .175. Note that question ordering may matter so that, for example, the context AH is treated different from HA. For simplicity, we only included a subset of all 12 possible 2-way tables. These 8 tables are sufficient to make our points.

2.1 Does a joint distribution exist?

The following question can be asked about Table 1: Does a single 4-way joint probability distribution exist that can reproduce Table 1? The 4-way joint probability distribution is defined by 4 binary random variables (A, H, I, U) that generate 16 latent joint probabilities that sum to one: $\pi(A = w \cap H = x \cap I = y \cap U = z)$, where, for example, A is a random variable with values w = 1 for yes and w = 0 for no, and similar definitions hold for the other three random

variables. For example, the relative frequency, p(YN|AI), of YN in the context of the pair AI is predicted by the marginal $\pi(A=1,I=0)=\sum_x\sum_z\pi(A=1\cap H=x\cap I=0\cap U=z)$. Note that this 4-way joint distribution is completely general (non-parametric), because no conditional independence or parametric distribution assumptions are imposed.

The answer to the above question is negative: There is no single 4-way joint distribution that can reproduce Table 1. First of all, the 4-way distribution requires the marginal distribution of a single random variable to be invariant across contexts. This requirement fails. For example, the marginal probability of yes to random variable I is not invariant: p(YY|IU) + p(YN|IU) = .355 which differs from p(YY|UI) + p(NY|UI) = .420. 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 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, HU. Suppose we restrict our question to only these four tables, can a 4-way joint distribution reproduce these 4 tables? Surprisingly, the answer is still negative. These 4 tables violate a consistency requirement of a single p-way joint distribution, called the Clauser, Horne, Shimony, and Holt (CHSH) inequality [for applications in psychology, see 11, 20].³ The CHSH inequality implies the following restriction on the joint probabilities required by the 4-way joint probability model: -1 < CHSH < 2, where

$$CHSH = E(A \cdot I) + E(H \cdot I) + E(H \cdot U) - E(A \cdot U), \tag{1}$$

and, for example, $E(A \cdot I) = \pi(A = 1 \cap I = 1)$ is the expectation of the product of the two random variables, A, I. If we set $\pi(A = 1 \cap I = 1) = p(YY|AI)$, $\pi(H = 1 \cap I = 1) = p(YY|HI)$, $\pi(H = 1 \cap I = 1) = p(YY|HI)$, $\pi(A = 1 \cap I = 1) = p(YY|AI)$, from Table 1, then the CHSH value computed from Table 1 equals CHSH = 2.25, which exceeds the bound required by the 4-way joint probability model. 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 inequality applies to 3-way joint distributions [44, 36]. Dzhafarov and Kujala [20] derive and provide a general summary of all these linear constraints required for a single joint distribution to reproduce a collection of contingency tables.

 $^{^3}$ The CHSH inequality is closely related to the Bell inequality, and the latter was derived for the Bohm experiment using a pair of entangled spin $\frac{1}{2}$ photons, which was used to test the famous Einstein Podolsky Rosen (EPR) paradox. There are different ways to derive the CHSH inequality, and we follow the derivation by Fine [23].

2.2 Non-parametric statistical tests of the joint distribution model

Suppose the data in Table 1 are based on a sample of N=100 independent observations for each 2×2 table. Then it is unclear whether the violations of the 4-way joint probability distribution, described above, are statistically significant. To address this issue, we propose the following general method: We compare the 4 - way joint probability model to a saturated model.⁴ The saturated model simply assumes that we have 8 independent 2-way tables, and each table has 4 probabilities that sum to one. The 4 - way joint probability model has 15 free parameters, because the 16 joint probabilities are constrained to sum to one. The saturated model has $8 \times 3 = 24$ parameters, because the probabilities sum to one 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 $G^2 = -2 \times log likelihood$ can be determined for each model. Then a likelihood ratio (i.e., chi-square difference) test can be used to compare models. Using this method with N=100 observations per table produces a chi-square difference equal to $G_{diff}^2 = G_{joint}^2 - G_{sat}^2 = 18.04$, which is a statistically significant difference with p = .031. Therefore, using this classical statistical test, the joint probability model is rejected. Note that this is a non-parametric test that requires no conditional independence or parametric distribution assumptions.

The above non-parametric 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 the joint distribution model. For example, if only the four 2×2 tables (AI,AU,HI,HU) are included in the design, then the saturated model has only $4 \times 3 = 12$ parameters, which is fewer than the 4-way joint distribution model (see, e.g., Bruza et al. 11). However, if four 1-way tables, produced by measuring each attribute alone, are included into the design to form a collection of 8 tables (A,H,I,U, AI,AU,HI, HU), then the saturated model has 16 parameters, which leaves df=1 for testing the joint probability model. It is worth nothing that the number of parameters in the joint probability model grows exponentially with the number of random variables, which makes it necessary to impose restrictions (e.g., using Bayesian networks or parametric distributional assumptions) to form testable models.

Including 1-way tables into the design provides direct tests of the marginal invariance assumption of the joint distribution model. For example, suppose attributes A,B are measured by binary choices, and the design included the three tables (A,B,AB). This simple design provides $2\ df's$ for testing the joint distribution model (see, e.g., Wang and Busemeyer 52). Alternatively, including different orders of presentation provides direct tests of commutativity. For ex-

 $^{^4\}mathrm{This}$ is what Dzhafarov and Kujala 21 call the all possible couplings model.

⁵Nevertheless, $G_{diff}^2 = 2.56$ after fitting the 4 - way model to Table 1, which reflects the violation of the CHSH inequality.

ample, if attributes A,B are measured on 9-point scales, and the design includes both 2 - way tables (AB,BA), then df = 80 for testing the joint probability model (see, e.g., Wang and Busemeyer 53).

On the one hand, an advantage of this non-parametric statistical test of the joint distribution model is that it tests all of the constraints imposed by the joint distribution model (including marginal invariance, absence of order effects, CHSH inequality, 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 versus a test of marginal invariance), but these additional tests are not described in detail here.

2.3 Previous research testing a joint distribution with multiple tables

The commutative property has been tested by using the pair of tables (AB,BA) that vary question order. It has long been known that question order effects commonly occur with human judgments [46, 43]. Recently, quantum models have provided good accounts for these effects [51, 55, 53]. The marginal invariance assumption has been tested by using a design with two tables (A,BA). These are also called tests of the law of total probability, or tests for interference effects. Several experiments have been conducted that demonstrate violations of marginal invariance [13, 18, 15, 33, 50, 52]. A number of experiments have been conducted to test the CHSH or similar inequalities required by a single joint distribution applied to a collection of several 2×2 tables [11, 4, 7, 26]. Although violations of the required inequalities were reported in the experiments testing the CHSH inequality, they were confounded with violations of marginal invariance [22]. It remains to be found out whether or not human judgments produce violations of the CHSH inequality in the absence of violations of order effects and marginal invariance [22].

3 Multidimensional Hilbert Space Modeling

3.1 Basics of quantum probability theory

It is helpful to introduce quantum probability theory by comparing it with 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 variables $(Y_i, i = 1, \dots, p)$ 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

⁶See Busemeyer and Bruza [12], Haven and Khrennikov [29], Khrennikov [31], van Rijsbergen [48] for introductions to quantum probability theory written for social and behavioral sciences.

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 $(Y_1 = y_i)$.

Classical theory begins with a universal set Ω containing all events, which is called the sample space; and quantum theory replaces this with a vector space \mathscr{H} containing all events, which is called the Hilbert space. Classical theory defines an event A as a *subset* of the sample space, whereas quantum theory defines an event A 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. The change from subsets to subspaces is where the logic of events differs between the two theories.

Classical theory assumes closure: If $A \in \Omega$ is an event, and $B \in \Omega$ 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 $A \in \mathcal{H}$, $B \in \mathcal{H}$ 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 events.

In quantum theory, a sequence of events, such as A and then B, denoted AB, is represented by the sequence of projectors P_BP_A . If the projectors commute, $P_AP_B = P_BP_A$, then the product of the two projectors is a projector corresponding to the subspace $A \cap B$, that is, $P_BP_A = P(A \cap B)$; and the events A and B are said to be *compatible*. However, if the two projectors do not commute, $P_BP_A \neq P_AP_B$, then neither their product is a projector, and the events are *incompatible*.

Classical theory defines a set function p that assigns probabilities to events, which is required to be an additive measure: $p(A) \ge 0$, $p(\Omega) = 1$, 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:⁷

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

Quantum probabilities also satisfy an additive measure: $p(A) \ge 0$, $p(\mathcal{H}) = 1$, and if $P_A P_B = 0$, then $p(A \lor B) = p(A) + p(B)$. In fact, Equation 2 is the unique way to assign probabilities to subspaces that form an additive measure for dimensions greater than 2 [24].

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

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

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

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

⁷A more general approach uses what is called a density operator rather than a pure state vector, but to keep ideas simple, we use the latter.

and so the probability of the sequence AB equals $p(AB) = p(A) \cdot p(B|A) = \|P_B P_A |\psi\rangle\|^2$. 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 $||P_C(P_BP_A)|\psi\rangle||^2$ for quantum theory.

3.2 Building Projectors

This section describes a general way to construct the projectors for events in the Hilbert space, and to formally describe the conditions that produce incompatibility. This section is somewhat abstract and technical, and a concrete application is provided in the next section where we build a simple model for Table 1. In the following, $|V\rangle$ denotes a vector in the Hilbert space, $\langle V|W\rangle$ denotes an inner product, $|V\rangle\langle V|$ denotes an outer product, and P^{\dagger} denotes a Hermitian transpose.

In general, a projector, denoted P, operating in an N-dimensional Hilbert space \mathscr{H} , is defined by the two properties $P=P^\dagger=P^2$. By the first property, P is Hermitian, and so it can be decomposed into N orthonormal eigenvectors; by the second property, P has only two eigenvalues, which are simply (0,1) [28]. Define $|V_j\rangle$, $j=1,\cdots,N$ as the set of N orthonormal eigenvectors of P. The projector P can be expressed in terms of the eigenvectors as follows

$$P = \sum_{j} \lambda_{j} |V_{j}\rangle \langle V_{j}|, \qquad (3)$$

where the outer product, $|V_j\rangle \langle V_j|$, is the projector that projects into the ray spanned by eigenvector $|V_j\rangle$, and $\lambda_j=1$ if $|V_j\rangle$ corresponds to an eigenvalue of 1, and $\lambda_j=0$ if $|V_j\rangle$ corresponds to an eigenvalue of 0. These N eigenvectors form an orthonormal basis that spans the Hilbert space. Every vector, such as $|\psi\rangle \in \mathcal{H}$ can be expressed as a linear combination of these basis (eigen) vectors

$$|\phi\rangle = \sum_{j}^{N} \phi_{j} \cdot |V_{j}\rangle \tag{4}$$

If two projectors, P_A , P_B share all of the same eigenvectors, then they commute [28]. In other words, two events A, B are compatible if they are described in terms of the same basis. If the two projectors do not share all of the same eigenvectors, then they do not commute, and the events A, B are described by two different bases. They are incompatible, and must be evaluated sequentially, because one needs to change from one basis to evaluate the first event, to another basis to evaluate the second event, making them incompatible.

Define $|V_j\rangle$, $j=1,\dots,N$ as the basis used to describe event A, and define $|W_j\rangle$, $j=1,\dots,N$ as the basis used to describe event B. We can change from one basis to another by a unitary transformation (a "rotation" in Hilbert space)

$$|W_i\rangle = U|V_i\rangle, j = 1, \cdots, N,$$
 (5)

where U is defined by $U^{\dagger}U = I$, that is, U is an isometric transformation that preserves inner products [28]. Therefore, the projector for event B can be reexpressed in terms of the event A basis $|V_i\rangle$, $j = 1, \dots, N$ as follows

$$P_{B} = \sum_{j}^{N} \lambda_{j} |W_{j}\rangle \langle W_{j}|$$

$$= U\left(\sum_{j} \lambda_{j} |V_{j}\rangle \langle V_{j}|\right) U^{\dagger}.$$
(6)

According to Equation 5, the unitary transformation U represents the transitions from state $|W_i\rangle$ to state $|V_j\rangle$ by the inner product $\langle V_j|W_i\rangle$.

So far, we have presented a general method for building the projectors by defining a basis for the vector space and by transforming from one basis to another using unitary transformation. Then the next question is how to build the unitary transformation? In general, any unitary transformation can be built from a Hermitian operator H as follows (Halmos, 1993):

$$U = exp(-i \cdot H). \tag{7}$$

The right hand side is exponential function of the Hermitian operator H (see appendix for details).

In summary, the HSM program selects a Hermitian operator H for Equation 7, and then uses the Hermitian operator to build the unitary operator U which provides the relation between projectors P_A and P_B for incompatible events. The beauty of using a vector space is that it provides an infinite number of ways to generate incompatible variables by unitary "rotation," and yet remain within the same N-dimensional space. This is how an HSM model maintains a low dimensional representation even when there are a large number of variables.

3.3 Building the Hilbert space

This section describes how we construct a Hilbert space to represent the p variables. This construction depends on the compatibility relations between the variables. For this section, we need to use the Kronecker (tensor) product between two matrices, denoted as $P \otimes Q$ (see the Appendix for a brief review).

To begin building the Hilbert space, suppose we measure a single variable, say Y_1 , that can produce n_1 values corresponding to the mutually exclusive and exhaustive set of events $(Y_1 = y_i), i = 1, \dots, n_1$. To represent these events in a Hilbert space, we partition the space into n_1 orthogonal subspaces. Each subspace, such as $(Y_1 = y_i)$, corresponds to a projector $P(Y_1 = y_i)$. The projectors for all of the events are pairwise orthogonal, $P(Y_1 = y_i)P(Y_1 = y_j) = \mathbf{0}$, and

complete, $\sum_i P(Y_1 = y_i) = \mathbf{I}$ (where \mathbf{I} is the identity that projects onto the entire Hilbert space). These n_1 events are all compatible, and the projectors are all commutative, because they are all orthogonal to each other. Each projector generates $N_1 \geq n_1$ eigenvectors, and the projectors all share the same eigenvectors, but with different eigenvalues. These N_1 eigenvectors provide the basis for spanning a N_1 -dimensional Hilbert space, \mathscr{H}_{N_1} .

Continuing with the case of a single variable represented by the Hilbert space \mathscr{H}_{N_1} , we can express each vector $|\phi\rangle\in\mathscr{H}_{N_1}$ in terms of its coordinates with respect to the eigenvectors of $P(Y_1=y_i)$ by using Equation 4. Using this basis, the coordinate representation of each projector, say $P(Y_1=y_i)$ is simply an $N_1\times N_1$ diagonal matrix, $M_1(i)$ with ones located in the rows corresponding to basis vectors that have an eigenvalue of one associated with the projector $P(Y_1=y_i)$, and zeros otherwise. The coordinate representation of $|\psi\rangle$ with respect to this basis is a $N_1\times 1$ column matrix ψ with coordinate ψ_i in row i, which satisfies $\psi^\dagger\psi=1$. Then the probability distribution over the values of Y_1 for $i=1,\cdots,n_1$ is given by

$$||P(Y_1 = y_i) \cdot |\psi\rangle||^2 = ||M_1(i) \cdot \psi||^2 = |\psi_i|^2.$$
 (8)

There is little difference between classical and quantum probability at this point. Next suppose we measure two variables, Y_1 with n_1 values and Y_2 with n_2 values, with $n_1 \geq n_2$. If these two variables are compatible, then the joint event $(Y_1 = y_i \cap Y_2 = y_i)$ is well defined for all pairs of values. Therefore the Hilbert space is partitioned into $n_1 \cdot n_2$ orthogonal subspaces. Each subspace corresponds to a projector $P(Y_2 = y_j)P(Y_1 = y_i) = P(Y_1 = y_i)P(Y_2 = y_j) = P(Y_1 = y_i)P(Y_2 = y_j)$ $y_i \cap Y_2 = y_i$). These projectors are pairwise orthogonal and complete, and every pair of projectors is commutative. Each projector shares $(N_1 \cdot N_2) \geq (n_1 \cdot n_2)$ eigenvectors, but with different eigenvalues, to span a Hilbert space $\mathscr{H}_{N_1 \cdot N_2}$. Using this basis, the projector $P(Y_1 = y_i)$ is represented by the Kronecker product $M_1(i) \otimes I_{N_2}$, where I_{N_2} is an $N_2 \times N_2$ identity matrix. The projector $P(Y_2 = y_i)$ is represented by the matrix Kronecker product $I_{N_1} \otimes M_2(j)$. Then $P(Y_2 = y_j)P(Y_1 = y_i) = P(Y_1 = y_i \cap Y_2 = y_j)$ is represented by the product $(M_1(i) \otimes I_{N_2}) \cdot (I_{N_1} \otimes M_2(j)) = M_1(i) \otimes M_2(j)$, which is simply a diagonal matrix with ones located in the rows corresponding to $(Y_1 = y_i \cap Y_2 = y_j)$ and zeros otherwise. The coordinate representation of $|\psi\rangle$ with respect to this basis is a $(N_1 \cdot N_2) \times 1$ column matrix, $(\psi, \psi^{\dagger} \psi = 1)$, with coordinate ψ_{ij} in row $n_2 \cdot (i-1) + j$. Then the joint probability for a pair of values equals

$$||P(Y_2 = y_j)P(Y_1 = y_i)|\psi\rangle||^2 = ||M_1(i) \otimes M_2(j) \cdot \psi||^2 = |\psi_{ij}|^2.$$
 (9)

There is still little difference between the classical and quantum theories at this point. Adding variables increases the dimensionality of the space, just like it does with a Bayesian model.

Now suppose that variables Y_1 (with n_1 values) and Y_2 (with $n_2 \leq n_1$ values) are incompatible. In this case, we cannot define the joint occurrence of two events $(Y_1 = y_i \cap Y_2 = y_j)$, and we can only represent a sequence of two single events, e.g., $(Y_1 = y_i)$ and then $(Y_2 = y_j)$ by the sequence of projectors

 $P(Y_2 = y_j)P(Y_1 = y_i)$. As before, we define $P(Y_1 = y_i)$ as the projector for the event $(Y_1 = y_i)$, and likewise, we define $P(Y_2 = y_j)$ as projector for the event $(Y_2 = y_j)$. Both projectors are represented with a Hilbert space, \mathscr{H}_{N_1} , of dimension $N_1 \geq n_1$. We can choose to express each vector $|\phi\rangle \in \mathscr{H}_{N_1}$ in terms of the coordinates with respect to the eigenvectors of $P(Y_1 = y_i)$ by using Equation 4. Using this basis, the coordinate representation of projector $P(Y_1 = y_i)$ is simply an $N_1 \times N_1$ diagonal matrix, $M_1(i)$ with ones located in the rows corresponding to basis vectors that have an eigenvalue of one associated with this projector, and zeros otherwise. Using Equation 5, the projector $P(Y_2 = y_j)$ can be expressed in terms of the $P(Y_1 = y_i)$ basis by a unitary matrix, U. Then the matrix representation of $P(Y_2 = y_j)$ is $(U \cdot M_1(j) \cdot U^{\dagger})$. Finally, the coordinate representation of the state vector $|\psi\rangle$ with respect to the Y_1 basis is a $N_1 \times 1$ column matrix ψ . The probability of the sequence of events $(Y_1 = y_i)$ and then $(Y_2 = y_j)$ equals

$$||P(Y_2 = y_j)P(Y_1 = y_i)|\psi\rangle||^2 = ||(U \cdot M_1(j) \cdot U^{\dagger}) \cdot M_1(i) \cdot \psi||^2.$$
 (10)

This is where a key difference between the classical and quantum theories occurs. Note that, unlike a Bayesian model, adding variable Y_2 does not increase the dimensionality of the space.

Finally suppose that we measure three variables, Y_1 with n_1 values, Y_2 with n_2 values, and Y_3 with n_3 values. Suppose Y_1 is compatible with Y_2 and Y_2 is compatible with Y_3 but Y_1 is incompatible with Y_3 . In this case, we can partition the Hilbert space using projectors $P(Y_1 = y_i \cap Y_2 = y_j), i = 1, \dots, n_1, j = 1, \dots, n_2$, which are pairwise orthogonal and complete, and every pair of these projectors is commutative. Using the eigenvectors of these projectors as the basis, the projector $P(Y_1 = y_i)$ is represented by the Kronecker product $M_1(i) \otimes I_{N_2}$, and the projector $P(Y_2 = y_j)$ is represented by the Kronecker product $I_{N_1} \otimes M_2(j)$. Using a unitary transformation, U, the matrix representation of the projector $P(Y_3 = y_k)$ is given $(U \cdot M_1(k) \cdot U^{\dagger}) \otimes I_{N_2}$. Then, the probability of the two compatible events $(Y_1 = y_i)$ and $(Y_2 = y_j)$ equals

$$||P(Y_2 = y_j)P(Y_1 = y_i)|\psi\rangle||^2 = ||M_1(i) \otimes M_2(j) \cdot \psi||^2.$$
 (11)

the probability of the two compatible events $(Y_2 = y_i)$ and $(Y_3 = y_j)$ equals

$$||P(Y_3 = y_k)P(Y_2 = y_i)|\psi\rangle||^2 = ||(U \cdot M_1(k) \cdot U^{\dagger}) \otimes M_2(i) \cdot \psi||^2,$$
 (12)

and the probability of the sequence of two incompatible events $(Y_1 = y_i)$ and then $(Y_3 = y_k)$ equals

$$||P(Y_3 = y_k)P(Y_1 = y_i)|\psi\rangle||^2 = ||(U \cdot M_1(k) \cdot U^{\dagger} \cdot M_1(i)) \otimes I_{N_2} \cdot \psi||^2.$$
 (13)

The methods described above generalize in a fairly straightforward manner for more variables. Note that when variables are compatible, quantum probability theory works like classical probability theory, and the Hilbert space

dimensionality increases exponentially as the number of compatible variables increases. However, when variables are incompatible, it is unlike classical probability theory, and the Hilbert space dimensionality remains constant as the number of incompatible variables increases. In short, incompatibility—a central concept in quantum theory and its application to psychology [39, 56]—produces simplification by rotating the basis to generate variables rather than adding new dimensions.

3.4 The Hilbert space multi-dimensional program⁸

An HSM model is built using the following programatic steps. All of these steps are illustrated in the next section with a concrete application to the artificial data set.

First, the researcher needs to determine which variables or attributes are commutative, and which are not. Referring to our artificial data set, we need to determine, for example, whether the attributes A and H commute or not. Recall that if they are compatible, then they can be defined simultaneously, and sequence does not matter; but if they are incompatible, they must be evaluated sequentially, because one needs to change from one basis to another basis for evaluating the sequence. One way to determine this is to observe whether or not a pair of variables produce order effects. Alternatively, one can statistically compare competing models with different hypothesized compatibility relations.

Second, the dimension N of the Hilbert space is determined. This depends first of all on the assumed compatibility relations. Given the compatibility relations, an HSM modeling procedure can begin with the lowest possible dimension, and only increase the dimension as required by model comparisons that favor a higher dimension.

Third, a basis is selected for representing the coordinates of the state vector $|\psi\rangle$ in terms of combinations of 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 from the data. In general, each coordinate can be complex, containing a magnitude and a phase. Therefore, if the dimension equals N, then the initial state requires $2\cdot N$ parameters. However, the initial 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.

Fourth, the projectors from unitary transformations are built, and the latter are obtained by selecting a Hermitian operator used in Equation 7. 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

⁸The word program here refers to the set of procedures that we formulated to build HSM 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 2 × 2 tables. The current codes are written in Matlab and they are available at http://mypage.iu.edu/~jbusemey/quantum/HilbertSpaceModelPrograms.htm

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.

Fifth, using quantum probability rules, the probability of a sequence of measurements is computed. Using the predicted probabilities, the model computes the log likelihood of the data given the model. The parameters for the initial state and the Hermitian operators are estimated from the data using maximum likelihood, and the model computes $G^2 = -2 \cdot loglikelihood$ statistics for model comparison.⁹ The number of model parameters is determined by the number of parameters used to build the initial state vector plus the number of parameters used to estimate the Hermitian operators used in Equation 7.

Sixth, the fit of the model returns parameters for the initial 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 transformations 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.

4 Application to the Artificial Data Set

Step 1. Determine compatibility of variables. Psychologically, this step determines whether two variables can be measured simultaneously (compatible) or they have to be measured sequentially (incompatible). Based on the order effects observed in Table 1, we infer that the pair of variables A,H were incompatible, as well as the pair I,U. The design did not include manipulations of order to test compatibility between variables A,I or H,U. In this case, another way to empirically test compatibility is to compare model fits that make compatibility vs. incompatibility assumptions about these variables. Here for the purpose of illustration, we assumed that they were compatible.

Step 2. Define the Hilbert space. Assuming that A,I are compatible means that we can define all of the events obtained from all of the combination of values of these two variables: $(A = w \cap I = y)$, for (w = 0, 1) and (y = 0, 1). Similarly, assuming that H,U are compatible means that we can define all of the events formed by the all of the combination of values of these two variables: $(H = x \cap U = z)$, for (x = 0, 1) and (z = 0, 1). However, we cannot define combinations for more variables because of the incompatibilities. The simplest model is a model that assumes that each event is represented by only one dimension, which

 $^{^9\}mathrm{Currently},$ we use a particle swarm method to estimate parameters in order to avoid local minimum.

produces a total of four dimensions. Therefore, the minimum size of the Hilbert space was set to four dimensions, and we started with this minimum.

Step 3. Define the initial state. We chose a basis that provided the most meaningful parameters for the initial state. For this application, we chose to use the basis defined by the combination of variables A and I. Using this basis, the initial state $|\psi\rangle$ is represented by

$$|\psi\rangle = \sum_{w,y} \psi_{wy} \cdot |A = w \cap I = y\rangle.$$
 (14)

The four coefficients in Equation 14 form a 4×1 column matrix

$$\psi = \begin{bmatrix} \psi_{11} \\ \psi_{10} \\ \psi_{01} \\ \psi_{00} \end{bmatrix}.$$

For example, $|\psi_{10}|^2$ equals the probability of yes to A and no to I when this pair of questions is asked in that order. The parameters in ψ are estimated from the data under the constraint that $\psi^{\dagger}\psi=1$. In general, the 4 coefficients can be complex valued, and so each coefficient contributes a magnitude and a phase. However, the magnitudes must satisfy the unit length constraint that $\psi^{\dagger}\psi=1$. Also, one phase for one coefficient can be set to an arbitrary value without changing the final choice probabilities. Therefore, only $4\times 2-2=6$ free parameters are required for the initial state. These parameters tell us what the initial state of the psychological system (e.g., initial belief or attitude towards attributes A and I) is before any measurement is taken on the system, and can be used to compute the probability of certain response to an attribute when it is measured alone. That is, we can estimate more "context free" responses from the respondents—free from influences from measurement effects from the other attributes—even though we didn't collect such actual empirical data.

Step 4. Define projectors and state transitions. Define $M_n = diag \begin{bmatrix} 1 & 0 \end{bmatrix}$, $M_y = diag \begin{bmatrix} 0 & 1 \end{bmatrix}$, and $I_2 = diag \begin{bmatrix} 1 & 1 \end{bmatrix}$. The 4×4 matrix representation of the projector P(A = y) is the Kronecker product $(M_y \otimes I_2)$, which picks out the coordinates in ψ that are associated with the answer yes to attribute A. Likewise, the 4×4 matrix representation of the projector P(I = y) is the Kronecker product $(I_2 \otimes M_y)$. The 4×4 matrix representation of the projector P(H = y) is the Kronecker product $U_{HA} M_y U_{HA}^{\dagger} \otimes I_2$, which requires the use of the unitary matrix U_{HA} that transforms coordinates from the A to the H basis. The 4×4 matrix representation of the projector P(U = y) is the Kronecker product $I_2 \otimes \left(U_{UI} M_y U_{UI}^{\dagger}\right)$, which requires the use of the unitary matrix U_{UI} that transforms coordinates from the I to the U basis.

The 2×2 matrix representations, U_{HA} and U_{UI} , were determined from Equation 5 by selecting two 2×2 Hermitian matrices, H_{HA} and H_{UI} . The parameters of each of these Hermitian matrices were estimated from the data. Each 2×2 Hermitian matrix has four coefficients, two real diagonal values and

one complex off diagonal. However, one diagonal entry can be arbitrarily fixed, and so only 4-1=3 parameters are required for each 2×2 unitary matrix to produce a total of 6 parameters. These parameters determine the rotation from the basis of one variable to the basis of another variable. Psychologically, they tell us the relationship between the variables or attributes being examined, and can reveal the similarity between these variables, independent of the initial state (i.e., step 3) of the person. In addition, based on the unitary matrices, we can compute the probabilities the transition probabilities between basis vectors. For example, assuming each answer is represented by a single dimension, we can compute the transition probability from a response to one attribute (e.g., say "yes" to the attribute of attractiveness) to a response to another attribute (e.g., say "no" to the attribute of honesty).

Step 5. Compute choice probabilities for each response sequence. The choice probabilities for each sequence were computed by the product of projectors corresponding to the sequence. For example, the probability of YY for the AU table equals

$$\|P(U=y)P(A=y)|\psi\rangle\|^2 = \left\|\left(I_2 \otimes \left(U_{UI}M_yU_{UI}^{\dagger}\right)\right) \cdot \left(M_y \otimes I_2\right) \cdot \psi\right\|^2, \quad (15)$$

the probability of NN for the HI table equals

$$||P(I=n)P(H=n)|\psi\rangle||^2 = ||(I_2 \otimes M_n) \cdot \left(\left(U_{HA}M_nU_{HA}^{\dagger}\right) \otimes I_2\right) \cdot \psi||^2, \quad (16)$$

the probability YN for the AH table equals

$$\|P(H=n)P(A=y)|\psi\rangle\|^2 = \left\| \left(\left(U_{HA} M_n U_{HA}^{\dagger} \right) \otimes I_2 \right) \cdot \left(M_y \otimes I_2 \right) \cdot \psi \right\|^2, \quad (17)$$

the probability NY for the HA table equals

$$||P(A=y)P(H=n)|\psi\rangle||^2 = ||(M_y \otimes I_2) \cdot \left(\left(U_{HA}M_nU_{HA}^{\dagger}\right) \otimes I_2\right) \cdot \psi||^2. \quad (18)$$

For the artificial data, assume that each 2×2 table was based on 100 independent observations. The Hilbert space model has a total of 12 free parameters, which is 3 less than the 4-way joint probability model. Nevertheless, the Hilbert space model almost perfectly fits all the relative frequencies in Table 1, and the $G_{diff}^2=G_H^2-G_{Sat}^2=7.81\times 10^{-6}$. Step 6. The 6 model parameters representing ψ , along with the 6 parameters

Step 6. The 6 model parameters representing ψ , along with the 6 parameters representing U_{HA} and U_{IU} , are presented in the Appendix. Recall that the data were artificially generated for illustration, and so the results are not to be taken seriously. However, they help to show the application of a HSM model. We defer a more detailed discussion of results until the next section, where we report the results from fitting the model to real data from an experiment.

The initial state, ψ , can be used to compute the probability of responses to each attribute under the condition that the attribute was measured alone (see Table 2). For example, the predicted probability of answering Y (versus N) to

Table 2: Predicted probabilities of yes when each variable is measured alone

Attribute	A	Η	I	U
Probability	.4462	.3691	.3551	.3843

Table 3: Transition matrices between basis vector for pairs of incompatible attributes

	$ A=0\rangle$	$ A=1\rangle$		$ I=0\rangle$	$ I=1\rangle$
$ H=0\rangle$.7739	.2261	$ U=0\rangle$.8454	.1546
$ H=1\rangle$.2261	.7739	$ U=1\rangle$.1546	.8454

the attribute A when A is measured alone equals .4462 (probability of N equals 1-.4462=.5538), which is higher than the probability of answering yes to all the other variables. Note that Table 2 is not equal to the marginal probability of a 2×2 or higher order table involving pairs of incompatible variables.

The variables for a compatible pair, such as AI, are logically independent, and the HSM model predicts the 2-way joint distributions for each of the compatible pairs. (In this artificial case, the results are almost perfectly predicted.) The variables for incompatible pairs, such as AH and UI, are logically dependent, and so they do not provide a joint distribution, and instead they produce a probability for a sequence.

The unitary matrices, U_{AH} and U_{UI} , can be used to describe the transitions between sequential measurements. The squared magnitudes of the entries in the unitary matrices describe the probability of transiting from a basis vector representing a column attribute (e.g., $|A=w\rangle\otimes|I=y\rangle$) to a basis vector representing a row attribute (e.g., $|H=x\rangle\otimes|I=y\rangle$). Table 3 presents the transition probabilities for the two incompatible pairs, AH, and IU. For example, the probability of transiting to a positive answer for H from a positive answer to A equals .7738, which is lower than the probability, .8454, of transiting from a positive answer to I to a positive answer to U. In other words, the variables I,U are more similar to each other than the variables A,H.

Step 7. The transition matrices produced by unitary matrices are always symmetric. This is because each entry in the unitary matrix contains the inner product between vectors from different bases, e.g. $\langle H|A\rangle$, and the squared magnitude is the same in both directions, i.e., $|\langle H|A\rangle|^2 = |\langle A|H\rangle|^2$. Given the assumption in step 2 that events $(A=w\cap I=y)$, for all w,y, are represented by one dimensional subspaces (i.e., rays spanned by basis vectors $|A=w\rangle\otimes |I=y\rangle$), this implies symmetry in the conditional probabilities, i.e., p(H=x|A=w)=p(A=w|H=x) and p(I=x|A=w)=p(A=w|H=x). This is a very strong and empirically testable property of this simple quantum model [9]. However, this symmetry does not hold generally; if events are represented by multi-dimensional projectors instead of rays, then the conditionals can be asymmetric (see, e.g., Pothos et al. 42).

5 An Empirical Application

This section applies HSM modeling to a real experiment that was designed in a manner similar to the artificial example. A total of 184 participants made judgments on four attributes of anti-smoking 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 PSA's: One stimulus type included 8 examples warning about smoking causing death (Death PSAs), and the other stimulus type included 8 PSA's warning about smoking causing health harm (Harm PSAs). Each participant judged each PSA under 12 contexts: 6 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 for example YN means Yes to Persuasive and No to informative). Thus each person provided responses to 16 (PSA's) \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, presented separately for each stimulus type, but pooled across participants and order, are presented in Table 4 (later we present analyses at the individual level that include order.) For example, when the Death PSA was presented, the relative frequency of Y to Persuasive and N to Likable was .201, and the corresponding result for the Harm PSA was .176. Each 2×2 table for a pair of attributes is based on $184\cdot 16=2944$ observations. However, this table of pooled results ignores order effects and important individual differences, and so the subsequent analyses were conducted at the individual level of analysis. ¹⁰

5.1 Test of the joint probability model

Each individual produced a table in the same form as Table 4, but with 16 observations per 2×2 table (192 observations in total for both types of stimuli). Recall that the joint probability model states that the 6 rows of 2×2 tables are produced by a joint distribution, $\pi(P = w \cap B = x \cap I = y \cap L = z)$ where w = 0, 1, x = 0, 1, y = 0, 1 and z = 0, 1, that has 16 - 1 = 15 free parameters per stimulus type or 30 parameters altogether. The saturated model requires 3 parameters for each 2×2 table, producing a total of 18 parameters per stimulus

 $^{^{10}}$ 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 hyper parameters. 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 empirical support for specific parametric distributions of individual differences and lack of informative priors on the hyper-parameters 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 hyper priors.

Table 4: Observed Relative frequencies of pairs of answers for 6 different pairs of attributes

01 000115 0005	Death PSA				Harm PSA				
Attributes	YY	YN	NY	NN	Attributes	YY	YN	NY	NN
PI	.529	.166	.072	.232	PΙ	.438	.134	.049	.379
PB	.612	.092	.074	.223	PB	.459	.099	.069	.374
PL	.501	.201	.064	.235	PL	.378	.176	.083	.362
IB	.539	.074	.128	.259	IΒ	.419	.078	.109	.394
IL	.441	.181	.127	.251	IL	.324	.169	.124	.383
BL	.495	.188	.086	.232	BL	.356	.184	.102	.359

Table 5: Probabilities predicted by the HSM model
Death PSA
Harm PSA

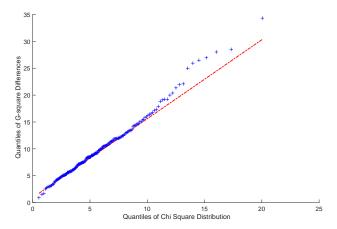
	Death	прл				main	IIDA		
Attributes	YY	YN	NY	NN	Attributes	YY	YN	NY	NN
PI	.544	.155	.065	.236	PΙ	.444	.122	.061	.373
PB	.610	.064	.055	.271	PB	.486	.064	.056	.394
PL	.507	.192	.069	.232	PL	.386	.180	.079	.355
IB	.539	.071	.132	.258	IΒ	.428	.077	.109	.386
IL	.441	.142	.124	.293	IL	.356	.142	.124	.378
BL	.493	.178	.083	.246	BL	.361	.176	.104	.359

type or 36 parameters altogether. Using maximum likelihood estimation for each person, we computed the G_{sat}^2 and G_{joint}^2 for each person. A quantile-quantile plot of the observed G^2 differences, $G_{diff}^2 = G_{joint}^2 - G_{sat}^2$ versus the χ^2 predicted by the null hypothesis is shown in Figure 1. As can be seen in Figure 1, the observed G_{diff}^2 exceeds the expected 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], and the difference is statistically significant $(\chi^2(2) = 78.84)$. We conclude that the 4-way joint probability model systematically deviates from the observed results for a substantial number of individuals.

5.2 Comparisons between Bayes net 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 Bayes net type of model. However, there may also be systematic deviations from a HSM model. Therefore, it is important to compare the fits of Bayes net versus HSM models. Because they two types of models are non-nested, we performed comparisons at the individual level using the Bayesian information criterion $BIC_{Model} = G^2_{Model} + p \cdot ln(192),$ where $G^2_{Model} = -2 \cdot loglikelihood,$ and p =number of model parameters.

Figure 1: Quantile - quantile plot of the observed versus predicted chi-square value



5.3 Simple Bayes net model

There are a large number of possible Bayes net type of models that one can construct for this application. We chose the 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 Bayes net models that perform better than the one here.

For the Bayes net type of model, we assumed that the two attributes, Informative (I) and Believable (B), are exogenous factors determined by the type of PSA's. Therefore, each type of stimulus produced a 2-way joint distribution with four joint probabilities $\pi(I = x \cap B = y|stimulus), x = 0, 1, y = 0, 1,$ and there are 2 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, which was represented by the conditional probabilities $\pi(P = w \cap L = z | I = x \cap B = y)$, for w = 0, 1 and z = 0, 1. However, this model produces the same number $(15 \cdot 2 = 30)$ of parameters as the 4 - way joint probability model. To simplify the model, we assumed independence, so that $\pi(P = w \cap L = z | I = x \cap B =$ $y) = \pi(P = w|I = x \cap B = y) \cdot \pi(L = z|I = x \cap B = y)$. We also assumed that the two conditionals, $\pi(P = w|I = x \cap B = y)$ and $\pi(L = z|I = x \cap B = y)$, did not depend on the stimulus type. Therefore, each of the two conditionals produces four parameters. Altogether, this model entails $(4-1) \cdot 2 + (4 \cdot 2) = 14$ parameters.

5.4 Simple HSM model

The same simple HSM model used to fit the artificial data in the previous section was applied to the real data from our experiment. First, we assumed that the attributes, Believable (B) and Informative (I) are compatible, which means we can think of these two attributes at the same time and the order of measuring the two attributes does not matter. This is consistent with the 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, and Likable (L) is a rotation of Informative. 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 simplicity, we assumed that each joint event for a compatible pair, such as $(B = x \cap I = y)$, is represented by a single dimension. We chose to represent the initial state and projectors by the basis described by the B,I events $(B = x \cap I = y)$.

To reduce the number of model parameters to a minimum, we restricted the coordinates of the initial state to be a real valued 4×1 matrix ψ with unit length $\psi^{\dagger}\psi = 1$. The unitary matrix for rotating between the incompatible basis vectors was constructed using a single parameter as follows:

$$H = i \cdot \theta \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

$$U = exp(-i \cdot H)$$

$$= \begin{bmatrix} cos(\pi \cdot \theta) & -sin(\pi \cdot \theta) \\ sin(\pi \cdot \theta) & cos(\pi \cdot \theta) \end{bmatrix}.$$
(19)

Parameters, θ_{PB} and θ_{LI} , were used to define rotation matrices U_{PB} for the P,B incompatible pair and U_{LI} for the I,L incompatible pair.

To account for the effect of type of stimulus, we allowed the initial state vector to vary across stimuli, ψ_{Death} and ψ_{harm} . However, according to the HSM model, 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. We tested this interesting prediction from HSM model 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 produced a total of $(3 \cdot 2) + 2 = 8$ parameters.

In sum, the HSM model starts with the 4-dimensional BI basis, which provides the coordinates that define the initial 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 4 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.

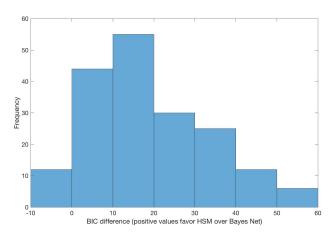


Figure 2: Frequency of BIC differences (BIC Bayes - BIC HSM)

5.5 Results of model comparisons

Maximum likelihood estimates and G^2 statistics were computed by fitting each model the model to the 192 observations separately for each participant. When comparing the 4-way joint probability model (30 parameters) to the constrained HSM model (8 parameter model), all 184 participants produced a BIC difference favoring the constrained HSM model. More interestingly, when comparing the Bayes net model (14 parameters) to the constrained HSM model (8 parameters), 148 out of 184 participants produced BIC differences favoring the constrained HSM model. The distribution of BIC differences ($BIC_{diff} = BIC_{Bayes} - BIC_{HSM}$) is presented in Figure 2.

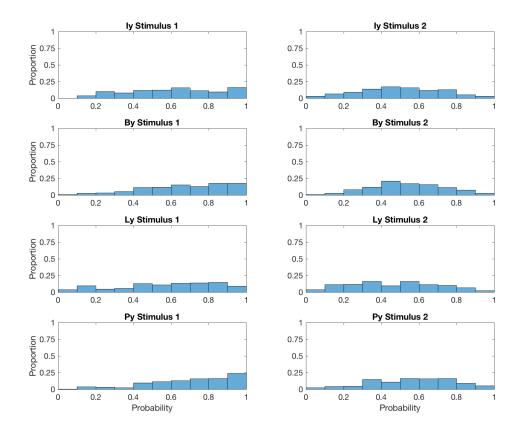
Finally, we tested the prediction of the HSM model that the rotation parameters are invariant across stimuli by comparing the $BIC_{constrained}$ to $BIC_{unconstrained}$ versions of the HSM model. In agreement with the prediction, the BIC difference favored the constrained model for 154 out of 184 participants.

The predictions generated by the constrained HSM model, pooled across participants and presentation order are presented in Table 5. As can be seen in the table, the constrained HSM model does a very good job of predicting the pooled results. The most important errors occur for the incompatible variables, where we constrained the model to use the same parameters across stimulus types.

5.6 Interpretation of parameters

The HSM model provides two sets of model parameters for each participant. One set, which is based on the initial 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 3 presents the relative frequency distribution

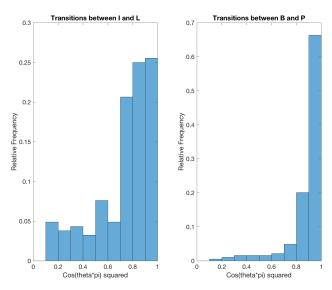
Figure 3: Relative frequency distribution for the probabilities of a "yes" answer to each of the I, B, L, and P attribute when measured alone (left panels, death PSAs; right panels, harm appeal PSAs)



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 3). In general, participants responded more positively towards 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

Figure 4: Relative frequency of transition probability from yes in one basis to yes in another



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 transitioning from basis vectors for B to basis vectors for P. Figure 4 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 4, 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. Interestingly, however, the similarity between P and B tends to be higher across all participants than that for L and I; in addition, there are larger individual differences for the L and I transitions since the parameter distribution is more widely distributed compared to that for the P and B transitions (see Figure 4).

6 Summary and Extensions

In this article, we presented the general procedures for building HSM models based on quantum probability theory. These 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. HSM models provide tools for modeling context effects, and the model parameters provide two psychologically meaningful and useful interpretations of these effects. First of all, the state vector of an HSM model provides an estimation of the respondents' initial 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 inter-relations between the p measurements, independent of the initial 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.

HSM models provide new contributions to the current set of probabilistic and statistical tools for contingency table analysis. Loglinear/categorical data models only apply to 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, and it is often the case that no complete 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 (a) violations of consistency requirements of the p-way joint distribution model, (b) a non-parametric statistical test of a p-way joint distribution for a collection of tables, and (c) illustrate the application of an HSM model to a concrete example. 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 Bayes net model using Bayesian information criteria. We conclude from these analyses that HSM models are empirically viable for modeling collections of contingency tables.

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 [30], 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 [17]. 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 [37]. 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

- [1] D. Aerts. Quantum structure in cognition. *Journal of Mathematical Psychology*, 53 (5):314–348, 2009.
- [2] D. Aerts and L. Gabora. A theory of concepts and their combinations ii: A hilbert space representation. *Kybernetes*, 34:192–221, 2005.
- [3] D. Aerts, J. Broekaert, and L. Gabora. A case for applying an abstracted quantum formalism to cognition. New Ideas in Psychology, 29:136–146, 2011.
- [4] D. Aerts, L. Gabora, and S. Sozzo. Concepts and their dynamics: A quantum theoretic modeling of human thought. *Topics in Cognitive Science*, 5:737–773, 2013.
- [5] A. Agresti and M. Katera. Categorical data analysis. Springer, 2011.
- [6] C. J. Anderson. Analysis of multivariate frequency data by graphical models and generalizations of the multidimensional row-column association model. *Psychological methods*, 7(4):446–467, 2002.
- [7] Masanari Asano, Takahisa Hashimoto, Andrei Khrennikov, Masanori Ohya, and Yoshiharu Tanaka. Violation of contextual generalization of the leggett-garg inequality for recognition of ambiguous figures. *Physica Scripta*, 2014(T163):014006, 2014.
- [8] Masanari Asano, Irina Basieva, Andrei Khrennikov, Masanori Ohya, and Yoshiharu Tanaka. A quantum-like model of selection behavior. *Journal* of Mathematical Psychology, 2016.
- [9] Thomas Boyer-Kassem, Sébastien Duchêne, and Eric Guerci. Testing quantum-like models of judgment for question order effect. *Mathematical Social Sciences*, 80:33–46, 2016.
- [10] P. D. Bruza, Z. Wang, and J. R. Busemeyer. Quantum cognition: a new theoretical approach to psychology. *Trends in cognitive sciences*, 19(7): 383–393, 2015.
- [11] Peter D Bruza, Kirsty Kitto, Brentyn J Ramm, and Laurianne Sitbon. A probabilistic framework for analysing the compositionality of conceptual combinations. *Journal of Mathematical Psychology*, 67:26–38, 2015.
- [12] J. R. Busemeyer and P. D. Bruza. Quantum models of cognition and decision. Cambirdge University Press, 2012.
- [13] J. R. Busemeyer, Z. Wang, and A. Lambert-Mogiliansky. Empirical comparison of markov and quantum models of decision making. *Journal of Mathematical Psychology*, 53 (5):423–433, 2009.

- [14] J. R. Busemeyer, E. M. Pothos, R. Franco, and J. S. Trueblood. A quantum theoretical explanation for probability judgment errors. *Psychological Review*, 118 (2):193–218, 2011.
- [15] E. Conte, A. Y. Khrennikov, O. Todarello, A. Federici, L. Mendolicchio, and J. P. Zbilut. Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems and Information Dynamics*, 16:1–17, 2009.
- [16] Adnan Darwiche. Modeling and reasoning with Bayesian networks. Cambridge University Press, 2009.
- [17] Kim De Roover, Eva Ceulemans, Marieke E Timmerman, Kristof Vansteelandt, Jeroen Stouten, and Patrick Onghena. Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological methods*, 17(1):100–119, 2012.
- [18] G. M. Di Nunzio, P. Bruza, and L. Sitbon. *Interference in Text Catego*rization Experiments, pages 22–33. Springer, 2014.
- [19] P. A. M. Dirac. *The principles of quantum mechanics*. Oxford University Press, 1930/1958.
- [20] E. Dzhafarov and J. V. Kujala. Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychol*ogy, 56:54–63, 2012.
- [21] E. N. Dzhafarov and J. V. Kujala. All-possible-couplings approach to measuring probabilistic context. *PloS one*, 8(5):e61712, 2013.
- [22] Ehtibar N Dzhafarov, Ru Zhang, and Janne Kujala. Is there contextuality in behavioural and social systems? *Phil. Trans. R. Soc. A*, 374(2058): 20150099, 2016.
- [23] A. Fine. Joint distributions, quantum correlations and commuting observables. *Journal of Mathematical Physics*, 23(7):1306–1310, 1982.
- [24] A. M. Gleason. Measures on the closed subspaces of a hilbert space. *Journal of Mathematical Mechanics*, 6:885–893, 1957.
- [25] P beim Graben and H. Atmanspacher. Complementarity in classical dynamical systems. *Foundations of Physics*, 36:291–306, 2006.
- [26] G. Gronchi and E. Strambini. Quantum cognition and bells inequality, a model for probabilistic judgment bias. *Journal of Mathematical Psychology*, 2016.
- [27] S. P. Gudder. Quantum Probability. Academic Press, 1988.
- [28] P. R. Halmos. Finite-Dimensional vector spaces. Springer Undergraduate Texts in Mathematics, 1993.

- [29] E. Haven and A. Khrennikov. *Quantum social science*. Cambirdge University Press, 2013.
- [30] Joel Huber, John W Payne, and Christopher Puto. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1):90–98, 1982.
- [31] A. Y. Khrennikov. *Ubiquitous quantum structure: From Psychology to Finance*. Springer, 2010.
- [32] A. N. Kolmogorov. Foundations of the theory of probability. N.Y.: Chelsea Publishing Co., 1933/1950.
- [33] P. D. Kvam, T. J. Pleskac, S. Yu, and J. R. Busemeyer. Interference effects of choice on confidence., 112 (34). Proceedings of the National Academy of Science, 112 (34):10645–10650, 2015.
- [34] P. La Mura. Projective expected utility. *Journal of Mathematical Psychology*, 53 (5):408–414, 2009.
- [35] A. Lambert-Mogiliansky, S. Zamir, and H. Zwirn. Type indeterminacy: A model of the 'kt' (kahneman-tversky)-man. *Journal of Mathematical Psychology*, 53 (5):349–361, 2009.
- [36] A. J. Leggett and A. Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks. *Physical Review Letters*, 54:857–860, 1985.
- [37] John J McArdle, Kevin J Grimm, Fumiaki Hamagami, Ryan P Bowles, and William Meredith. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological methods*, 14(2):126–149, 2009.
- [38] I Pitowski. Quantum probability, Quantum logic., volume 321 Lecture Notes in Physics. Springer, 1989.
- [39] Arkady Plotnitsky. Niels Bohr and complementarity: An introduction. Springer Science & Business Media, 2012.
- [40] E. M. Pothos and J. R. Busemeyer. A quantum probability model explanation for violations of 'rational' decision making. *Proceedings of the Royal Society*, B., 276 (1665):2171–2178, 2009.
- [41] E. M. Pothos and J. R. Busemeyer. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36: 255–274, 2012.
- [42] E. M. Pothos, J. R. Busemeyer, and J. S. Trueblood. A quantum geometric model of similarity. *Psychological Review*, 120 (3):679–696, 2013.

- [43] H. Schuman and S. Presser. Questions and answers in attitude surveys: Experiments on question form, wording, and content. NY: Academic Press, 1981.
- [44] P. Suppes and M. Zanotti. When are probabilistic explanations probable? *Synthese*, 48(191-199), 1981.
- [45] Patrick Suppes. Probability concepts in quantum mechanics. *Philosophy of Science*, 28(4):378–389, 1961.
- [46] R. Tourangeau, L. J. Rips, and K. A. Rasinski. *The psychology of survey response*. Cambridge University Press, 2000.
- [47] J. S. Trueblood and J. R. Busemeyer. A quantum probability account for order effects on inference. *Cognitive Science*, 35:1518–1552, 2010.
- [48] C.J. van Rijsbergen. *The Geometry of Information Retrieval*. Cambridge University Press, 2004.
- [49] J. Von Neumann. Mathematical Foundations of Quantum Theory. Princeton University Press, 1932/1955.
- [50] Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. Exploration of quantum interference in document relevance judgement discrepancy. *Entropy*, 18(4):144, 2016.
- [51] Z. Wang and J. R. Busemeyer. A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4):689–710, 2013.
- [52] Z. Wang and J. R. Busemeyer. Interference effects of categorization on decision making. *Cognition*, 150:133–149, 2016.
- [53] Z Wang and JR Busemeyer. Comparing quantum versus markov random walk models of judgements measured by rating scales. *Phil. Trans. R. Soc.* A, 374(2058):20150098, 2016.
- [54] Z. Wang, J. R. Busemeyer, H. Atmanspacher, and E. M. Pothos. The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, 5(4):672–688, 2013.
- [55] Z. Wang, T. Solloway, R. M. Shiffrin, and J. R. Busemeyer. Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the USA*, 111 (26): 9431–9436, 2014.
- [56] Zheng Wang and Jerome Busemeyer. Reintroducing the concept of complementarity into psychology. Frontiers in psychology, 6(1822), 2015.
- [57] Forrest W Young. Multidimensional scaling: History, theory, and applications. Psychology Press, 2013.

[58] V. I. Yukalov and D. Sornette. Decision theory with prospect interference and entanglement. *Theory and Decision*, 70:283–328, 2011.

7 Appendix

7.1 Matrix exponential function

Suppose H is the matrix representation of a Hermitian operator. 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

$$exp(H) = V \cdot exp(\Lambda) \cdot V^{\dagger},$$

$$exp(\Lambda) = diag \left[e^{\lambda_1} \cdots e^{\lambda_j} \cdots e^{\lambda_N} \right].$$

7.2 Kronecker product

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

$$P \otimes Q = \left[\begin{array}{cccc} p_{11} \cdot Q & \vdots & p_{1n} \cdot Q \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & p_{ij} \cdot Q & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ p_{m1} \cdot Q & \vdots & p_{mn} \cdot Q \end{array} \right].$$

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):

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

7.3 Parameters used to fit artificial data

$$\psi = \begin{bmatrix} .5203 \cdot e^{0}. \\ .4189 \cdot e^{i \cdot 2.2920} \\ .2904 \cdot e^{i \cdot 0.9383} \\ .6852 \cdot e^{i \cdot 0400} \end{bmatrix}$$

Note that $.5203 = \sqrt{1 - (.6852^2 + .2904^2 + .4189^2)}$.

$$H_{HA} = \begin{bmatrix} -.5911 & -.5037 \cdot e^{i \cdot .8862} \\ -.5037 \cdot e^{-i \cdot .8862} & 0 \end{bmatrix}$$

Table 3 is obtained by squaring the magnitudes of the entries.

$$H_{UI} = \begin{bmatrix} -1.2405 & -.4334 \cdot e^{i \cdot 1.2976} \\ -.4335 \cdot e^{-i \cdot 1.2976} & 0 \end{bmatrix}$$

Table 3 is obtained by squaring the magnitudes of the entries.