Quantum-Like Structure in Multidimensional Relevance Judgements

Sagar Uprety[1(✉)], Prayag Tiwari[2], Shahram Dehdashti[3], Lauren Fell[3], Dawei Song[1,4], Peter Bruza[3], and Massimo Melucci[2]

[1] The Open University, Milton Keynes, UK
sagar.uprety@open.ac.uk
[2] University of Padova, Padua, Italy
[3] Queensland University of Technology, Brisbane, Australia
[4] Beijing Institute of Technology, Beijing, China

Abstract. A large number of studies in cognitive science have revealed that probabilistic outcomes of certain human decisions do not agree with the axioms of classical probability theory. The field of Quantum Cognition provides an alternative probabilistic model to explain such paradoxical findings. It posits that cognitive systems have an underlying quantum-like structure, especially in decision-making under uncertainty. In this paper, we hypothesise that relevance judgement, being a multidimensional, cognitive concept, can be used to probe the quantum-like structure for modelling users' cognitive states in information seeking. Extending from an experiment protocol inspired by the Stern-Gerlach experiment in Quantum Physics, we design a crowd-sourced user study to show violation of the Kolmogorovian probability axioms as a proof of the quantum-like structure, and provide a comparison between a quantum probabilistic model and a Bayesian model for predictions of relevance.

Keywords: Multidimensional relevance · User behaviour · Quantum Cognition

1 Introduction

Relevance in Information Retrieval (IR) is widely accepted to be a cognitive feature, driving all our information interactions. All areas of research within IR thus strive to improve relevance of documents to a user's information need (IN). These research areas of IR can be broadly divided into two: system-oriented and user-oriented IR. Whereas the system-oriented viewpoint ties relevance to be an objective property of the document and query content, the user-oriented approach to IR views relevance as a cognitive property. Although IR fundamentally involves user interaction and decision-making, the user-oriented approach has been found harder to implement, especially in evaluating performance of IR systems. This is because of the variability in user judgements of relevance [5].

J. M. Jose et al. (Eds.): ECIR 2020, LNCS 12035, pp. 728–742, 2020.
https://doi.org/10.1007/978-3-030-45439-5_48

System-oriented IR thus sought to standardise IR evaluation, in which the user-cognitive notion of relevance was replaced by an objective, topical relevance. This led to evaluation methodologies based on the Cranfield and TREC type test collections. The user and all of his/her contexts were removed from the evaluation process.

Recent surge in availability of online user data has led to incorporation of more user context in the computation of relevance, e.g. in learning based ranking algorithms. This context is based on the user's past interactions with the system, in addition to user attributes like age, interests, etc. and current attributes like location, type of device, etc. The common feature in these various contexts is that they are static. They are determined before the point of user's interaction with the IR system. However, the process of IR is interactive and dynamic. In this paper, we focus on another type of context driving user interactions - dynamic context. Dynamic context is one which changes user's cognitive state *during* information interaction.

One well-known example of when a dynamic context affects relevance is the phenomenon of Order Effect [8]. Order effects have been investigated and found to exist in IR in the presentation order of documents [4,6,9,24]. For example, in a recent study reported in [22], two groups of participants were presented with a pair of documents D_1 and D_2 in two different orders. For some of such pairs, it was found that the relevance of a document judged by users is different depending on the order it was presented. Although the phenomenon may appear to have an intuitive explanation, it violates one of the fundamental assumptions of classical probability theory - joint distributions, where, for two random variables representing relevance of the documents - R_1, R_2, $P(R_1, R_2) = P(R_2, R_1)$, i.e., the order of judging the documents does not matter. Order effects violate this fundamental assumption. Such order effects have also been investigated and reported in between the different dimensions of relevance, like Topicality, Understandability, Reliability, etc. [1,19,20], where different orders of dimensions considered to judge a document lead to different relevance judgements.

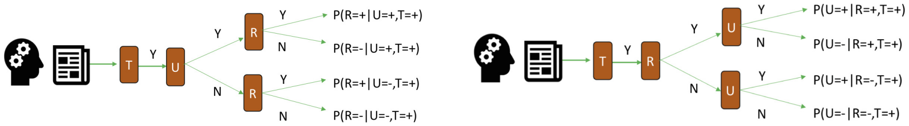
The field of Quantum Cognition [2] offers a generalised framework to model probabilistic outcomes of human decision-making. It has been successful in modelling and predicting order effects [16,23] and other paradoxical findings where axioms of classical probability theory are violated [3,14]. Conceptually, it challenges the notion that cognitive states have pre-defined values and that a measurement merely records them. Instead, the act of measurement creates a definite state out of an indefinite state and in doing so, changes the initial state of the cognitive system. In terms of relevance, we cannot pre-assign relevance of a document for a user. Instead, relevance is defined only at the point of interaction of the user's cognitive state with the document. Therefore, judgement of document D_2 first, changes user's initial state and the subsequent judgement of relevance of D_1 is different than when D_1 is judged before D_2. Should relevance of the documents for a user be a pre-defined entity, it would not be influenced by judgement of other documents and a joint distribution over relevance of the two documents would exist. We also say that these two measurements of relevance are incompatible with each other. That is, it is not possible to jointly consider the relevance of the two documents, at the same time. At the mathematical level, measurements in quantum theory are represented by operators, which in general, do not commute with each other.

In a classical system, all measurements will commute with each other. However, conversely, commutativity of measurements does not necessarily imply that the system is classical. Therefore, the type of measurements becomes imperative in identifying a quantum system. Even then, not all measurements on quantum systems generate data violating the classical probability theory. The system needs to be probed in a way which exploits the underlying quantum structure. In physics, this was done by experiments such as Stern-Gerlach and double-slit experiments [15] which showed the violation of classical probability principles for microscopic particles like electrons and photons. In cognitive science too, several experiments performed by Tversky, Kahneman and colleagues showed such violations in human decision-making under uncertainty [17].

Recently, an experiment protocol inspired by the Stern-Gerlach experiment in Physics has provided a new way to probe cognitive systems such that they exhibit a quantum-like structure [7]. By quantum-like structure we mean the representation of a system using the mathematical framework of quantum theory in order to model and predict the experimental data. In [19], this experiment was performed in an IR scenario involving judgement of relevance with respect to different dimensions. Extending from the Stern-Gerlach protocol, in this paper we design a new experiment to show the violation of classical probability theory in multidimensional relevance judgements. We hypothesise that multidimensional relevance judgement has an underlying quantum-like structure, which when subject to appropriate measurement design can exhibit violations of classical probability theory. Specifically, we investigate the violation of a particular axiom of Kolmogorovian probability theory [11]. Our results show that the experimental data indeed violates classical probability theory, and a quantum framework provides more accurate predictions to describe the data. This experiment not only shows the necessity of the quantum framework as an alternative for constructing probabilistic models, but also gives novel insights into user behaviour in IR. This understanding can contribute to improvement of interactive IR systems and we also discuss such implications in this paper.

2 Stern-Gerlach Inspired Protocol for Multidimensional Relevance

The basis of the research reported in this paper is the cognitive analogue of the Stern-Gerlach (S-G) experiment, originally conducted in [19]. The S-G experiment [15] was an important milestone in quantum physics as it showed the non-classical behaviour of microscopic systems. The key was a particular design of the experiment which exploited the incompatibility between measurement of electron spin states along different axes. An electron has a particular property



(a) Asking three questions in TUR order

(b) Asking three questions in TRU order

Fig. 1. S-G type experiment to construct a complex-valued Hilbert space

called spin, having two possible values - up (+), down (-), which can be measured along different axes. An electron may have spin state + along the x-axis but state − along y-axis. So the outcome of measurement of the spin property of the electron depends upon the axis of measurement. Also, any measurement of spin disturbs the system. If a measurement of spin is made along X axis and Z axis, then a third measurement along X axis may give a different answer than the first one. This phenomena is called measurement incompatibility, where two measurements cannot be jointly conducted on a system - one measurement disturbs the system and the other would then measure the changed system.

The S-G experiment also describes the minimum number of measurements required from a system to construct a complex-valued Hilbert Space structure. In particular, we need three incompatible measurements each with two mutually exclusive outcomes. We can use this arrangement of measuring properties of a quantum system to measure relevance of a document in IR. For this, we consider three dimensions of relevance: Topicality (T) - whether a document is topically relevant to a query, Understandability (U) - how easy it is to understand the content of the document, and Reliability (R) - how much can the document be relied upon. Each of these three dimensions can be posed as questions requiring a Yes/No type answer (denoted as + and − respectively) for a document. These three dimensions are important factors considered by users for deciding relevance. Besides, they are tied to a single document, unlike diversity or novelty, which is always considered in comparison with other documents. Certain dimensions like Interest, Habit, etc. are difficult to ascertain via crowdsourcing. As reported in [1], the different relevance dimensions can exhibit incompatibility for certain query-document pairs.

In [19], three query-document pairs were designed in such a way as to potentially exhibit incompatibility between judgement of relevance with respect to different dimensions. The content of the documents was altered to introduce uncertainty in judging each of the three dimensions. The participants were presented with three questions related to three relevance dimensions, for each query-document pair, in line with the S-G design. Figure 1 shows the three questions asked to two different groups in different orders. More details about this design can be found in [19] and [7]. This setup enables one to construct a complex-valued Hilbert space, which models the quantum-like structure of the user's cognitive state during information interaction.

2.1 Constructing Complex-Valued Hilbert Space

The first step in building a quantum probabilistic model is to construct a representation for the user's cognitive state. In the quantum framework, a complex-valued Hilbert space is used to represent a quantum system, and the state of the system is represented as a vector in this Hilbert space.

Following the convention used in Quantum Physics, we represent any complex-valued vector A in a finite dimensional Hilbert space as a ket vector $|A\rangle$ and its complex conjugate as a bra vector $\langle A|$. The norm of this vector is the square root of its inner product with its conjugate - $|\langle A|A\rangle|^{1/2}$. For two such vectors, their projection onto each other is given as the square of their inner product - $|\langle A|B\rangle|^2$. Each vector is written as a linear combination of the vectors of the basis in which it is represented. For the purpose of representing the cognitive state of a person judging a document as topically relevant or topically irrelevant, we consider a basis formed by two orthogonal vectors $|T+\rangle$ and $|T-\rangle$ respectively. Before a user considers a judgement of topicality, the cognitive state is indefinite with respect to considering the document as topically relevant or irrelevant. Both potentialities exist. We say that the cognitive state collapses to either $|T+\rangle$ or $|T-\rangle$ after the judgement. Before the judgement, we can represent the indefinite cognitive state in terms of probabilities of its potential responses. This is represented as a linear combination of the two basis states, weighted each by real or complex coefficients (called probability amplitudes), such that the square of the probability amplitude gives the probability of collapsing to the respective state. The initial state S is thus written as:

$$|S\rangle = t\,|T+\rangle + \sqrt{1-t^2}\,|T-\rangle \tag{1}$$

Query	P(T)	P(U|T)	P(R|U,T)	P(R|T)	P(U|R,T)
Query 1	P(T=+) = 0.762	P(U = +|T+) = 0.578	P(R = +|U=+,T=+) = 0.587	P(R = +|T+) = 0.546	P(U = +|R=+,T=+) = 0.600
			P(R = -|U=+,T=+) = 0.413		P(U = -|R=+,T=+) = 0.400
		P(U = -|T+) = 0.422	P(R = +|U=-,T=+) = 0.370	P(R = -|T+) = 0.454	P(U = +|R=-,T=+) = 0.407
			P(R = -|U=-,T=+) = 0.630		P(U = -|R=-,T=+) = 0.593
Query 2	P(T=+) = 0.671	P(U = +|T+) = 0.802	P(R = +|U=+,T=+) = 0.844	P(R = +|T+) = 0.731	P(U = +|R=+,T=+) = 0.882
			P(R = -|U=+,T=+) = 0.156		P(U = -|R=+,T=+) = 0.118
		P(U = -|T+) = 0.198	P(R = +|U=-,T=+) = 0.526	P(R = -|T+) = 0.269	P(U = +|R=-,T=+) = 0.480
			P(R = -|U=-,T=+) = 0.474		P(U = -|R=-,T=+) = 0.520
Query 3	P(T=+) = 0.899	P(U = +|T+) = 0.977	P(R = +|U=+,T=+) = 0.738	P(R = +|T+) = 0.646	P(U = +|R=+,T=+) = 0.963
			P(R = -|U=+,T=+) = 0.262		P(U = -|R=+,T=+) = 0.037
		P(U = -|T+) = 0.023	P(R = +|U=-,T=+) = 0.000	P(R = -|T+) = 0.354	P(U = +|R=-,T=+) = 0.889
			P(R = -|U=-,T=+) = 1.000		P(U = -|R=-,T=+) = 0.111

Fig. 2. Probabilities for the questions of TUR and TRU for the three queries

In the S-G inspired experiment design, we ask the user sequential questions about judgement of Topicality (T), Understandability (U) and Reliability (R) in the order TUR or TRU, as shown in Fig. 1. Therefore we represent the cognitive state w.r.t Understandability and Reliability in term of Topicality:

$$|U+\rangle = u\,|T+\rangle + \sqrt{1-u^2}\,|T-\rangle\,, |U-\rangle = \sqrt{1-u^2}\,|T+\rangle - u\,|T-\rangle \tag{2}$$

$|U-\rangle$ is constructed using the fact that $|U+\rangle$ and $|U-\rangle$ are orthogonal. u^2 is the probability that users judge a document Understandable, given that they have judged it as Topically relevant.

Refer to [19] (Section 3) or [15] (Chapter 1) for the necessity of using a complex-valued probability amplitude in the representation of Reliability in term of Topicality:

$$|R+\rangle = r\,|T+\rangle + \sqrt{1-r^2}e^{i\theta_r}\,|T-\rangle\,, |R-\rangle = \sqrt{1-r^2}e^{-i\theta_r}\,|T+\rangle - r\,|T-\rangle \tag{3}$$

The parameters (u, r and θ_r) comprise the construction of the Hilbert space for user's cognitive state w.r.t the interaction between the three dimensions. The parameter t defines the initial state. The experiment design of Fig. 1 was carried out in [19] for three queries. The results are listed in Fig. 2.

3 Formulation of Research Hypotheses

Using the complex-valued Hilbert Space of multidimensional relevance, this paper aims to design an extended experiment to test the following research hypotheses: (1) Fundamental axioms of classical Kolmogorov probability are violated in a multidimensional relevance judgement scenario; (2) Probabilities obtained from the experiment can be better predicted with quantum than classical (Bayesian) probabilistic models. In the following two subsections, we mathematically formulate these hypotheses.

3.1 Violation of Kolmogorov Probability and Quantum Correction

Quantum probabilities are generalisation of Kolmogorov probabilities. In fact, Kolmogorov probabilities are related to set theory which formalises Boolean logic. The following proposition gives one of their fundamental properties [11]:

$$0 = \delta = P(A \vee B) - P(A) - P(B) + P(A \wedge B) \tag{4}$$

where A, B are subsets of the set of all alternatives Ω, and $P(A)$, $P(B)$ are the corresponding probabilities. The axiom will be violated if the value of δ is different from zero.

In the quantum probability theory, the computation of probabilities are represented by projection operators for the events $U\pm$ and $R\pm$ corresponding to relevance or non-relevance with respect to Understandability and Reliability.

The analogue of relation (4) in quantum mechanics is given by the following definition [21]:

$$\mathcal{D}(U\pm, R\pm) = \Pi(U \pm \vee R\pm) - \Pi(U\pm) - \Pi(R\pm) + \Pi(U \pm \wedge R\pm) \tag{5}$$

where projection operators $\Pi(U\pm)$ and $\Pi(R\pm)$ are given by:

$$\Pi(U\pm) = |U\pm\rangle \langle U\pm| \,, \qquad \Pi(R\pm) = |R\pm\rangle \langle R\pm| \tag{6}$$

It is possible to prove that this quantum correction term $\mathcal{D}(U\pm, R\pm)$ is proportional to the commutator of the projection operators of $U\pm$ and $R\pm$ [21] and can be thus obtained as:

$$\mathcal{D}(U\pm, R\pm) = \left[\Pi(U\pm), \Pi(R\pm)\right] \left(\Pi(U\pm) - \Pi(R\pm)\right)^{-1} \tag{7}$$

where $[A, B]$ stands for the commutator for two operators A and B. The projection operator $\Pi(U+)$ is equal to the outer product of the state $|U+\rangle$ with itself, where the vector $|U+\rangle$ is computed using Eq. 2. In order to construct the vector, first the Topicality basis is represented as the standard basis and hence the orthogonal vectors $|T+\rangle$ and $|T-\rangle$ are given as:

$$|T+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |T-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8}$$

Thus, vectors $|U+\rangle$ and $|U-\rangle$ are given as:

$$|U+\rangle = \begin{pmatrix} u \\ \sqrt{1-u^2} \end{pmatrix}, \quad |U-\rangle = \begin{pmatrix} \sqrt{1-u^2} \\ -u \end{pmatrix} \tag{9}$$

Then the projector $\Pi(U+)$ is given as:

$$\Pi(U+) = |U+\rangle \langle U+| = \begin{pmatrix} u \\ \sqrt{1-u^2} \end{pmatrix} \begin{pmatrix} u & \sqrt{1-u^2} \end{pmatrix} = \begin{pmatrix} u^2 & u\sqrt{1-u^2} \\ u\sqrt{1-u^2} & 1-u^2 \end{pmatrix}$$

Similarly, $\Pi(R+)$ is:

$$\begin{aligned} \Pi(R+) = |R+\rangle \langle R+| &= \begin{pmatrix} r \\ \sqrt{1-r^2}e^{i\theta_r} \end{pmatrix} \begin{pmatrix} r & \sqrt{1-r^2}e^{-i\theta_r} \end{pmatrix} \\ &= \begin{pmatrix} r^2 & r\sqrt{1-r^2}e^{-i\theta_r} \\ r\sqrt{1-r^2}e^{i\theta_r} & 1-r^2 \end{pmatrix} \end{aligned}$$

From the values of u, r and θ_r obtained in [19], these projection operators can be constructed. The quantum analogue of δ, can then be calculated from Eq. (7). Value of δ obtained from our experiment is compared to that predicted by the classical (always zero) and quantum probability frameworks.

3.2 Quantum Probabilities vs Classical Probabilities

The violation of Kolmogorovian probability axiom by a given system would likely lead to inaccurate predictions on the system using Kolmogorovian probability. This subsection formulates computation of conditional probabilities of relevance judgement along one dimension given another, using classical vs. quantum frameworks. They will be compared for our experimental data in Sect. 5.

For an initial state of the system $|S\rangle$, the probability of event $|T+\rangle$ in the quantum framework is given by $P(T+) = |\langle T+|S\rangle|^2 = t^2$, i.e., square of projection of vector $|S\rangle$ onto vector $|T+\rangle$. The probability for sequence $U+$ following $T+$ is given as [2]:

$$P(U+, T+) = |\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2 \tag{10}$$

The quantum framework does not define joint probability of events T and U, as in general $P(T+, U+) \neq P(U+, T+)$. As we can see $P(T+, U+) = |\langle T+|U+\rangle|^2 |\langle U+|S\rangle|^2$, which for $\langle U+|S\rangle \neq \langle T+|S\rangle$ is not equal to $P(U+, T+)$ in Eq. 10. The conditional probabilities are given according to Luder's rule [2,10] as:

$$\begin{aligned} P_q(U+|T+) &= P(U+, T+)/P(T+|S) \\ &= \frac{|\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2}{|\langle T+|S\rangle|^2} \\ &= |\langle U+|T+\rangle|^2 = u^2 \end{aligned} \tag{11}$$

Note that subscript q is added to distinguish from classical conditional probability. Then $P_q(R+|U+, T+)$ is given as (see [19] Sect. 4.2 for derivation):

$$\begin{aligned} P_q(R+|U+, T+) &= |\langle R+|U+\rangle|^2 \\ &= (ur)^2 + (1-u^2)(1-r^2) + 2ur\sqrt{(1-u^2)(1-r^2)}\cos\theta_r \end{aligned} \tag{12}$$

In contrast, classical probability theory has the basic assumption of commutativity of two events. Therefore the joint probability distribution always exists, which is the basis of calculating conditional probabilities in Bayes' rule. Consequently, for events T, U and R we have:

$$P(U+, R+, T+) = P(R+, U+, T+) \tag{13}$$

which can be written in terms of conditional probabilities as:

$$P(T+)P(R+|T+)P(U+|R+, T+) = P(T+)P(U+|T+)P(R+|U+, T+) \tag{14}$$

This enables calculation of conditional probabilities using the Bayes rule:

$$P(R+|U+, T+) = \frac{P(U+|R+, T+)P(R+|T+)}{P(U+|T+)} \tag{15}$$

Similarly, the other conditional probabilities can be obtained. Again, note that the probabilities in Eqs. (15) and (12) are different because of the difference in the underlying assumption of commutativity or joint probability.

4 Experiment

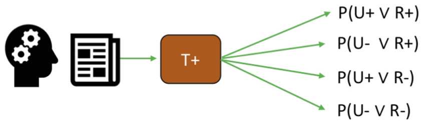
4.1 Methodology

The main aim of this experiment is to investigate the violation of Eq. 4. We already have the single question probabilities from the experiment in [19] and we need to obtain the probabilities of conjunction and disjunction. We do so by posing questions about Understandability and Reliability at the same time, as a pair, rather than sequentially. Each of the dimensions have two outcomes (e.g. Reliable or Not Reliable) and therefore we construct four pairs of statements, as listed in Fig. 4. For the disjunction measurement, we ask the participants to select whether they agree with at least one of the two statements or none of them (corresponding to a Boolean Or condition). For a conjunction measurement on each of the four statement pairs, we ask the participants whether they agree with both of the questions or not. Figure 5a and b show the designs for the disjunction and conjunction questions for a query-document pair. We now have a total of eight such questions and we follow a between-subjects design such that a participant is shown only one of these eight questions randomly. Note that we are able to use the probabilities from the experiment in [19] because our experiment is a between-subjects design. The same participant is not asked all the questions - to avoid memory bias. The design is summarised in the following steps for each of the three query-document pairs:

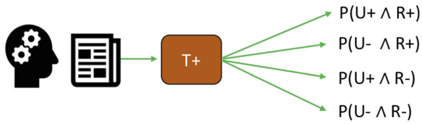
1. The participants are shown information need, query and document snippet.
2. Next, they are asked a Yes/No question about the Topicality of the document. This is to prepare the cognitive state of all participants by projecting their initial/background state onto the Topicality subspace of the underlying Hilbert space constructed in the previous experiment in [19].
3. Lastly, they are randomly shown one of the eight possible conjunction or disjunction questions and asked to choose the appropriate answer (Fig. 3).

4.2 Participants and Material

We recruited 335 participants for the experiment using the online crowd-sourcing platform Prolific (prolific.ac). The study was designed using the survey platform Qualtrics (qualtrics.com/uk). The participants were paid at a rate of £6.30/h. We sought the participants' consent and complied with the local data protection guidelines. The study was approved by The Open University UK's Human Research Ethics Committee with reference number HREC/3063/Uprety.



(a) Design for disjunction question



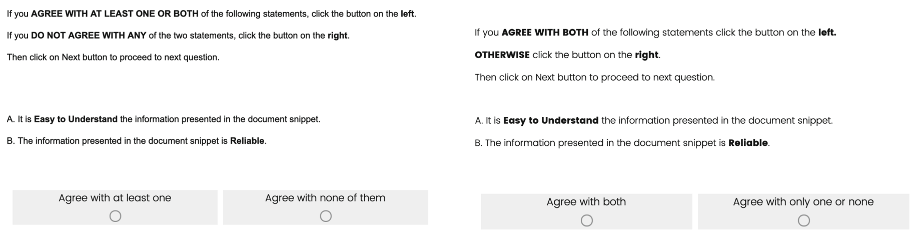
(b) Design for conjunction question

Fig. 3. Experiment design

Statement Pairs
A. It is Easy to Understand the information presented in the document snippet. B. The information presented in the document snippet is Reliable.
A. It is Not Easy to Understand the information presented in the document snippet. B. The information presented in the document snippet is Reliable.
A. It is Easy to Understand the information presented in the document snippet. B. The information presented in the document snippet is Not Reliable.
A. It is Not Easy to Understand the information presented in the document snippet. B. The information presented in the document snippet is Not Reliable.

Fig. 4. Four pairs of statements for conjunction and disjunction questions

We use the same set of three query-document pairs for our experiment as used in [19], as we have reused some of their data. Each participant was shown the three queries (and the documents) and were asked to judge the topicality of the document and one of the eight questions (so we obtain probabilities like $P(U+\vee R+|T+)$, etc.) Thus the participants can be said to be divided into eight groups for a between-subjects design.



(a) Design for disjunction question (b) Design for conjunction question

Fig. 5. Conjunction/disjunction question design

5 Results and Discussion

5.1 Violation of Kolmogorov Probability Axiom

The probabilities of conjunction and disjunction of the Understandability and Reliability questions are reported in Fig. 6. In order to compute the δ reported in Eq. 4, we also need the two probabilities related to single questions $U+$ and $R+$, apart from the conjunction and disjunction probabilities. These single question probabilities are obtained from the results in [19] (listed in Fig. 2). Then, we calculate $\delta = P(U \pm \vee R \pm |T+) + P(U \pm \wedge R \pm |T+) - P(R+|T+) - P(U+|T+)$. In Fig. 6 we see that δ is different from zero for all the three queries, although according to classical probability we expect that δ would be zero in all cases. Eq. (7), based on the projection operators in quantum probability, gives predictions of δ, as are shown in the last column of the table.

The violation of classical probability is a result of non-commutative structure of operators for U and R. As we can see, if operators of U and R commute with each other, the quantum correction term in the Eq. (7) approaches zero (the commutator is zero). In fact, the probability values obtained may violate some of the other basic axioms of classical/Kolmogorovian probability. For example, for Query 2, we can see that $P(U{-}{\wedge}R{+}|T{+}) = 0.414$ and $P(U{-}|T{+}) = 0.198$ which clearly violates $P(A, B) < P(A)$. Also, for this query, $P(U{-}{\wedge}R{-}|T{+})$ is greater than both $P(U-|T+)$ and $P(R-|T+)$. This type of violation has been termed as conjunction fallacy in the cognitive science literature [18]. Quantum models have been previously used to explain such violation [3] where the fundamental notion of incompatibility in judgements is identified as the potential cause.

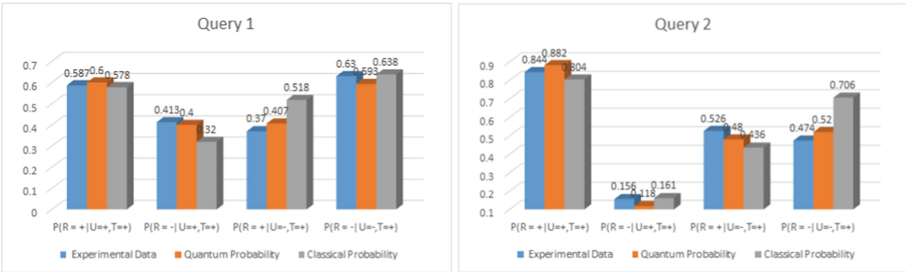
Query	P(U ∨ R I T =+)	P(U ∧ R | T = +)	δ	D(U,R)
Query 1	P(U+ ∨ R+|T=+) = 0.641	P(U+ ∧ R+|T=+) = 0.308	-0.175	-0.124
	P(U- ∨ R+|T=+) = 0.826	P(U- ∧ R+|T=+) = 0.393	0.251	0.032
	P(U+ ∨ R-|T=+) = 0.774	P(U+ ∧ R-|T=+) = 0.241	-0.017	-0.032
	P(U- ∨ R-|T=+) = 0.656	P(U- ∧ R-|T=+) = 0.435	0.215	0.124
Query 2	P(U+ ∨ R+|T=+) = 0.792	P(U+ ∧ R+|T=+) = 0.692	-0.049	-0.533
	P(U- ∨ R+|T=+) = 0.714	P(U- ∧ R+|T=+) = 0.414	0.199	0.071
	P(U+ ∨ R-|T=+) = 0.692	P(U+ ∧ R-|T=+) = 0.321	-0.058	-0.071
	P(U- ∨ R-|T=+) = 0.536	P(U- ∧ R-|T=+) = 0.368	0.437	0.533
Query 3	P(U+ ∨ R+|T=+)= 0.943	P(U+ ∧ R+|T=+) = 0.700	0.02	-0.623
	P(U- ∨ R+|T=+) = 0.562	P(U- ∧ R+|T=+) = 0.234	0.127	0.331
	P(U+ ∨ R-|T=+) = 0.907	P(U+ ∧ R-|T=+) = 0.378	-0.046	-0.623
	P(U- ∨ R-|T=+) = 0.535	P(U- ∧ R-|T=+) = 0.283	0.354	0.331

Fig. 6. Probabilities for conjunction and disjunction questions and associated violation from Kolmogorovian probability

5.2 Comparison of Quantum and Classical Probability Predictions

Figure 7 shows a comparison between quantum and classical probabilities with the experimental data for first two queries. The data for Query 3 had many probabilities close to 0 (see Fig. 2) and hence the sample became too small for a meaningful comparison. The probabilities are calculated for prediction of judgement of Reliability given the participant has judged Understandability and Topicality (positively), using equations derived in Sect. 2.1. Bayesian probabilities, in some cases, are significantly different from experimental data ($P(R+|U-, T+)$ for query 1 and $P(R-|U-, T+)$ for query 2). Quantum probabilities are consistently closer to the experimental data.

The Bayesian probabilities, as mentioned earlier, are based on the chain rule $P(R+, U+, T+) = P(R + |U+, T+)P(U + |T+)P(T+)$. The fundamental assumption here is that the variables corresponding to R, U and T can be jointly measured. In terms of the judgement process, this implies that a user can jointly consider information regarding the Reliability, Understandability and Topicality of a document with respect to the query. The incompatibility revealed in [19] and the order effects shown in [1] suggest that this is not always the case in general. Therefore we see Bayesian predictions deviate from the experimental data. As the quantum probability theory based on the Hilbert space model is free from this assumption of compatibility, it provides a promising alternative model that gives predictions closer to the experimental data. In fact, the modelling of incompatibility of different judgement perspectives forms one of the pillars of the Quantum Cognition research framework.



(a) (b)

Fig. 7. How quantum and classical probabilities compare with the experimental data for query 1 and query 2

6 Implications for IR

Quantum models can capture richer cognitive interactions, by way of generalising some of the constraints of classical models like commutativity. Here we discuss a few cases where our findings can inform the design of IR systems and algorithms.

The impossibility of jointly modelling Reliability and Understandability (which leads to the Kolmogorovian axiom violations) can be attributed to the fact that humans make decisions in a sequential manner and consideration of one dimension affects the judgement of the next dimension. Therefore, different orders of consideration of dimensions would lead to different final relevance judgements, *making the order a factor in the variability of relevance judgements by users*. When using an IR system to perform a task or make an important decision, there might be a particular order of dimensions which can lead the user to make an optimal decision. For example, for a health related query, a user might find a document difficult to understand, which may affect his or her judgement of Reliability and hence the overall relevance. However, if another user first judges reliability and finds it highly reliable, the judgement of understandability might be different. The IR system can help users to consider the optimum sequence of dimensions and thus maximise the utility, by providing extra information. For example, if the system can also provide information about the Reliability of the document in terms of a Reliability score or ratings by other users, it can reduce uncertainty in judgement and thus minimise the influence of judgement of other dimensions. Thus, for the given medical document, the low understandability might not affect the perception of Reliability.

Secondly, quantum probabilistic models can replace Bayesian models used in IR algorithms for ranking and evaluation. For example, in [13], a multidimensional evaluation metric is proposed where the gain provided by a document is written as a function of the joint probability of relevance with respect to different dimensions, e.g. $P(T, U, R, ...)$. Similar assumptions have also been made in [12,25]. For documents exhibiting incompatibility between different dimensions, predictions from such a model will be inaccurate. A probabilistic model based on non-commutative operator algebra, accounting for the incompatibility between different dimensions, needs to be considered.

Finally, these results of violation of classical probability theory calls for further user behaviour experiments to be conducted in IR that further exploit the Quantum-like Structure in human judgements. It would require novel experimental protocols like that of Stern-Gerlach, Double-slit experiment, etc., to generate data beyond the modelling capacity of classical probability theory. Such experiments in themselves might lead us to new insights into user behaviour in IR and information based decision-making in general.

7 Conclusion

Extending a quantum-inspired experiment protocol, in this work, we begin with the hypothesis that the multidimensional property of relevance has an underlying quantum cognitive structure which can be shown as violation of certain classical (Kolmogorovian) probability axioms. A particular experimental design is reported which can exploit the quantum cognitive structure. The data shows violation of one of Kolmogorovian probability axioms. We further show that quantum probability theory is a better alternative to model multidimensional relevance judgements than its classical counterpart, i.e. Bayesian model. Finally, we highlight important implications of our research findings to the design of IR algorithms system and user experiments.

Acknowledgements. Authors affiliated to the universities in UK, Italy and China are funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 721321, National Key Research and Development Program of China (grant No. 2018YFC0831704) and Natural Science Foundation of China (grant No. U1636203). Authors affiliated to QUT, Australia are supported by the Asian Office of Aerospace Research and Development (AOARD) grant: FA2386-17-1-4016.

References

1. Bruza, P., Chang, V.: Perceptions of document relevance. Front. Psychol. **5**, 612 (2014). https://doi.org/10.3389/fpsyg.2014.00612
2. Busemeyer, J.R., Bruza, P.D.: Quantum Models of Cognition and Decision, 1st edn. Cambridge University Press, New York (2012)
3. Busemeyer, J.R., Pothos, E.M., Franco, R., Trueblood, J.S.: A quantum theoretical explanation for probability judgment errors. Psychol. Rev. **118**(2), 193–218 (2011). https://doi.org/10.1037/a0022542
4. Clemmensen, M.L., Borlund, P.: Order effect in interactive information retrieval evaluation: an empirical study. J. Doc. **72**(2), 194–213 (2016). https://doi.org/10.1108/JD-04-2015-0051
5. Cool, C., Belkin, N.J.: Interactive information retrieval: history and background. Facet, 1–14 (2011). https://doi.org/10.29085/9781856049740.003
6. Eisenberg, M., Barry, C.: Order effects: a study of the possible influence of presentation order on user judgments of document relevance. J. Am. Soc. Inf. Sci. **39**(5), 293–300 (1988)
7. Fell, L., Dehdashti, S., Bruza, P., Moreira, C.: An experimental protocol to derive and validate a quantum model of decision-making. In: Proceedings of the 41st Annual Meeting of the Cognitive Science Society (COGSCI 2019) (2019)
8. Hogarth, R.M., Einhorn, H.J.: Order effects in belief updating: the belief-adjustment model. Cognitive Psychol. **24**(1), 1–55 (1992). https://doi.org/10.1016/0010-0285(92)90002-j
9. Huang, M.H., Wang, H.Y.: The influence of document presentation order and number of documents judged on user's judgments of relevance. J. Am. Soc. Inf. Sci. Technol. **55**(11), 970–979 (2004). https://doi.org/10.1002/asi.20047
10. Khrennikov, A.: Basics of quantum theory for quantum-like modeling information retrieval. In: Aerts, D., Khrennikov, A., Melucci, M., Toni, B. (eds.) Quantum-Like Models for Information Retrieval and Decision-Making. SSTEAMH, pp. 51–82. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25913-6_4
11. Kolmogorov, A.N.: Foundations of the Theory of Probability. Martino Fine Books, Eastford (2013)
12. Palotti, J., Goeuriot, L., Zuccon, G., Hanbury, A.: Ranking health web pages with relevance and understandability. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 2016, pp. 965–968. ACM, New York (2016). https://doi.org/10.1145/2911451.2914741
13. Palotti, J., Zuccon, G., Hanbury, A.: MM: a new framework for multidimensional evaluation of search engines. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management CIKM 2018, pp. 1699–1702. ACM, New York (2018). https://doi.org/10.1145/3269206.3269261
14. Pothos, E.M., Busemeyer, J.R.: A quantum probability explanation for violations of 'rational' decision theory. Proc. R. Soc. B: Biol. Sci. **276**(1665), 2171–2178 (2009). https://doi.org/10.1098/rspb.2009.0121
15. Sakurai, J.J., Napolitano, J.: Modern Quantum Mechanics. Cambridge University Press, Cambridge (2017)
16. Trueblood, J.S., Busemeyer, J.R.: A quantum probability account of order effects in inference. Cognitive Sci. **35**(8), 1518–1552 (2011). https://doi.org/10.1111/j.1551-6709.2011.01197.x

17. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. Science **185**(4157), 1124–1131 (1974). https://doi.org/10.1126/science.185.4157.1124
18. Tversky, A., Kahneman, D.: Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol. Rev. **90**(4), 293–315 (1983). https://doi.org/10.1037/0033-295x.90.4.293
19. Uprety, S., Dehdashti, S., Fell, L., Bruza, P., Song, D.: Modelling dynamic interactions between relevance dimensions. In: Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval - ICTIR 2019. ACM Press (2019). https://doi.org/10.1145/3341981.3344233
20. Uprety, S., Song, D.: Investigating order effects in multidimensional relevance judgment using query logs. In: Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval ICTIR 2018, pp. 191–194. ACM, New York (2018). https://doi.org/10.1145/3234944.3234972
21. Vourdas, A.: Probabilistic inequalities and measurements in bipartite systems. J. Phys. A: Math. Theor. **52**(8), 085301 (2019). https://doi.org/10.1088/1751-8121/aafe97
22. Wang, B., Zhang, P., Li, J., Song, D., Hou, Y., Shang, Z.: Exploration of quantum interference in document relevance judgement discrepancy. Entropy **18**(12), 144 (2016). https://doi.org/10.3390/e18040144
23. Wang, Z., Busemeyer, J.R.: A quantum question order model supported by empirical tests of an a priori and precise prediction. Topics Cognitive Sci. **5**(4), 689–710 (2013)
24. Xu, Y., Wang, D.: Order effect in relevance judgment. J. Am. Soc. Inf. Sci. Technol. **59**(8), 1264–1275 (2008). https://doi.org/10.1002/asi.20826
25. Zuccon, G.: Understandability biased evaluation for information retrieval. In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 280–292. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_21