Investigation and Modelling of Quantum-like User Cognitive Behaviour in Information Access and Retrieval

Thesis

Investigation and Modelling of Quantum-like User Cognitive Behaviour in Information Access and Retrieval

By

Sagar Uprety

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Computing and Communications, Faculty of STEM
The Open University, Milton Keynes, United Kingdom

AUGUST 2020

Supervisors: Dawei Song, Yijun Yu, Shailey Minocha

ABSTRACT

This thesis is fundamentally about using conceptual and mathematical constructs from the area of Quantum Theory in Information Retrieval (IR). The need and motivation for this is two-fold – firstly, it has been increasingly shown in decision sciences that human decision-making does not always conform to the norms of traditional probability and logic framework. The quantum framework offers a generalised probability and logic framework which can model decisions or judgements under dynamic context and ambiguity. Secondly, there is a need in IR for theories and models which improve our understanding of user behaviour. Hence it is worth exploring the combination of the quantum framework and IR, especially focused on the user aspects of IR. The overarching research question is whether there is evidence of user behaviour in IR scenarios which warrants the need for a quantum based approach by way of showing the limitations of the traditional (classical) approach. The methodology involves analysing data to detect quantum-like phenomena of interference, contextuality, incompatibility, etc. from two common data sources in IR – standard datasets like query log data, and through crowdsourced user studies designed similar to some quantum physics or cognitive science experiments. While the evidence of quantum-like phenomena from standard datasets is not convincing, we find that some of the user studies reveal the quantum-like structure of document judgements. One of the key findings which has implication for IR is the dynamic interactions between the different dimensions of relevance. For example, a user's judgement of reliability of a document depends significantly on whether they found it understandable or not. Thus, the consideration of one relevance dimension or document feature can provide a context for another dimension, contrary to the current IR models which consider these features to be independent of each other and an objective property of the document. The quantum framework has been especially designed to deal with such scenarios where properties of systems or objects do not exist independent of measurement context. The thesis concludes with suggestions about incorporating quantum mathematical constructs into state-of-art IR algorithms.

Acknowledgements

I would like to firstly thank my principle supervisor, Professor Dawei Song in his support throughout the three years of my PhD. Right from the giving me the opportunity to work in this exciting area of research, to constantly encouraging me with my ideas and providing freedom to develop them.

I would also like to thank my supervisors Dr. Yijun Yu and Professor Shailey Minocha for their help and inputs whenever I needed them and also for the great discussions with them on research related things.

I am grateful to the faculty at the Open University's School of Computing and Communications for organising the PG forum for valuable research insights, the AI and NLP Reading group where I had some great discussions with Dr. Alistair Willis and Dr. Paul Piwek, who was also my third party monitor. I am also thankful to Professor Arosha Bandara for his general help and support. I also learned a lot from my fellow PhD students in the school.

I am also grateful to QUARTZ project consortium for the many project meetings all over Europe and all of my colleagues, especially Dimitris Gkoumas, Prayag Tiwari and Amit Kumar Jaiswal, with whom I spent the most time travelling and talking about things research and beyond.

I am highly indebted to my collaborators and secondment hosts at QUT, Brisbane - Professor Peter Bruza, Dr. Shahram Dehdashti and Lauren Fell. I had some great discussions about Quantum Cognition and learned a lot during my visits to Australia.

Lastly, nothing would have been possible without the constant help and support of my family. I am indebted to my parents - Ija and Babba, for two of the most important qualities for a researcher and in life in general - honesty and curiosity. My younger brother Rohan, who inspired me to strive to become someone he can look up to. And finally my wife, Ayushi, who has stood by me throughout these difficult times of the Covid-19 pandemic, being a companion and constant pillar of support.

AUTHOR'S DECLARATION

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at the Open University or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at the Open University or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

SIGNED: ....Sagar Upreti.................. DATE: ....31/08/2020....

Table of Contents

List of Tables

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 The gist of Information Retrieval

Information Retrieval (IR) is the process of finding information in forms of text, images, items, profiles, etc. which is relevant to the need of a user. A user's information need is usually expressed in terms of a query. An IR system's task is to find documents relevant to the query. In the present day, its easier to talk of and explain IR in terms of Internet search engines which are frequently used by billions of people all over the world. Web search is one of the widely researched areas of IR. Another area in IR is Medical IR - retrieving information from medical documents, reports, etc. In legal IR, information is retrieved from large corpus of legal documents, e.g. to find precedence to lawsuits. Apart from these domains, there are also vertical search engines - searching for an item in an e-commerce website, searching for a friend on a social networking website, or searching for files in your desktop. Traditionally, the most focused area of research in IR has been to improve the matching algorithms between documents and queries and rank the documents in decreasing order of matching scores. As the similarity matching techniques have become better over these years, focus of IR research has shifted from representation and matching to other factors in IR, especially to the user. The difference between IR and a database retrieval lies in the specificity of the information need. In a database retrieval, an information need is explicit and it can be precisely translated into a query. IR is essentially a user-oriented task and in general the information need of the user cannot be precisely translated to a natural language query. Most of the time the user does not know how to formulate the appropriate query, and the information need is only partially specified.

Consider this example - Its a Sunday and all libraries in the city of Milton Keynes are closed. I want a place to sit down and finish my report. I like working in an environment which is quiet enough yet there are some people working around, so that I am motivated to focus on work. How do I find the perfect place to complete my work? I open up a search engine, but I am not quite sure how to proceed. I have something in my mind to accomplish but how should I convert it into a succinct query? Or even an elaborated query? I start with something like "Working spaces near me which are open on a Sunday" and see results about co-working spaces. I acquire the information about co-working spaces by going through the documents and judge them as not relevant to my need, since they are meant for small companies. I refine my query as "Individual Working spaces near me which are open on a Sunday". Although the results do not seem to be about the topic of my information need, I find one document from a very credible source. In the past too I have found relevant information from this blog. So, I open this blog, believing that it might contain relevant information. Reading several documents it appears that mostly likely I would have to go to London to work, so I also search to see if there are any libraries open in London on Sundays. Now I start getting some relevant documents, as I find not only the libraries which are open on Sundays but articles about Study spaces in London, which also include names of certain cafes where there are desk spaces available. Now I am sure of what I am looking for, and start searching about the specific cafes and libraries to know more about them - location, timings, etc. After narrowing down one place, I proceed to find the ways of getting there and book tickets.

From the above example we see that IR is in general an exploratory task which involves information discovery. Initially I was not thinking specifically about libraries and cafes, but as I read more documents, my search path meandered towards the correct choices.

There are a few important points to be considered in the above example:

- User's **information need is in general underspecified** and difficult to express in natural language queries.
- The **information need evolves** as the search process goes on. Starting with looking for working spaces, I ended up looking for specific cafes and libraries.
- **Relevance Judgement is a multi-criteria decision-making process**. While looking for relevant information, my reasons for clicking a document were not only

whether it contains information pertaining to the topic of my need. Greater trust in the source of the document altered my expectations of its topic matching my need. There can be a number of different criteria I might have considered in deciding whether a document is relevant or not, e.g. reliability, novelty, etc.

- **Judgement of documents influences information need**. My subsequent search queries are influenced by the documents I judge for the current query. When I see documents about co-working spaces coming up, I refine my query for individual working spaces. On seeing that most working spaces are in London, I refine my query to include libraries. Finally I learn that certain cafes provide working spaces to individuals and thus my subsequent searches are about those cafes.

IR systems can be improved upon to reduce the steps needed to resolve an information need. For example, diversity search is an area of research in IR in which IR systems present diverse documents for the user IN. For the above example, documents about co-working spaces, libraries and cafes can all be shown on the same page. This will help the user make decision faster. It is a challenging task to identify the diverse categories of documents for a query. Another way to help users with better and faster decision making is for the IR system to understand them better. Having personalised knowledge about the user and the context - like the location, time, user interests, etc. can help in better results. User modelling is already a very popular topic among IR researchers. Thirdly, IR systems can do better by improved understanding of natural language. For example, if instead of the phrase "Working spaces", I had used "Study spaces", I would have got results about libraries and cafes in the first instance itself. A user is not always aware about the related terms which appear in documents. It is the task of the IR system to understand the meaning of the query and look for semantically similar terms to add to the query, or suggest improved queries.

1.2 Motivation for Using Quantum Theory to Model Human Judgements

In the previous section I have considered some challenging problems in IR. Various approaches are being used to tackle such issues including neural networks [47, 91, 114], reinforcement learning [81, 98, 145, 174] etc. My project is about improving IR systems, particularly user interactions using the mathematical framework of Quantum Theory (QT). Although QT originated in Physics as a way to model behaviour of microscopic particles, it is also a generalised theory of probability. It offers a new way of calculating probabilities of which the classical probabilities are a special case (Born's rule [20]). It also generalises the vector space models used in classical theory by using a complex vector space (also called Hilbert Space). The combination of these allows physicists to model unique properties of the Quantum world like Superposition, Quantum Interference and Quantum Entanglement. QT has been successfully applied to model and predict some normatively irrational and paradoxical human decision making [33, 34, 126–128, 152, 175]. Applications of QT to model human behavioural and cognitive data is now studied under the upcoming field of Quantum Cognition [32].

This thesis is neither about Quantum Physics, nor is it about Quantum Computing. Instead it seeks to employ some of the methodologies developed in the field of Quantum Cognition to model user judgements and decisions in IR. In the following points, I discuss the motivation for using QT to model behavioural and decision-making data [32]. This helps to give an intuition as to why QT can be a useful candidate for data and behavioural modelling in IR.

- **Human judgements are based on indefinite states**

  According to the commonly used models used in decision making, such as Bayesian networks or Markov models, a system changes state from moment to moment, but any given point of time, it is in a definite state with respect to some judgement. For example, suppose one is a jury member tasked to decide whether a person is guilty or not. One has heard conflicting evidence and based on that one has to make the decision. Suppose the jury member's decision fluctuates between guilty and not guilty as he or she weighs the evidence. Current classical models of decision making will assume that the jury member is in an definite state at each moment of time. However, due to the complexity of determining it precisely, a probability is assigned to each state at each moment in time. So for example, at a certain moment in time, there might be a 80% chance that the jury member thinks that the person is guilty and a 20% chance that the person is not guilty. But it is definitely one or the other.

  A Quantum probabilistic model works differently, in that it allows for an indefinite state to exist. In QT, it is called a Superposition state and the jury member is not in a state which is either definitely guilty or definitely not guilty. Instead, he or she is in a confused or ambiguous or conflicted state and at any moment in time does not necessarily think that the person is guilty or not guilty. Both potentialities exist at the same time. The potential (probability) for guilty maybe more than the potential for not guilty, and these potentials may change over time. In the classical probabilistic modelling, the cognitive system follows a trajectory of definite states across time, like a point particle following a definite, measurable path. In the Quantum model, it is not a single path but rather a wave travelling across time. However, when a decision is finally made, the cognitive state collapses to a definite state of guilty or not guilty.

It seems as the wave nature of cognitive states better resembles the experiences of conflict, ambiguity and confusion. The particle like modelling captures the state of conflict resolution, decision and certainty. However, for a major part of the decision making journey, a person is in the indefinite state and therefore modelling of these indefinite states can lead to better decision predictions.

In IR, before a user judges a document, the user is in a superposition state of relevance and non-relevance with respect to the document. Only once the judgement is made, does the state become definite. The advantage of using a superposition state of decisions is that it can be used to model ambiguity in making the decisions, when a user is not certain of the judgements.

- **Constructive Nature of Judgements**

In the classical probabilistic models of definite decision states - which keep on changing every moment - what is recorded at a given moment reflects the state of the system immediately before the measurement. For example, suppose after watching a thrilling scene of a movie, a person is asked "Are you afraid", or "Are you excited". Then the answer to the question reflects the state of thought just before each of these questions. In Quantum Theory, where systems exist in a superposition of all possible states, the definite state is obtained only on measurement of the system, as the state collapses to one of the potentialities. Immediately before measurement, the Quantum system is in an indefinite state, instead of a definite state which a classical model assumes. For example, a person may have ambiguous feelings after watching the disturbing scene, but when asked the question "Are you excited", the cognitive state becomes more definite. He or she is able to answer "Yes, I am" or "No, not at all". Thus the answer is constructed from the interaction of the indefinite state and the question (measurement). This is also in line with the psychological theory of preference construction [146]. According to this theory, preferences or beliefs or choices are not simply read off some master list but are constructed on the spot by an adaptive decision maker.

Consider the following scenario taken from [159]: "the discussion of the meaning of preference and the status of value may be illuminated by the well-known exchange among three baseball umpires. "I call them as I see them," said the first. "I call them as they are," claimed the second. The third disagreed, "They ain't nothing till I call them." Analogously, we can describe three different views regarding the nature of values. First, values exist-like body temperature-and people perceive and report them as best they can, possibly with bias (I call them as I see them). Second, people know their values and preferences directly-as they know the multiplication table (I call them as they are). Third, values or preferences are commonly constructed in the process of elicitation (they ain't nothing till I call them)." The superposition and measurement principles of QT are analogous to this third view of preference as a constructive, context-dependent process.

As mentioned earlier, many times the user information need (IN) is also not definite. It can be assumed to be in superposition state of many different INs. As the user goes on interacting with documents, it becomes more and more specific. This contextual evolution of user IN can be modelled using the Quantum framework and predict user IN for improved ranking of documents.

- **Judgements are influenced by each other**

  Consider the following question asked to a young adult - "Are you happy?", the typical answer is "Yes, everything is fine." However, if the question asked is "When was the last time you dated someone?" and if the answer is "Many years ago", now asking the question about how happy they are may not produce the same answer. This is because answering the 'date' related question has changed the context of the second answer. There are many such examples where judgements are influenced by other judgements. For example, if a man has to buy a car and his choice is a BMW while his wife's choice is a Mercedes. So he is definite about his choice but when he considers what his wife likes, then he comes into a state of confusion about which of the brands to buy - an uncertain or indefinite state. The question about wife's preference of brand disturbs the man's own judgement and creates an uncertain state. In Quantum Theory, there is the concept of incompatibility of observables, where it is not possible to construct a joint probability distribution for the variables and we can only assign probabilities to sequence of measurements. It is impossible to be in a definite state with respect to two incompatible questions, because the definite state for one is an indefinite state for another. The Uncertainty Principle is a famous example of this phenomena from Quantum Theory. The position and momentum of an electron are two incompatible properties. They cannot be jointly measured.

This phenomena can be very easily seen in the context of IR, as judgement of a document is influenced by what other document the user is seen. This also means that relevance of a document is constructive, because if the context of surrounding documents influence relevance, in what way can it be said to be defined independently? Incompatibility can also manifest in the individual factors contributing to relevance - e.g. topicality, reliability, novelty, understandability, etc. - also called different dimensions of relevance. Reliability of a document may be influenced by whether a user finds it understandable or not, or whether it is newly produced or not.

These were some of the examples of human judgements and decision making which can be inherently captured using the framework of Quantum Theory. As discussed in the previous section, Information Retrieval is essentially a decision making process from the user perspective. This makes it quite an exciting task to see if we can apply the Quantum probabilistic model in Information Retrieval and related user-centric areas.

1.3 Quantum Theory and Information Retrieval - Hypothesis and Main Motivations of the thesis

The most common approach in IR for measuring relevance stems from the Cranfield paradigm [42] which typically involves assigning a fixed relevance label to documents as judged by a majority of a small sample of users. This paradigm is also present in its modern form in TREC evaluations. It comes under the system view of relevance. That is, the relevance of a document is pre-fixed, irrespective of what a new user might judge it to be. The user and its context is therefore removed from relevance modelling. Considering relevance from each user's point of view makes it difficult to measure and predict due to varying cognitive factors associated with human judgements. The same document may be relevant for one user and irrelevant for another, for the same query or information need. In the arguments to follow, I consider relevance from the user point of view and formulate research questions linking relevance to contextuality.

I hypothesise that the nature of relevance of documents to queries is similar to that of quantum systems. One cannot pre-assign values of relevance (i.e. relevant, not relevant, etc.) to documents and they are realised only when a user interacts with a document. The underlying intuition for this is the argument that cognitive processing of information maybe subject to ambiguity. Here ambiguity needs to be carefully explained. Document judgements may manifest ambiguity from two point of views - System ambiguity and user ambiguity. System ambiguity is when a system cannot predict whether a user would judge a document relevant or not due to incomplete context. As the system gains more information about the user and its context, the predictions become more accurate. This is similar to the hidden variables concept. Lack of knowledge of all possible hidden variables introduces uncertainty in the predictions and once the hidden variables are known, the system can accurately predict user judgements. For example, for the query 'Things to do near me', a document about an activity won't be relevant if it is an outdoor activity and the forecast is of heavy rain. For another location with clement weather, the same document will be more relevant to the query. Assume that users of the system who judge the document are uniformly split between the two locations. If information about the location and its weather is not available to the system, it may find that half of the users find a document relevant and the other half do not. The system is ambiguous about its prediction, with both types of judgements equally possible. It might be tempting to model such a document as in a superposition state with fifty percent probability that it will be judged relevant. But here is the catch - for the users, judgement of the document poses no ambiguity. Half of the users are certain that the document is relevant and the other half are certain that it is not relevant. Hence, even if we assume the query-document-user to be a quantum system and model document to be in a superposition state, the system will not manifest other quantum properties such as incompatibility and interference.

Let us consider the ambiguity in relevance judgement from the user point of view. Ambiguity will arise if the user cannot decide with certainty whether the document is relevant or not and both options seem almost equally likely. In common usage, the user can be said to be 'uncertain', 'conflicted', 'confused' or 'in two minds' about the document. It may be because the user lacks access to information needed to make the judgement (e.g. background knowledge) or because different aspects of the document offer conflicting signals regarding relevance (e.g. the document maybe highly topical but not reliable). The underlying user cognitive model is of a state which allows both relevant and non-relevant judgements to have potential for being expressed. This is an indefinite state because it does not have a fixed value. Classical probability theory assumes a definite state where each state is pre-defined and probabilities arise only due to incomplete information about the state. In the indefinite state, although both the relevance and non-relevance values can be expressed when the judgement takes place, the user can be more inclined towards one than the other. For example, the potentiality for judgement of relevance can be more than the judgement of non-relevance, say 80% versus 20%. In QT, these potentialities are modelled as probabilities of the two possible states. These probabilities can be empirically estimated as proportions obtained by measuring a large sample of identically prepared indefinite states. In the Hilbert space representation of a quantum system, states are represented as vectors in the Hilbert space. So, mutually exclusive states are represented as orthogonal vectors and the dimensionality of the Hilbert space is the cardinality of the set of possible states of the system. Probability of the system to be in a particular state on measurement is calculated as the square of the inner product of that state and the current system state.

Similarly, if a sample of a large number of IR system users with identical cognitive states (i.e., with same potentialities for the two relevance states to be expressed on judgement) are considered, we can obtain a distribution of judgements where 80% users judge the document relevant and 20% users judge it non-relevant. This is because we can expect the ambiguous users to randomly decide on one of relevance or non-relevance in the proportion of the potentialities expressed in their indefinite cognitive states.

Hence we can see how a user's cognitive state in information interaction can be modelled similar to a quantum system. The system under consideration here is not the user's cognitive state alone. It is the user along with the components of information interaction - query, context and document. Document judgement is analogous to measurement of the quantum system. The analogy also extends to the phenomena of collapse of the quantum system to a particular state on measurement wherein the system state does not change even after repeated measurements. Similarly, if a user judges a document as relevant/non-relevant or reliable/not-reliable, they will not change their judgement when asked to judge the same property again (Of course, they may if the collapse was a partial collapse, i.e. they were still not certain of the judgement. But for the moment I want to keep the model simple).

However, the real quantum advantage presents itself when two different properties are measured sequentially where the second measurement can change the outcome of the first measurement. This will be shown in the Stern-Gerlach experiment in Chapter 5. In the case of IR, this can manifest when a user judges two documents or considers two relevance dimensions to judge a document. Phenomenon like order effects are evidence that such measurements are incompatible and cannot be jointly measured. Incompatibility is forbidden in classical probability theory and quantum modelling of such scenarios can prove more effective in capturing the richness of interactions.

It is be noted that from a information interaction point of view, uncertainty in making a definite relevance judgement is important for quantum-like contextuality. With insufficient information accessible in the document (or a snippet in a list of results) to form a judgement, peripheral information tends to be used and it is in these cases that contextuality is most likely to be present, as judgements cannot be considered to be pre-determined. In contrast, query-document pairs with little to no uncertainty present may be expected to behave in a classical, deterministic manner. For example, consider the query, 'In which year was Mahatma Gandhi born', the documents containing the correct answer can be pre-assigned as relevant. This is because relevance in this case is tied to a single, clear fact that is easily identifiable in the document.

QT has been applied to IR since more than a decade now. However, most of the research has been from a system-oriented IR point of view. As I will discuss in the next chapter of literature review, researchers have utilised most of the quantum concepts of superposition, interference, entanglement, etc. to build representation and ranking models in IR. However, the biggest drawback of these models is their lack of interpretability. It is not clear from the models whether the underlying phenomena are quantum-like and whether there is a quantum advantage and need for using QT.

As mentioned before, QT was developed because of the need to explain and model certain experimental data which violated constraints of existing physical theories. Hence, before using QT to construct IR models, it is important to find whether the data warrants the need of models based on the quantum framework. This means looking for violations of certain classical constraints, e.g. axioms of classical probability or of boolean logic, etc. Such violations would mean that models built upon these, like Bayesian or Markov models are rendered insufficient for modelling such data. Also, when we would know that the data behaves in ways similar to those from quantum systems, it would be easier to explain the workings of the quantum-based IR models, interpret their predictions and justify their need and advantages. This has always been an issue with the current research in Quantum-inspired IR.

Lastly, it is easy to see in light of the above discussions that QT is best suited to model human cognitive states during decision-making, especially in presence of ambiguity and multi-criteria decision-making. Current research in Quantum-inspired

IR has also focused on modelling textual data (also other modalities). Such kind of data is deterministic in nature, hence one can argue whether it is possible that phenomena like superposition, interference exist in such data. However, these properties can come up when users interact with such data. Hence, in this thesis, user is placed at the centre of the IR process.

The above discussion forms the motivation for this thesis. The next section channels these motivations and intuitions into actionable research questions.

1.4 Research Questions

The overall aim of the thesis is to lay a foundation for QT based models in IR by gathering evidence for the need of such models. I firmly believe that this should be the first step before proceeding to build mathematical models. This thesis has the following general research question:

RQ: Is there evidence of quantum-like phenomena in user cognitive behaviour in IR?

To make the question clearer, I need to define some of the terms used in the question. Firstly, by ***evidence*** I mean empirical evidence which is obtained by carrying out experiments on IR data. This data is usually about document judgements labels (relevant vs non-relevant, etc.) available via either standard IR datasets and query logs or collected by myself using crowdsourced user studies. The evidence is in form of insufficiency of classical probability methods in modelling the obtained IR data. The term ***classical probability*** above means the probabilities calculated based on the Kolmogorovian axioms of probability [86]. Any method based on a set-theoretic representation of events, following Boolean logic, is a ***classical method***, including Bayesian models. Thus, most of the current IR ranking and representation models can be termed as classical.

Quantum-like phenomena refers to a set of phenomena unique to Quantum Theory (QT) which set it apart from classical physics. These are - superposition, contextuality, incompatibility and interference. Others like entanglement, etc. are not relevant from a cognitive and decision-making point of view. A similar word ***quantumness*** will be encountered throughout the thesis. Quantumness of a system or phenomenon or quantumness of data means that classical methods are not sufficient to model the system or phenomenon or data.

User cognitive behaviour refers to those set of user interactions with IR systems which involve cognitive processing for judgement or decision-making. For example, reconciling different criteria for judging a document such as credibility, novelty, etc., in order to decide its utility or relevance. The focus is on cognitive processing which may involve ambiguity and uncertainty, instead of user interactions like clicks, bookmarking, etc.

The answer to this question is sought by breaking it down into the following sub-research questions:
RQ 1: Can standard IR datasets or query logs provide evidence of incompatibility between judgement of different dimensions of relevance?

Multidimensional relevance [93, 182, 191] bears a lot of similarity with quantum properties and it is likely that quantum phenomenon of incompatibility manifests. Incompatibility is in-turn manifested in human decision-making as order effects which has been successfully modelled and predicted using the quantum framework [32]. However, most of these experiments have been using data from small-scale user studies. My first experiments for this thesis investigate whether it is possible to observe incompatibility from large scale query log data and if so, can it be utilised to build IR ranking models. The experiments in Chapter 3 answer this question.

RQ 2: How to verify quantumness of IR data using no-go theorems of Quantum Theory?

Since the beginning of the 20th century, a large number of experiments in physics showed that quantum systems violated laws of classical physics and that they could not be modelled using classical physics. However, it was many years later in the 1960's that QT was theoretically proven to violate classical laws using certain no-go theorems such as Bell's Theorem [17] and Kochen Specker Theorem [85]. These theorems were formulated as certain inequalities which were supposed to be always obeyed under classical laws. Quantum systems violated these inequalities and since then these inequalities have been used to demonstrate the quantumness of a system. Therefore, I also intend to use these inequalities for IR data, which provide a formal way to prove quantumness of the data generated by user cognitive information interactions. Experiments using these inequalities are performed in Chapter 4.

RQ 3: How to adapt existing experiments from quantum theory to study dynamic interactions between relevance dimensions so as to reveal quantum-like nature of user cognitive states?

As mentioned before, multidimensional relevance is analogous to those quantum properties (e.g. electron spin and photon polarisation) which exhibit incompatibility. The Stern-Gerlach (S-G) experiment in Physics [135] studies such properties to establish quantumness of these systems. In Chapter 5, I adapt the S-G experiment in IR scenarios to see whether interactions between different relevance dimensions give rise to quantum-like phenomena of incompatibility and interference. The S-G experimental protocol also helps in constructing representation and predictive models for multidimensional relevance.

RQ 4: Do the quantum effects observed in the interaction between relevance dimensions have any effect on the final decision of document relevance?

In most of the experiments performed so far, the focus is on the quantum phenomena generated through judgement of different relevance dimensions. In the last series of experiments, I investigate the impact of various quantum-like phenomena on the final judgement of relevance itself. In IR, modelling and prediction of relevance is more important than that of the individual relevance dimensions. Hence, any quantum-like phenomena observed should affect relevance judgements, otherwise they are of not much use in IR.

1.5 Main Contributions and Thesis Structure Overview

1.5.1 Main Contributions of the Thesis

The main contributions of this thesis can be summarised as follows:

- **A method for extracting Hilbert Space structure from query log data.** Hilbert space structure has been used in IR for smaller scale user study data or as toy examples [21, 65]. In Chapter 3, I devise a method to represent query log data in a Hilbert space with different bases for different relevance dimensions. Hilbert space is the building block in the quantum framework and lays the foundation for quantum probabilistic models for representation and prediction. It can also be used to further test for quantumness of IR data, e.g. by modelling order effects and interference in the data. The methods developed to model query log data into a Hilbert space representing multidimensional relevance judgements is a novel contribution of the thesis.

- **Method to formulate relevance judgement in the form of Bell-type inequalities.** Bell-type inequalities are a formal method to establish quantumness of systems or data generated by systems. Hence it is important that any claims of quantumness of systems are verified through these Bell-type inequalities. While cognitive scientists have explored ways to use Bell-type inequalities [7, 57] for other human decision-making data, it has not been used for IR data before. This thesis devises a way to formulate relevance judgement data both from query logs and from user studies in terms of Bell-type inequalities.

- **Complex-valued representation of user cognitive states.** Complex numbers are essential to QT and are instrumental in differentiating quantum probability from classical probability via the interference term [64]. Complex numbers have been previously used in quantum-inspired IR but have been heuristically defined [131, 180, 199]. In the experiments in Chapter 5, complex numbers naturally arise in the modelling of user cognitive states. They are also utilised in modelling incompatibility/order effects and in quantum probability calculations. This is the first experiment in IR where complex numbers are used to represent aspects of user judgements.

- **Strong evidence of quantumness in multidimensional judgements** The main research question of the thesis is to find evidence of quantumness in IR data. I find strong evidence of quantum-like phenomena of incompatibility and interference in judgements of different dimensions of relevance, like topicality, reliability and understandability. For example, the judgement of understandability might be interfered by a previous judgement of reliability. Or, the order in which these dimensions are considered changes the judgement of these dimensions. Hence there is evidence that the relevance dimensions are not objective property of documents, but rather the judgements are constructive and created at the point of consideration depending upon the context.

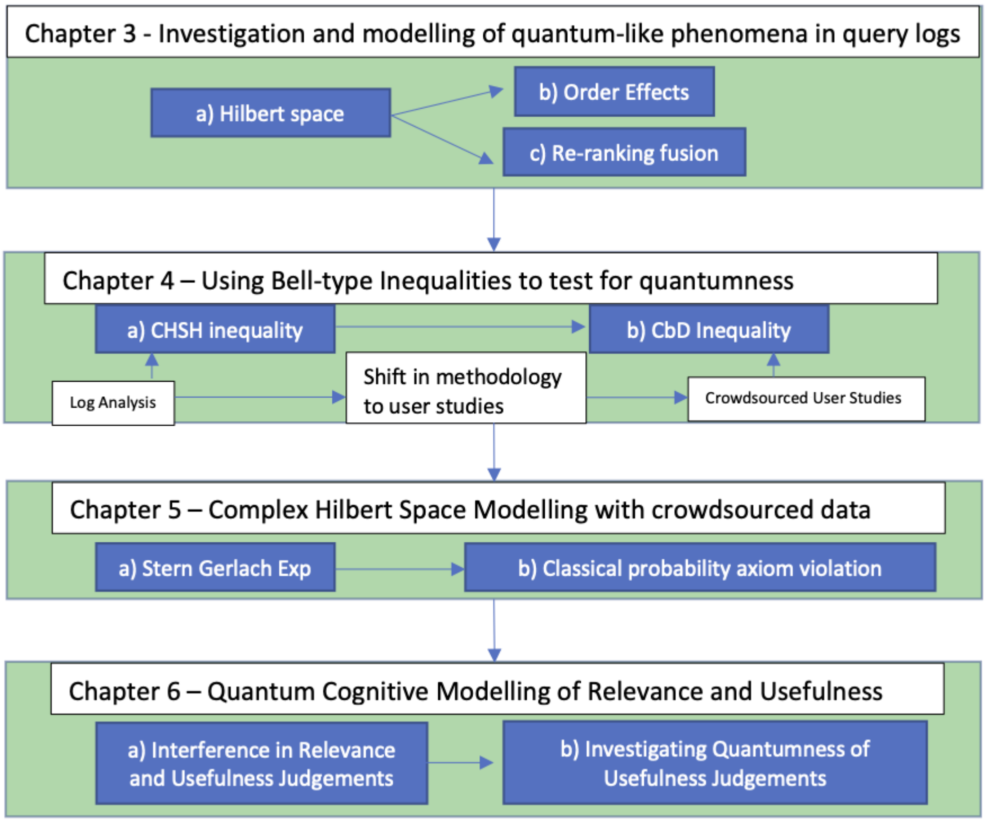


Figure 1.1: Thesis overview

1.5.2 Thesis Structure Overview

Figure 1.1 shows an overview of the experiment based chapters of the thesis. In Chapter 3, I explore query log data which is a commonly used source of user behaviour data in IR. I start with constructing Hilbert Space representations of a document with different basis representing relevance in different dimensions. I utilise these representations to model Order Effects in judgement of a pair of documents. Lastly, I combine the Hilbert Space representations and the quantum concept of collapse of a state to re-rank documents in session search. While the Hilbert Space structure of multidimensional relevance offers a novel representation which combines document properties with user cognitive state, the evidence of incompatibility as order effects is weak. Thus I conclude that I need a more principled way of detecting quantumness in IR data. This chapter answers research question **RQ 1**.

In Chapter 4, I move towards using Bell-type inequalities as a go/no-go tool for detecting the quantumness of data. I use the same query log datasets as in Chapter

3 and integrate the Hilbert Space representation into the CHSH inequality to check whether it is violated (which is a sign of quantumness). However, I do not find any violation of the inequality. Careful thought reveals that one of the reasons it is not violated is because of the lack of context in the data. While query logs contain various user interaction data such as query modification and clicks, it does not sufficiently capture the different contexts in which users judge documents. Here, my research takes a significant turn and I decide to use crowdsourced user studies to collect my own data. This gives us the opportunity to collect contextual data using particular experiment designs. Crowdsourcing gives us a larger sample of data than lab based studies so that I can carry out statistical analysis. Using the crowdsourced data, I study the violation of another Bell-type inequality - the Contextuality-by Default (CbD) inequality. This inequality is a modified CHSH inequality for decision-making experiments. While there is a violation of the inequality found, I discuss some recent developments pointing to some limitations in the modification assumptions of CbD. Therefore, I proceed to find other ways to test the quantumness of crowdsourced data. This chapter answers research question **RQ 2**.

In Chapter 6, complex-valued multidimensional Hilbert Space models are constructed using crowdsourced data. As CHSH and CbD inequalities do not provide direct ways to represent and quantify the quantum-like structure of cognitive states which I aim to model in this thesis, I turn to the Stern-Gerlach protocol of Physics to construct quantum states. Cognitive analogue of the Stern-Gerlach experiment of Quantum Physics is utilised to design user studies. Now we start seeing more evidence of quantumness - presence of complex phase, negative values in Wigner function, non-commutative operators. I decide to test another inequality to further test the quantumness. This time, the violation of a fundamental classical probability axiom is examined- violation occurs. This chapter answers research question **RQ 3**

Having found evidence of quantum-like phenomena existing in user judgements of different dimensions of document relevance, the question to ask if whether the evidence is useful for IR systems. For example, the presence of incompatibility in judgement of different relevance dimensions help us in predicting document relevance or does it effect the perception of usefulness of documents. This is investigated in Chapter 7, which focuses on quantum phenomena of interference of relevance dimensions on final judgement of document relevance or usefulness. Research question **RQ 4** is answered in this chapter.

Chapter 7 concludes the thesis by discussing the key findings of the research, implications to IR and future work which can be build upon it.

1.6 Origins

The following publications co-authored by me form the basis of the chapters in this thesis:

- Chapter 3 is based on:
  - Uprety et.al [165]: "Modeling Multidimensional User Relevance in IR Using Vector Spaces. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval"
  - Uprety et.al [164]: "Investigating order effects in multidimensional relevance judgment using query logs. In Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval"
- Chapter 4 is based on Uprety et.al [163]: "Investigating Bell Inequalities for Multidimensional Relevance Judgments in Information Retrieval. In Quantum Interaction. Springer International Publishing, Cham, 177–188"
- Chapter 5 is based on:
  - Uprety et.al [161]: "Modelling dynamic interactions between relevance dimensions. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval"
  - Uprety et.al [166]: "Quantum-Like Structure in Multidimensional Relevance Judgements. In Advances in Information Retrieval. Springer International Publishing, Cham, 728–742"

CHAPTER 2

Background and Literature Review

In this chapter, concepts relevant to the research topic are defined and briefly explained, with appropriate citations where needed. It starts with a short overview of IR - its definition, methodologies and current challenges. A brief introduction to Quantum Theory is provided. The review also discusses how and why the framework of Quantum Theory is useful for mathematical modelling in general. Finally, I survey the research done in application of the Quantum framework in IR and related areas.

2.1 Introduction to Information Retrieval

Information Retrieval (IR) is finding material (documents, videos, audio, etc.) of an unstructured nature that are relevant to an information need of the user, from within large collections (of document, videos, images, etc.). Information is also obtained or extracted from databases, however databases store and retrieve information in a structured manner. In a structured model of information extraction, a user has an exact knowledge of its information need. Such an information need can be exactly translated into a query, such that two users having the same information need will construct an identical query. The data to be retrieved is organized into related tables or hierarchical objects which makes it easier for artificial queries to extract it deterministically. On the other hand, Information Retrieval involves naturally occurring data in the form of articles, web-pages, images, etc. The need of the user is most of the times vague and is only approximated in terms of a query. This vagueness also results in the use of probabilistic models to rank the different information(documents) retrieved for a query.

2.1.1 Formal Concepts

The problem of IR can be formally defined using the following concepts:

Vocabulary: $V = \{w_1, w_2, ..., w_N\}$ The universal set consisting of all possible words to be considered by a user and an IR system.

Query: $q = q_1, ..., q_m$ where $q_i \in V$

Document: $d_i = d_{i1}, ..., d_{im}$ where $d_{ij} \in V$

Collection: $C = d_1, ..., d_M$

Relevant Documents for Query q: $R(q) \subseteq C$

IR task: Compute R'(q) which is an approximation of R(q)

The main task in IR is to compute this set $R(q)$. This can be broadly broken down in two steps (Figure 2.1):

1. **Document filtering**
   - $R'(q) = \{d \in C | f(d,q) = 1\}$, where $f(d,q) \in 0,1$ is a binary classifier.
   - The system must narrow down a set of relevant documents from the whole collection
2. **Document Ranking**
   - $R'(q) = \{d \in C | f(d,q) > \theta\}$, where $f(d,q)$ is a real number and θ is a cut-off determined by user actions.
   - All retrieved documents are not equally relevant as information need is not precise. Thus a ranking is preferred and an IR system needs to compute relative relevance of documents.

2.1.2 Retrieval Models

In IR, retrieval and ranking go together, due to the vagueness of the information need, as discussed in the previous section. The task is not only to retrieve relevant documents, but simultaneously rank them in decreasing order of relevance. Here I briefly discuss two most prevalent models in IR - one is the similarity model, of which the Vector Space Model is an example, and the other one is the Probabilistic model, of which the Language model is an example.

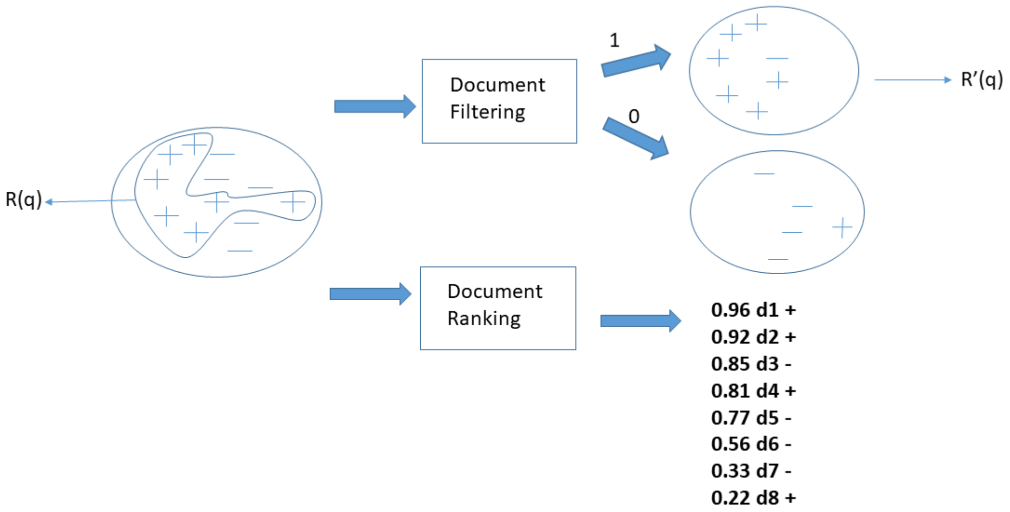


Figure 2.1: Basic IR task

2.1.2.1 Vector Space Model

A Vector Space Model(VSM) is a geometric model where each term(word or a phrase) of a query or a document is represented as a vector in a high dimensional space. Each term corresponds to a dimension of this space and the term vectors are orthogonal to each other. A query or a document is also defined as a vector in this space and can be expressed as a weighted sum of the term vectors $q = (x_1, x_2, ..., x_N)$ and $d = (y_1, y_2, ..., y_N)$ where x_i and $y_i \in R$. Relevance between query q and document d is estimated as the similarity between the vectors q and d. So, we have

$$f(q,d) = sim(q,d) = q.d = x_1y_1 + x_2y_2 + ... + x_Ny_N = \sum_{i=1}^{N} x_iy_i \tag{2.1}$$

There are many different ways to define the term weights for the query and document vectors. One can use one-hot vector approach where each element of the vector is either 0 or 1 and indicates the presence or absence of a term. Using it in the above formula would mean a simple counting of the number of terms which occur both in a query and a document. Another way is to use the term frequency(TF) [95, 96] for each term in the query or document. Besides counting the number of terms occurring together, TF gives more weights to terms which occur more frequently. However, the natural language comprises of many frequently used words and expressions, such as "the", "of", "in", etc. Using TF to create vectors will give higher weights to such terms in the documents, diluting its semantic representation. Thus, another parameter called the inverse Document Frequency(IDF) [82, 137] is used, which is the inverse of the number of documents of the collection a term occurs in. So matching a rare term will have a higher contribution to relevance than a commonly occurring term.

Thus combining the TF-IDF weighting scheme, the vector for a document looks like $y = c(w_i, d) * IDF(w_i)$. The IDF can be roughly expressed as $\frac{M}{k}$ where M represents the number of documents in the collection and k is the number of documents a term occurs in. Taking the logarithm helps smoothen the penalty on frequent terms and is found to be a better approach. The formula for calculating relevance between a query and a document is

$$f(q,d) = \sum_{i=1}^{N} x_i y_i = \sum_{w \in q \cap d} c(w,q)c(w,d)\frac{\log(M+1)}{k} \tag{2.2}$$

2.1.3 Probabilistic Models

2.1.3.1 The Probability Ranking Principle

The idea of ranking documents for a query was formalized in terms of the Probability Ranking Principle(PRP) in [133]. It essentially states that ranking of documents has to be done in the decreasing order of probability of relevance of documents. This is the optimal strategy of ranking with the following two assumptions:

- The utility of a document to a user is independent of the utility of other documents in the ranked list. That is, the judgement of a document is not influenced by other documents seen before.

- User browses the documents sequentially in the order of the ranked list

The PRP is extended in [67] as the Probability Ranking Principle for Interactive Information Retrieval(PRP-IIR) which focuses on the complete user interaction process and takes into account the costs of different activities like document judgment, query reformulation, etc.

2.1.3.2 BM25 Model

In addition to term weighting, document-length normalization is performed in the BM25 [134] model because documents with a large number of terms will most likely satisfy any query. It smoothens the TF for high frequency terms to avoid their dominance in the relevance calculation. The BM25F [132] and BM25+ [99] are further improvements over the BM25.

2.1.3.3 Language Model

The basic idea of a probabilistic retrieval model is that the ranking function $f(d,q)$ is estimated as a probability of relevance $P(R=1|d,q)$ where $R \in \{0,1\}$. Query likelihood model asks the question - "Given the user likes a document d, how likely it is to formulate the query q". Thus I seek the probability $P(q|d)$. Statistical Language Modelling [40, 125, 147] is one way to calculate these probabilities. A language model(LM) is a probability distribution over word sequences. The probability of a sequence of N words $P(w_1 w_2 .. w_N) = P(w_1) * P(w_2) * ... * P(w_N)$ (Unigram LM). So for a query comprising n words, $q = w_1..w_n$, $p(q|d) = p(w_1|d) * ... * p(w_n|d)$. Here d represents the document language model distribution and $p(w_i|d)$ is the likelihood that the word w_i is generated by the document model d. The ranking function uses the logarithm of the query likelihood:

$$
\begin{aligned}
f(q,d) &= \log(p(q|d)) \\
&= \sum_{i=1}^{N} \log(p(w_i|d)) \\
&= \sum_{w \in V} c(w,q) \log(p(w|d))
\end{aligned}
\tag{2.3}
$$

where $c(w,q)$ is the count of the word in the query. Now the retrieval problem narrows down to the estimation of $p(w_i|d)$, i.e., the document language model. Different estimation methods lead to different ranking functions. The most commonly used method to estimate the language model is the count based method, where the probability that a word is generated by a document is given by its relative frequency. This is the maximum likelihood estimate $p_{ML}(w|d) = \frac{(w,d)}{|d|}$. However for words which do not occur in the document $c(w,d) = 0$ and our language model will not work properly. To overcome this, a technique called smoothing is used which takes into account unseen words. The probability of an unseen word is considered proportional to its probability given by a reference language model.

$$
p(w|d) = \begin{cases} p_{seen}(w|d), & \text{if w is seen in d} \\ \alpha_d p(w|C), & \text{otherwise} \end{cases}
$$

Here $p(w|C)$ is the probability of the word in a background language model. Thus I get

$$\log p(q|d) = \sum_{w \in V} c(w,q) \log p(w|d) = \sum_{w \in V, c(w,d)>0} c(w,q) \log p_{seen}(w|d) + \sum_{w \in V, c(w,d)=0} c(w,q) \log \alpha_d p(w|d) \tag{2.4}$$

One can rewrite the second term in terms of all the query words in the vocabulary minus the query words matched in the document. That is

$$\sum_{w \in V, c(w,d)=0} c(w,q) \log \alpha_d p(w|d) = \sum_{w \in V} c(w,q) \log \alpha_d p(w|d) - \sum_{w \in V, c(w,d)>0} c(w,q) \log \alpha_d p(w|d) \tag{2.5}$$

Thus the final ranking function in a probabilistic model with smoothing is:

$$\log p(q|d) = \sum_{w \in V, c(w,d)>0} c(w,q) \log \frac{p_{seen}(w|d)}{\alpha_d p(w|C)} + |q| \log \alpha_d + \sum_{w \in V} c(w,q) \log p(w|C) \tag{2.6}$$

2.1.4 Query Expansion and Relevance Feedback

Query Expansion and Relevance Feedback are ways to improve the ranking of documents after the first round of retrieval is performed.

2.1.4.1 Query Expansion

IR systems can be affected by the problem of synonymy, i.e. having multiple words for the same concept. For example, someone might be searching for laptops in documents which only use the word notebook. The IR system needs to be aware of different terms the user can use when searching for information related to notebooks. Query expansion is the method which uses additional inputs of terms to rewrite the query or changing the weights assigned to terms. Query expansion is divided into two types - Global methods and local methods of query expansion [101]. Global methods are techniques for expanding the query, independent of the specific query and document results. A common method is to maintain a thesaurus and for each query term t, add to query terms from the thesaurus which are synonyms of t or related to it. Instead of using a manual thesaurus, automatically derived methods [130, 142] using word co-occurrence approaches can also used. Another approach is query log mining to exploit the query reformulations manually done by previous users and use it to suggest queries to a new user. This requires a huge log of queries, and is more appropriate in the web search scenarios. Query expansion in general has the advantage of increasing the recall of the IR system, although sometimes

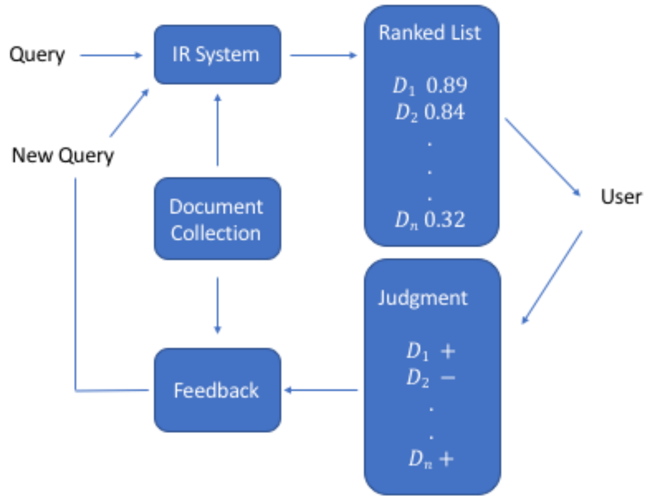


Figure 2.2: Relevance Feedback Process

the original meaning of the query can be significantly changed by adding or removing terms. This problem known as Query Drift [193] is addressed in the later sections using Quantum inspired models. Local methods of query expansion adjust the query relative to the documents which appear in the first round of retrieval for the query. Relevance feedback is a technique for local query expansion.

2.1.4.2 Relevance Feedback

Relevance feedback(RF) also intends to improve the ranking of documents by involving the user in the process. User feedback is captured over the initial result set. RF broadly involves the following steps [101]

- User query
- Initial document set
- User judgement of some documents as relevant or non-relevant
- Re-computation of relevance scores by the system based on the user feedback
- A new document set displayed to the user

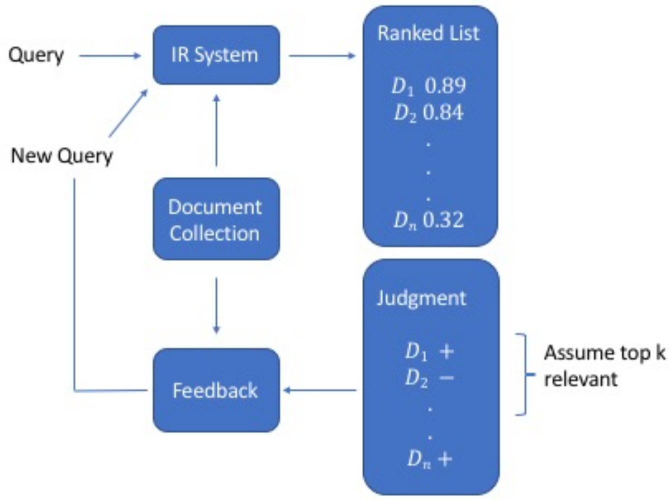


Figure 2.3: Pseudo Relevance Feedback Process

There might be several iterations of the above set. While global query expansion uses the query to re-rank documents, RF uses the documents with relevance judgments provided by the user. RF can be used to track the user's evolving or dynamic information need. The Rocchio algorithm [136] is a classic example of implementing RF. Since I know the relevant and non-relevant documents using user feedback, the task then becomes to maximize similarity of the query with the relevant documents and minimize its similarity with the non-relevant documents. In a VSM, this comes out to be the difference between the centroids of the relevant and non-relevant documents respectively. Figure 2.2 illustrates the basic RF process.

Another method called the Pseudo Relevance Feedback automates RF by assuming the top k ranked documents to be relevant and then performs RF as above (Figure 2.3). In Implicit Relevance Feedback(IRF), all the documents clicked by the user are considered as relevant and then RF is performed (see Figure 2.4). Unlike RF, it is not explicit that the clicked documents are considered as relevant by the user or not.

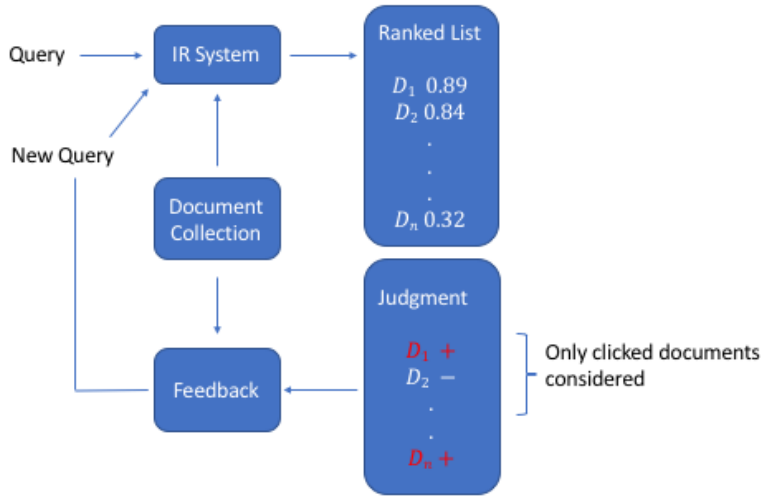


Figure 2.4: Implicit Relevance Feedback Process

2.1.5 Evaluation in IR

In the previous sections I have seen various models for IR. Evaluation is a method to compare the different models and any new model for measuring their effectiveness. The basic idea behind evaluation of document retrieval models is to have a test collection which contains a set of documents, a set of information needs expressed as queries and a set of relevance judgments generally expressed in binary as relevant or non-relevant for each query-document pair. Common test collections are TREC(Text Retrieval Conference), CLEF(Cross Language Evaluation Forum), Reuters newswire articles, etc. Common evaluation metrics are:

- Precision(P) defined as the fraction of retrieved documents which are relevant.

$$P = \frac{\text{No. of relevant documents retrieved}}{\text{Total no. of documents retrieved}} \tag{2.7}$$

- Recall(R) is the fraction of relevant documents which are retrieved.

$$R = \frac{\text{No. of relevant documents retrieved}}{\text{Total no. of relevant documents}} \tag{2.8}$$

- F-measure measures the trade-off between Precision and Recall. Its the weighted harmonic mean of Precision and Recall.

$$F = \frac{1}{\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}} \tag{2.9}$$

- Average Precision(AP) is the average of precision values of the top k documents calculated after each document is retrieved. Mean Average Precision(MAP) takes the average of AP over all the queries. Thus for a query q_j, if the set of relevant documents are $\{d_{1j}, ..., d_{mj}\}$, R_{jk} is the set of ranked documents from the top till document d_k,

$$MAP(q) = \frac{1}{|q|}\sum_{j=1}^{|q|}\frac{1}{m_j}\sum_{k=1}^{m_j} P(R_{jk}) \tag{2.10}$$

- Discounted Cumulative Gain(DCG) - IR systems ranking highly relevant documents lower in the search list should be penalized. For relevance judgments with graded relevance scores(e.g. 0-4), DCG reduces the graded relevance value of a document logarithmically proportional to its value in the search result. Thus DCG at position p is

$$DCG = \sum_{i=1}^{p}\frac{2^{rel_i}}{\log_2(i+1)} \tag{2.11}$$

- Normalized DCG(NDCG) - Document lists vary in length for different queries. Thus DCG is not ideal to compare ranking performance across different queries. So an ideal DCG(IDCG) is first computed using an ideal ranking for documents. Ideal ranking ranks documents in order of decreasing relevance. Then in the original ranking for a query, the DCG is calculated at each position and then normalized using IDCG. Thus

$$NDCG_p = \frac{DCG_p}{IDCG_p} \tag{2.12}$$

2.1.6 Multidimensional Relevance

The concept of relevance lies at the heart of Information Retrieval (IR) and is fundamentally a cognitive notion, part of our cognitive ability. The underlying intent behind all the advances in IR has been to improve the relevance of information presented to the user.

One of the main attributes of relevance is that it is a relation. There is always, implicitly or explicitly, the word ‘to’ associated with relevance [140]. It relates information or an information object to a context or situation. Relevance is also believed to manifest itself in different ways, with each manifestation indicating a different relation. Earlier works defined these manifestations at an abstract, philosophical level such as system relevance (related to the algorithmic query-document matching), topical relevance (related to subject expressed in the query), cognitive relevance (related to pertinence), situational relevance (related to utility), affective relevance (related to motivation/intent), etc. [19, 45, 46, 138, 139]. In recent years, this plurality of relevance has been studied in terms of the judgement criteria considered by users. Apart from 'Topicality', there have been 'Reliability', 'Understandability', 'Novelty', 'Interest', etc. These relevance criteria are also called dimensions of relevance.

Several works have investigated these different factors, other than the query-document topical match, which users might consider in assessing relevance. One of the earliest works [44] investigating different relevant criteria identified 38 variables which effect relevant judgement. Later on, several studies were carried out in which users were asked to specify their judgement criteria [14, 15, 109, 110, 117]. Scores of criteria such as depth/scope, accuracy, presentation quality, currency, tangibility, reliability, etc. were reported by users. In recent years, certain criteria or 'dimensions' are widely accepted as the most important considerations for user judgements of relevance. These include reliability [113, 143, 177, 183], understandability [115, 195], novelty [38, 186], effort [79, 167, 185], etc. A multidimensional relevance model was proposed [182, 191] which defined five such dimensions and was extended to seven dimensions in [93], including 'interest' and 'habit' dimensions. [80] reported positive correlations between multiple relevance dimensions and user experience measures. Relevance judgement as an aggregate of the judgements under different dimensions was investigated in [49, 50].

2.1.7 Summary and Discussion

In this section I discussed basic IR terminologies, processes and some algorithms for representation and ranking. Methods to improve ranking are also discussed including those involving user interactions. Finally a section on evaluative measures in IR is discussed. This basic introduction to IR is sufficient to help understand the experimental work which will be presented in the next chapter, along with the knowledge of basic Quantum Theory, which is discussed in the next section.

2.2 Introduction to Quantum Probability

Quantum Mechanics is a theory for calculating probabilities, which was developed in the first half of the twentieth century to explain the counter-intuitive results of experiments on microscopic particles. These results could not be explained using standard probabilistic models. It was later axiomatically organized by John von Neumann [168], thus enabling it to be used as an abstract mathematical framework even outside of Physics. At around the same time, the classical probability theory that I know of, was organized into axioms by Kolmogorov [87]. The fundamental difference between the classical and Quantum probabilities lies in the representation of events. In the classical probability theory, events are represented as subsets of a sample space. In the Quantum probability theory, events are represented as subspaces of an abstract vector space. As such, the Quantum probability theory is a generalization of the classical probability theory and can be useful in calculating the probabilities of events which cannot be represented in a set-theoretic formalism due to their inherent structure. The use of Quantum Theory for applications beyond Physics was first suggested by Niels Bohr [18], one of the founding fathers of Quantum Mechanics. He mentioned the existence of complementary variables in Psychology as similar to the incompatible properties of quantum systems. As I will discuss in this chapter, the Quantum framework provides a method to model incompatible variables naturally. In the sections to follow, a brief description of the need for the Quantum Probabilistic framework is provided and the formal concepts underlying Quantum Theory are discussed.

2.2.1 The Double Slit Experiment

The earliest experiment on microscopic particles which puzzled physicists was the Double Slit Experiment. Consider Figure 2.5 in which microscopic particles, say electrons are fired from a source to a screen consisting of two slits. On the right of this screen is another screen made up of detectors, which can detect the arrival of a particle as a function of the distance x from the center of this screen. By measuring the mean number of pulses, one can measure the probability of the electron reaching the detector screen as a function of x. The probability distribution obtained when both slits A and B are open is as shown in Figure 2.6(D). It is a complicated curve having several maxima and minima indicating that there are some locations on the detector screen that electrons never register. A traditional attempt to explain the structure of this curve would be as follows.

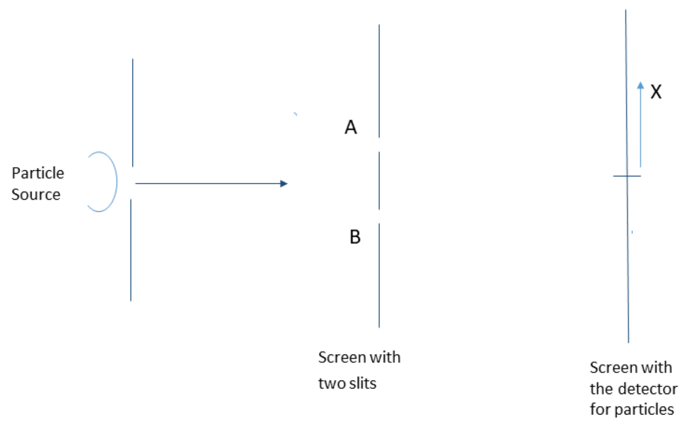


Figure 2.5: Double slit experiment setup

The experiment is repeated with only one slit open at a time. Thus, when slit A is open, I get the probability of an electron registering at a point x from the center of the detector as $P(A)$. Similarly, when only slit B is opened, this probability is $P(B)$. The distributions of $P(A)$ and $P(B)$ are shown in Figure 2.6(A) and 2.6(B). Now, when both slits are opened, the electron registers at a particular location of the detection arriving either from slit A or slit B. By the law of total probability, $P(X) = P(A) + P(B)$ and the two probability distributions add up as shown in Figure 2.6(C). In what was also of great surprise to the Physicists in the 1920s, this is not the same as the distribution obtained experimentally in Figure 2.6(D)

However, the complicated curve in Figure 2.6(D) was familiar to Physicists as it is the distribution of intensity of waves in the interference pattern obtained when a wave passes through two slits and impinges on a screen. In wave mechanics, two related properties attributed to the waves are their Amplitude and Intensity. In simple terms, amplitude is the disturbance of the wave perpendicular to its direction of travel, which can be modeled as crests and troughs. Intensity of a wave is the rate of energy delivered to a unit surface area. In the case of light waves, intensity is generally referred to by the brightness of light. Mathematically, intensity is calculated as the square of the amplitude. Amplitude is in general a complex number. The interference pattern is

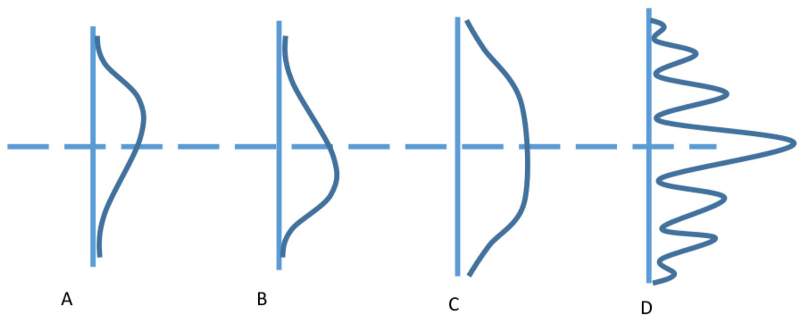


Figure 2.6: Double slit Experiment Probability Distributions

produced by adding the amplitudes of two waves and the squaring the sum up to get the intensity. If the amplitudes of two waves is given by ϕ_a and ϕ_b, then the intensity of the individual waves are $I_1 = |\phi_a|^2$ and $I_2 = |\phi_b|^2$. The intensity when the waves interfere is given by $I = |\phi_a + \phi_b|^2$. If $\phi_a = a \exp^{i\theta_a}$ and $\phi_b = b \exp^{i\theta_b}$, then I get:

$$
\begin{aligned}
I &= |\phi_a + \phi_b|^2 = |\phi_a|^2 + |\phi_b|^2 + |\phi_a| * |\phi_b|^\dagger + |\phi_a|^\dagger * |\phi_b| \\
&= a^2 + b^2 + 2ab\cos(\theta_a - \theta_b)
\end{aligned}
\tag{2.13}
$$

where the third term in the expression of intensity is the interference term and depends upon the phase difference between the complex amplitudes of the waves. This phase difference term is what gives rise to the interference pattern which we see in Figure 2.6(D).

The puzzling findings of the Double slit experiment was explained by assuming that the electron behaves like a wave when travelling from the source through the slits to the detector screen. In doing so, it is as if a single electron goes through both the slits at the same time - a fundamental Quantum property called superposition - which I will discuss shortly. The probability that an electron is detected at the screen is calculated in the same manner as the intensity of a wave. A complex quantity called the probability amplitude is ascribed to the electron corresponding to the two possible paths. Let ϕ_a be the probability amplitude for the path $S->A->X$ and ϕ_b be the probability amplitude of the electron for the path $S->B->X$, when the slits A and B are opened respectively. The amplitudes differ because of the difference in the complex phase for the two paths taken. The probabilities are calculated, according to the Born rule [20], as the square of the amplitudes(for complex amplitudes, as the product of the amplitude and its complex conjugate). Thus the probability of detecting an electron at a position X from the centre of the detector screen, when slit A is open, is $P(A) = |\phi_a| * |\phi_a|^\dagger = |\phi_a|^2$. In the case both slits are open, the probabilities are calculated by following the Law of total amplitude, a generalisation of the Law of total probability [64]. The probability amplitudes for the two paths are added up and then the probability is calculated by taking the square of the sum

$$
\begin{aligned}
P(X) &= |\phi_a + \phi_b|^2 \\
&= |\phi_a|^2 + |\phi_b|^2 + 2|\phi_a| * |\phi_b| \\
&= P(A) + P(B) + 2\sqrt{P(A)}\sqrt{P(B)}\cos(\theta)
\end{aligned}
\tag{2.14}
$$

where θ is the phase difference between the two paths. For $\theta = \frac{\pi}{2}$, we get the law of total probability as a special case.

Thus we see that the origins of Quantum Probabilities lie in the law of total amplitude and the Born rule. When a quantum entity can take one or more paths, it takes all of them at the same time, and the quantum entity is said to be in a superposition state of all possible paths. These paths influence each other, in a phenomenon called Quantum Interference, which gives rise to the extra terms in the calculation of probabilities.

It should be noted that when I say Quantum probabilities, the concept of probability remains the same as classical probabilities. The probability of a certain outcome of an experiment is p, then if the experiment is repeated many times one expects that the fraction of those which give the desired outcome is p. What changes in Quantum Theory is only the method of calculating probabilities.

2.2.2 The Axioms of Quantum Theory

In this section I present the Quantum phenomena described above as axioms and show the link between geometry and Quantum probability.

2.2.2.1 Representation of Events

Quantum Theory is generally concerned about assigning probabilities to events. In the classical method of calculating probabilities, we assume a finite sample space consisting of N points. The collections of all the points in the space are described as a set $X = \{X_1, X_2, ..., X_N\}$. An event is any subset of X, say $A \subseteq X$. For two such events $A \subseteq X$ and $B \subseteq X$, $A \cup B$ and $A \cap B$ are also events. Atomic events are given by singletons.

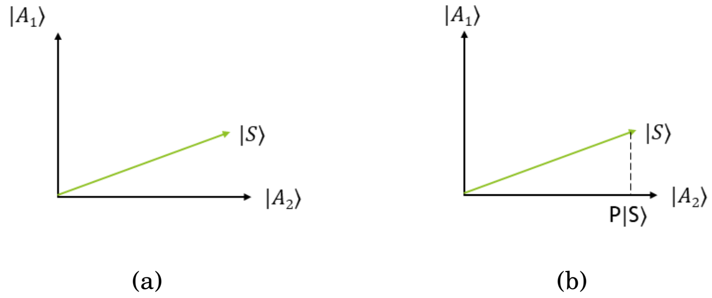


Figure 2.7: Sample Hilbert Space

In the Quantum framework, instead of the sample space of events, we have a complex Hilbert space of infinite dimensions. For simplicity I deal with a finite dimensional Hilbert space here. An N-dimensional Hilbert space is spanned by N orthonormal basis vectors $X = \{|X_i\rangle, i = 1,...,N\}$. The choice of basis is arbitrary and there can be any number of basis for a Hilbert space. Here $|X\rangle$ is the way to denote a vector in the Dirac [51] notation. An event A is defined by a subspace spanned by the subset $X_A \subseteq X$ of basis vectors. If A is an event spanned by $X_A \subseteq X$ and B is an event spanned by $X_B \subseteq X$, the intersection of the two events, also called the "meet", $A \wedge B$ is given by the span of vectors in the subset $X_A \cap X_B$. Similarly, the union of the events, called the "join", $X_A \vee X_B$, is given by the span of vectors in $X_A \cup X_B$. Note how the set theoretical intersection and union of points are replaced by the span of the intersection and union of vectors. This structural property leads to the violation of the distributive axiom. Before talking about that, I first discuss the concept of state and projectors in the Quantum framework.

2.2.2.2 States of a Quantum System

In the classical framework, we have the concept of a probability distribution function $p(X_i)$, which assigns real numbers to each point X_i of a sample space. In the Quantum framework, we define a state vector $|S\rangle$ of unit length in the Hilbert space X which induces a probability distribution over the subspaces of a Hilbert space(Figure 2.2(a)). A subspace is represented in terms of a projection operator P, which are Hermitian($P\dagger = P$) and Idempotent($PP = P$). The probability induced by a state vector $|S\rangle$ onto a subspace is given by the square of the projection of the vector onto the subspace and is calculated as:

$$|P|S\rangle|^2 = \langle S|P^\dagger P|S\rangle = \langle S|P|S\rangle \tag{2.15}$$

Figure 2.2(b) shows a two-dimensional Hilbert space with state vector $|S\rangle$ projected onto a one-dimensional subspace. In this case the projector is given by $P_{A_2} = |A_2\rangle\langle A_2|$ and the probability distribution of the state given by the vector $|S\rangle$ is:

$$\begin{aligned} |P_{A_2}|S\rangle|^2 &= \langle S|P_{A_2}|S\rangle \\ &= \langle S|A_2\rangle\langle A_2|S\rangle \\ &= |\langle A_2|S\rangle|^2 \end{aligned} \tag{2.16}$$

Here the quantity $\langle A_2|S\rangle$ is the probability amplitude of the state $|S\rangle$ for the event A_2. The state of a quantum system $|S\rangle$ is in general a superposition of all possible events. As discussed before, the events are given by all the vectors of an orthonormal basis. In the basis $\{|A_1\rangle, |A_2\rangle\}$, the state of the system is represented as:

$$|S\rangle = a_1|A_1\rangle + a_2|A_2\rangle \tag{2.17}$$

where $a_1 = \langle A_1|S\rangle$ and $a_2 = \langle A_2|S\rangle$ are the probability amplitudes and a_1^2, a_2^2 represent the probabilities for events A_1 and A_2 to occur for the state $|S\rangle$. Hence $a_1^2 + a_2^2 = 1$.

2.2.2.3 Superposition and Collapse of a Quantum State

In models based on the classical probability theory, like Bayesian networks, a state of a system evolves from moment to moment, but any given point of time the system is in a definite state. To deal with uncertainty about which state the system is in, probabilities are assigned to each state. Thus, a dynamic system is in a definite state at each point of its evolution and is governed by a probability distribution over the states.

A quantum system is different from classical systems because of its ability to be in a superposition of all the possible states at the same time. This superposed state is a new state, which is not equal to any of the possible states of a classical system. Rather, it encapsulates the possibilities of being in the possible states when a measurement is performed on the system. Hence measurement is an important part of the Quantum framework and is used to calculate the probabilities induced by the state vector of a system. Upon measurement, the superposed state collapses into one of the possible states with a certain probability. As an example, the system with state $|S\rangle$ in Figure 2.2(a) is a superposition of the basis vectors $|A_1\rangle$ and $|A_2\rangle$. It is neither in state $|A_1\rangle$ nor in $|A_2\rangle$. It is a new state with probabilities a_1^2 and a_2^2 for the system to be in state $|A_1\rangle$ or $|A_2\rangle$ upon measurement. Changing the probability amplitudes a_1 and a_2 leads to a new state, different from $|S\rangle$. This concept of collapse of a superposed state into one of the constituent states is called the Copenhagen Interpretation of Quantum Theory.

2.2.2.4 Update of the Quantum State

In the classical probability theory, when an event $A \subseteq X$ is observed, the new probability distribution is given by the conditional $P(X_i|A) = \frac{P(X_i \cap A)}{P(A)}$, where $P(A)$ is the normalizing factor which ensures that the probability assigned to the entire sample space remains one.

In Quantum theory, the observation(measurement) of an event A causes the collapse of the state and the revised state is given by

$$|S_A\rangle = \frac{P_A\,|S\rangle}{|P_A\,|S\rangle\,|} \tag{2.18}$$

In the case of state $|S\rangle$, if event A_2 occurs, the final state becomes

$$\begin{aligned} |S_A\rangle &= \frac{|A_2\rangle\,\langle A_2|S\rangle}{|\,|A_2\rangle\,\langle A_2|S\rangle\,|} \\ &= \frac{|A_2\rangle\,\langle A_2|S\rangle}{|\,|A_2\rangle\,| * |\,\langle A_2|S\rangle\,|} \\ &= |A_2\rangle \end{aligned} \tag{2.19}$$

since $|A_2\rangle$ is a unit vector. Thus we say that upon observation of event A_2, the state of the system collapses to A_2.

2.2.2.5 Compatible and Incompatible Events

Classical systems follow the principle of unicity [73], which states that for any given experiment, we have one sample space and all the events are contained in this sample space. Therefore a single probability distribution function is sufficient to calculate the probabilities for all the events.

In the Quantum framework, a state vector is represented as a superposition of all the basis vectors. One can choose to represent this state vector in any arbitrary basis. Thus the same state vector is expressed in different basis and each basis represents a particular property of the quantum system. The state vector induces different probabilities onto

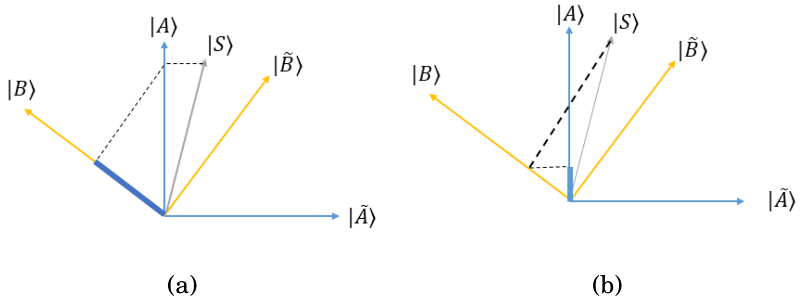


Figure 2.8: Incompatible measurements

different basis of the Hilbert space. The state vector is thus an abstract entity. It does not have any fixed representation. A particular representation conceptualises when we talk of a particular basis.

In Figure 2.8, I show a Hilbert space with two basis. One with orthonormal vectors $|A\rangle$ and $|\widetilde{A}\rangle$ and another basis with orthonormal vectors $|B\rangle$ and $|\widetilde{B}\rangle$. Consider the following events, in this particular sequence- A and B. To calculate the probability that these two events occur, the state vector $|S\rangle$ is projected onto the vector $|A\rangle$ and the new collapsed state is projected onto the vector $|B\rangle$. Hence we get the probability for A and B to occur as $P(A,B) = |P_B P_A |S\rangle|^2$, and using equations 2.16 and 2.19, we get

$$P(A,B) = |\langle B|A\rangle|^2 . |\langle A|S\rangle|^2 \tag{2.20}$$

Now if the same two events occur in the reverse order, B and then A, then the probability of them occurring is given by $P(B,A) = |P_A P_B |S\rangle|^2$, which, using equations 2.16 and 2.19, is

$$P(B,A) = |\langle A|B\rangle|^2 . |\langle B|S\rangle|^2 \tag{2.21}$$

Now, the equations 2.20 and 2.21 are different when the terms $\langle B|S\rangle$ and $\langle A|S\rangle$ are different. And they are different if A and B are vectors in different basis. In the classical theory, we can assign joint probability distributions to two events occurring together regardless of their order, i.e., $P(A,B) = P(B,A)$. But such a joint probability distribution does not exist for events belonging to two different basis in a Hilbert space. I call these events as Incompatible events. In the language of linear algebra, the projectors belonging to these events do no commute. Thus $P_A P_B \neq P_B P_A$. A geometrical explanation can be obtained from Figure 2.8. In Figure 2.8(a), the order of projections is $S->A->B$ and in Figure 2.8(a), it is $S->B->A$. We can see that the final projections in the two cases are different and depend upon the geometry of the Hilbert space(specifically, the angle between the vectors).

2.2.2.6 Violation of Distributive Axiom

In the classical probability theory, for a sample space $X = \{A, B\}$, the distributive axiom states that $A = A \cap (B \cup \widetilde{B}) = (A \cap B) \cup (A \cap \widetilde{B})$, where $\widetilde{B}$ is the complement of event B. This axiom leads to the law of total probability:

$$\begin{aligned} p(A) &= p(A \cap X) = p(A \cap (B \cap \widetilde{B})) \\ &= p((A \cap B) \cup (A \cap \widetilde{B})) \\ &= p(A \cap B) + p(A \cap \widetilde{B}) \\ &= p(B)p(A|B) + p(\widetilde{B})p(A|\widetilde{B}) \end{aligned} \tag{2.22}$$

In simple terms, this law states that if an event A occurs, it can occur along with B or without B. In the Quantum framework, consider a two dimensional Hilbert space with two basis vectors $|A_1\rangle$ and $|A_2\rangle$, as in Figure 2.9. Denoting the one-dimensional subspaces of this Hilbert spaces by their projectors P_S, P_{A_1}, P_{A_2}, the meet of two subspaces $P_S \wedge P_{A_1} = 0$ and $P_S \wedge P_{A_2} = 0$. The intersection or meet of the subspaces is defined in the same way as the intersection of events in the set-theoretic model. The difference comes in the definitions of union and complement. The union of events, defined by the join of two subspaces is the vector space spanned by the union of the set of vectors in the two subspaces. Hence $P_{A_1} \vee P_{A_2}$ is the whole two-dimensional Hilbert Space, not just the two vectors $|A_1\rangle$ and $|A_2\rangle$, which is the case in the union of two sets. Thus we get

$$P_S \wedge (P_{A_1} \vee P_{A_2}) = P_S \tag{2.23}$$

which clearly violates the distributive axiom.

2.2.2.7 Density Matrices and Trace Rule

Another way of representing a Quantum state, apart from the vector representation is the density matrix or the density operator ρ. For a state $|\psi\rangle$, the density matrix is given by

$$\rho = |\psi\rangle\langle\psi| \tag{2.24}$$

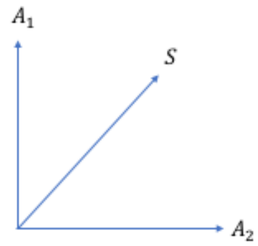


Figure 2.9: Basic 2-D Hilbert Space

which is a square matrix. The probability induced by the state represented by ρ onto a subspace represented by the projector P is given by

$$Pr = tr(\rho P) \tag{2.25}$$

where $tr(x)$ is the trace of a matrix x, the sum of its principle diagonal elements. If we denote $P = |\phi\rangle\langle\phi|$, then the trace can be written as

$$tr(\rho P) = tr(\rho\,|\phi\rangle\langle\phi|) = \langle\phi|\,\rho\,|\phi\rangle$$

which, for $\rho = \langle\psi|\psi\rangle$ is

$$\mathrm{tr}(\rho P) = |\langle\psi|\phi\rangle|^2 \tag{2.26}$$

which is the same probability as calculated using the Born rule described earlier.

Density Matrix gives us the advantage of representing a mixture of classical and quantum systems. For example, if n quantum systems are in an ensemble with each system occurring with a probability p_i, then this mixed system of quantum states can be represented by the density matrix

$$\rho = p_1\rho_1 + p_2\rho_2 + ... + p_n\rho_n \tag{2.27}$$

where ρ_i is the density matrix of a pure state which has a classical uncertainty of p_i associated with it.

2.2.2.8 Composite Quantum Systems

Multiple Quantum systems can be considered as a single system by combining their Hilbert spaces using a tensor product of the individual Hilbert spaces. If $|\psi_1\rangle, |\psi_2\rangle, ..., |\psi_n\rangle$ represent the states of n distinct quantum systems, the state of the composite quantum system of all these individual systems is given by $|\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle$, also denoted as $|\psi_1\rangle |\psi_2\rangle .. |\psi_n\rangle$ For example, consider two quantum systems represented by two dimensional Hilbert spaces with $|0\rangle$ and $|1\rangle$ as the basis vectors. The state of the systems are given by

$$|\psi\rangle_1 = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$

and

$$|\psi\rangle_2 = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$

Then the composite system is given by

$$|\psi\rangle_1 \otimes |\psi\rangle_2 = \frac{1}{2}(|0\rangle|0\rangle + |0\rangle|1\rangle + |1\rangle|0\rangle + |1\rangle|1\rangle) \tag{2.28}$$

The above composite state is a separable state. It can be factorized into two separable components, as $(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)\otimes(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)$. There exists some composite systems in nature where it is not possible to separate the composite state back into single systems. A famous example of such states are Bell states:

$$\begin{aligned} |\psi^{\pm}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle \pm |1\rangle|1\rangle) \\ |\phi^{\pm}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle \pm |1\rangle|0\rangle) \end{aligned} \tag{2.29}$$

These states are called Entangled states and this property of Entanglement is a unique and a fundamental feature of Quantum Physics. When a measurement is performed on one part of the entangled system, the state of the other system can be known instantaneously, even if the two individual components are separated by large distances. For example, consider two experimenters Alice and Bob who possess quantum states which are entangled with each other - $|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B \pm |1\rangle_A|0\rangle_B)$, where subscripts A and B denote that the states are possessed by Alice and Bob respectively. Now initially both the systems are in a superposition state. One cannot tell if it is in state $|0\rangle$ or state $|1\rangle$. If these were not entangled, measurement on the systems would have led to the collapse to state $|0\rangle$ and $|1\rangle$ with equal probability. However, in this entangled state, if Alice measures her system and it collapses to, say, state $|0\rangle_A$, then the state of the composite system collapses to state $|0\rangle_A|1\rangle_B$. Alice can instantaneously know that Bob's state has collapsed to state $|1\rangle$.

2.2.3 Quantum Cognition and Human Decision-Making

In providing an explanation for the puzzling findings of the Double-Slit experiment and many other such experiments involving microscopic particles, physicists created an entirely new theory of probabilistic and dynamic systems, of which the older classical theories were a special case. On the other hand, there have been many examples of human decision making which cannot be modelled by classical probabilistic and logic models. Decades of research by cognitive scientists have shown that in some cases human judgement under uncertainty violates the classical (Bayesian) Probability theory and other logic models [83, 158]. Such judgements have often been termed as 'irrational', because they do not follow the normative theories of rational choice. Some of these types of judgements which systematically occur are called cognitive biases [75]. The discipline of Quantum Cognition mentioned earlier seeks to model and such explain normatively irrational decision-making behaviour and cognitive biases.

Here, I am going to discuss some cognitive biases and non-normative decision-making behaviour which have been modelled in Quantum Cognition and also which come up later in this thesis in modelling of relevance judgements scenarios in IR.

2.2.3.1 Context and Order Effects

In human inference, order of information presented has an effect on the final judgement[76]. It may be due to a recency bias, where recent information has a disproportionate importance, or primacy bias where past information has disproportionate importance. Such biases lead to formation of comparative contexts where decision-makers compare the information presented to that previously seen. This in turn leads to attraction and repulsion (or contrastive) effects [107]. A example of attraction effect is a document appearing to be more relevant when presented after an irrelevant document, than when presented at first. In repulsion effect, when a lesser relevant document follows a highly relevant document, its relevance may be perceived even lower than before. All these types of effects are different types of context effects, because the presentation of information or a judgement changes the context of judgement. In case of sequential judgements, such context effects are called order effects, as different orders correspond to different contexts. Change of context by the act of judgement is similar to the change of the states of quantum systems on measurement. Therefore, QT is being increasingly applied in recent times to model such context effects [13, 153, 175].

It needs to be noted that another type of bias called position bias [48] heavily influences relevance judgements of search engine users. Here, the documents listed at the top ranks of a search engine result page get more clicks than the bottom ranked documents due to their higher position and the top-down reading of web pages. Position bias can lead to context effects in the form of anchoring bias [155], which is another type of cognitive bias in which the decision-makers base or anchor subsequent decisions on the initial information presented. In a real-world scenario, these biases will be interacting with each other. Position bias will lead to priming and anchoring bias which will affect the subsequent document judgements. However, this may be counteracted by recency bias which reduces the affect of the top ranked documents.

A famous example of order effects in decision-making is the Gallup Poll in 1997 wherein 1002 participants were asked yes/no questions about two US presidential candidates in different orders [107]. The proportion of participants answering yes/no were significantly different in different orders, as shown in figure 2.10. The two orders of questions setup different comparative contexts such that the difference between the responses of the two questions changes significantly from 18% to -3%. Now, this behaviour deviates from norms of classical probability theory because its fundamental assumption is that conjunction of two events is commutative. That is, $P(A \cap B) = P(B \cap A)$. Thus one can always form a joint probability distribution to assign probabilities to conjunction of events, but we see from the data of the 1997 Gallup Poll that a joint distribution over the decisions of participants does not exist. A simple Hilbert Space model based on the quantum mathematics can explain and model this data, as shown in figure 2.11. An initial user cognitive state vector S projects unto the representation vectors for Al-Gore (A) and Bill Clinton (B) to reproduce the probabilities in the non-comparative contexts (i.e. when the questions are asked first). In the comparative contexts, the sequential projections also reproduce the probabilities. The different angles between the basis and the initial states alone can parameterise different scenarios of non-commutative measurement.

In more related case of document judgements, consider an example where a user is asked to judge an online forum post based on two decision perspectives: (a) Is the post relevant to the topic of discussion (denoted as T) and (b) Is the sentiment of the post positive (denoted as $+$). Essentially one has to calculate $P(T,+)$. Assume that the user's cognitive state with respect to the post is uncertain, with the user being 90% certain that the post is topically relevant and 40% certain that the post has a positive sentiment. Using the braket notation, the user's cognitive state as a vector is denoted as $|S\rangle$ in a two dimensional Hilbert space. Positive and negative sentiments are orthogonal vectors

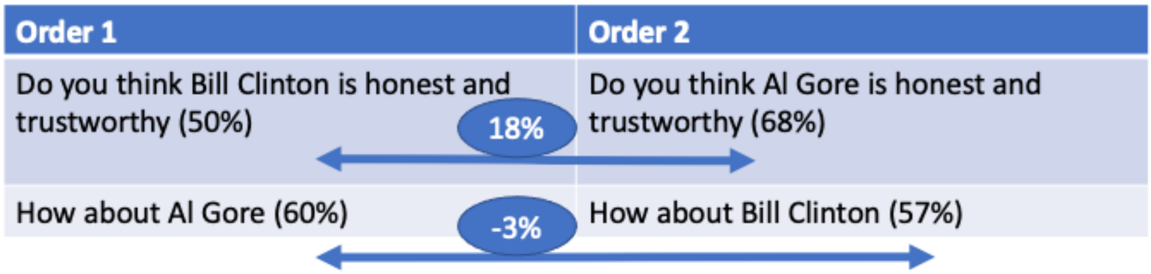


Figure 2.10: Order effects in Gallup poll data

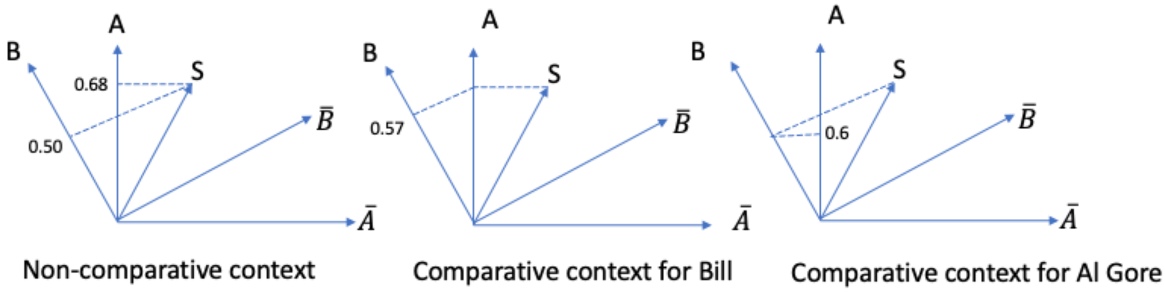


Figure 2.11: A simple quantum-inspired model for order effects in Gallup poll data

which span the Hilbert space. Also, Topical and Not-topical are another set of orthogonal vectors in the same Hilbert space. Thus, these two basis represent two different ways of evaluating the forum post - one from the topical perspective, and another from the sentiment perspective. Thus:

$$\begin{aligned} |S\rangle &= 0.9487\,|T\rangle + 0.3162\,\big|\widetilde{T}\big\rangle \\ &= 0.6325\,|+\rangle + 0.7746\,|-\rangle \end{aligned} \tag{2.30}$$

where $|T\rangle$ and $\big|\widetilde{T}\big\rangle$ represent the vectors for the post being Topical and not topical respectively. As discussed in the previous section, the probability that the post is topically relevant for the given user state, is the square of the projection of the user state $|S\rangle$ on the vector for Topicality $|T\rangle$, which gives us $|\langle T|S\rangle|^2 = 0.9487^2 = 0.90$.

Using the derivation in Appendix A, we also get:

$$|+\rangle = 0.8449\,|T\rangle - 0.5349\,\big|\widetilde{T}\big\rangle \tag{2.31}$$

Now let us calculate the probability of the post being topically relevant and of positive sentiment in two ways - (a) The user evaluates the post first considering its topicality

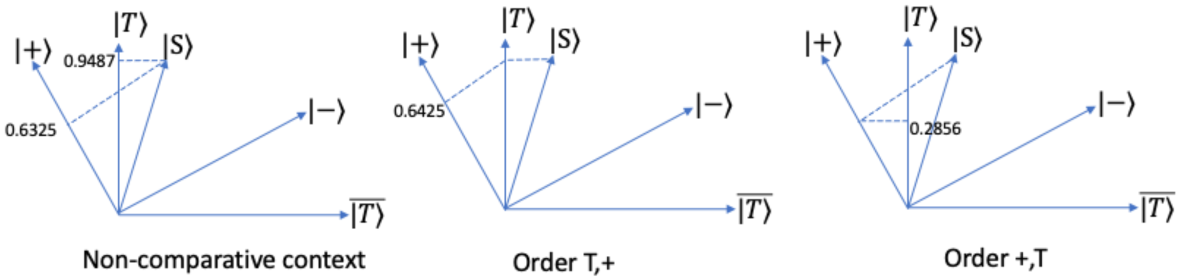


Figure 2.12: Quantum model for a toy example

and then the sentiment, (b) The user first considers the sentiment of the post and then the topicality. For case (a), we take the user's cognitive state and first project it onto the vector for Topicality($|T\rangle$) , and then project the resulting state onto the vector for positive sentiment. Thus we take the path $|S\rangle -> |T\rangle -> |+\rangle$. We get,

$$P(T,+) = |\langle T|S\rangle|^2.|\langle +|T\rangle|^2 = 0.9487^2 * 0.8449^2 = 0.6425$$

For the path $|S\rangle -> |+\rangle -> |T\rangle$, we get,

$$P(+,T) = |\langle +|S\rangle|^2.|\langle T|+\rangle|^2 = 0.6325^2 * 0.8449^2 = 0.2856 \tag{2.32}$$

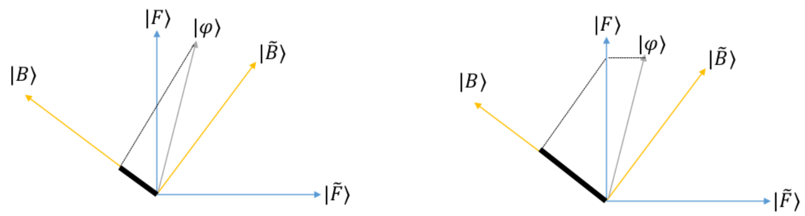
Figure 2.12 shows these projections.

Thus, when users are uncertain about judging the document because the two attributes of topicality and sentiment are not compatible, there will be an order effect. Different order of consideration of these attributes leads to different judgement and they cannot be considered jointly. QT framework is built to model such processes, by defining the two decisions of topic and sentiment as different basis. Seen from the norms of quantum probability theory, such behaviour no longer appears irrational. The bias in decision-making is no longer a feature of human cognition but of the structure and nature of the information presented in the document.

The simplistic model of two sequential judgements can be extended using a higher dimensional, complex-valued Hilbert space. For example, in a search engine result page, the interaction between different types of context effects and biases over a list of documents can be modelled using series of vector projections, provided the angle and complex phases are accurately determined.

2.2.3.2 Conjunction Fallacy

In a famous experiment [156] participants were presented with the following text:



(a) (b)

Figure 2.13: Conjunction Fallacy explained using Quantum Framework

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with the issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable: (a) Linda is a bank teller (b) Linda is active in the feminist movement and is a bank teller

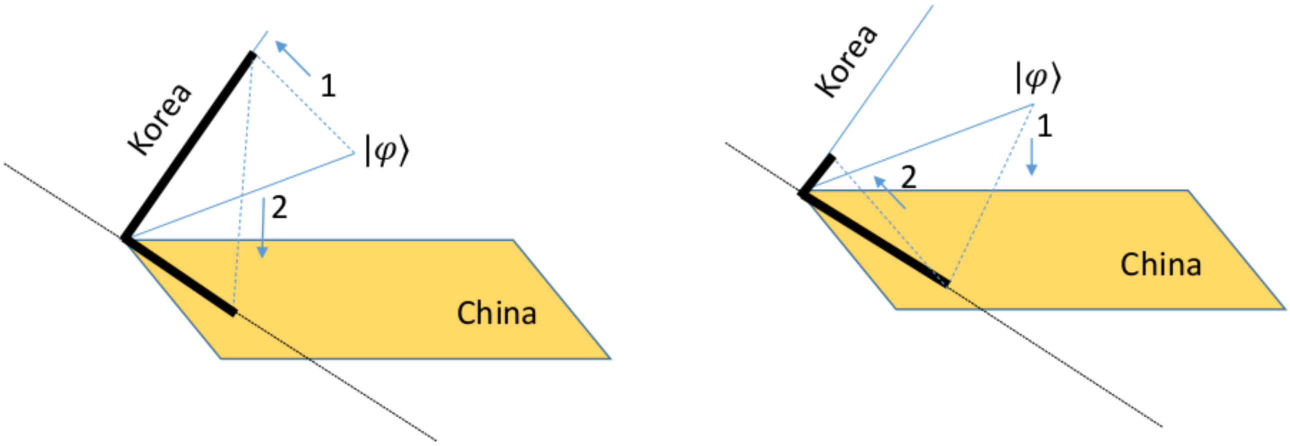
The participants consistently rated the probability of event (b) as more than that of (a). This violates the axioms of probability theory, according to which the probability of conjunction of two events is always less than that of any of the single events. In the set theoretical formalism of probability, of the sample space of all possible Lindas who are Bank Tellers, only a subset of it will be both Bank Teller and Feminist.

These findings, termed as the Conjunction Fallacy, have been investigated a lot since then [144]. Experiments have been conducted with various kinds of stories, even using words like "betting" instead of "Probability", indicating that this judgement error is not due to ignorance or misunderstanding of the concept of Probability. There is also another example of similar behaviour called the Disjunction Fallacy, where humans rate the probability of disjunction as less than that of individual events. It is concluded that the classical probability Theory cannot explain such judgements. In the Quantum probabilistic framework, the sample space is a finite or infinite dimensional Hilbert space, which is an abstract Vector Space with inner products. Each event is represented as a subspace of the Hilbert space. For example, consider the event "Linda is active in the feminist movement". In a two dimensional Hilbert space, let F be the vector(a one dimensional subspace) denoting this event. I rather denote it as $|F\rangle$, to be consistent with the Dirac notation of Quantum Theory. The negation of this event, that Linda is not active in the feminist movement is given by an orthogonal vector, denoted as $|\widetilde{F}\rangle$.

Together, these two vectors span the two dimensional Hilbert space, thus forming an orthogonal basis. We have another event "Linda is a bank teller", $|B\rangle$. Now this event is not mutually exclusive to $|F\rangle$ or $|\widetilde{F}\rangle$, nor is same as them. So I denote it in the same Hilbert space as a separate vector. $|B\rangle$ and $|\widetilde{B}\rangle$ form another orthonormal basis of the Hilbert space. The Quantum equivalent of the probability distribution function - which assigns classical probabilities to each event, is an abstract state vector. As discussed in the previous subsections, probability of an event is calculated by projecting the state vector onto the event subspace and taking the square of the projection obtained. The closer an event subspace is to the state vector, the larger the projection, and hence larger the probability. The essential difference between Quantum and classical probabilities lies in the concept of incompatible events. It is not possible to specify a joint probability distribution for incompatible events. Being certain about the outcome of one event induces an uncertain state regarding the outcomes of other events. In terms of cognition, incompatible events means that a cognitive agent cannot think about two events at the same time, thus assesses them one after the other. Incompatibility induces a sequence of judgements, instead of a joint decision. For compatible events, the Quantum framework gives the same results as the classical one. In case of the Conjunction Fallacy [33], consider the Hilbert space in Figure 2.13. The state vector $|\psi\rangle$ represents user's cognitive state prior to evaluating the two questions posed about Linda in the previous section. Note that $|\psi\rangle$ is closer to $|F\rangle$ and almost orthogonal to $|B\rangle$, indicating that for the user, the probability that Linda is a feminist is high and that Linda is a bank teller (option (a) in the problem described above) is low. As the two events described in option (b) are represented as incompatible, the user cannot consider their joint probability and evaluates them sequentially. We, therefore, first project the state $|\psi\rangle$ onto $|F\rangle$ and then onto $|B\rangle$ (Figure b) This final projection is larger than the direct projection from $|\psi\rangle$ to $|B\rangle$ (Figure a). For this alignment of vectors, the Quantum model explains the Conjunction Fallacy.

2.2.3.3 Similarity Judgements

Another paradoxical finding from the works of Amol Tversky is that similarity judgements by humans violate metric axioms. In some cases, the similarity of A and B is not the same as similarity of B and A. As an example, the similarity of Korea(North Korea) to China was judged greater than the similarity of China to Korea [154]. The explanation proposed by Tversky was that most of the features associated with Korea are similar to China. So Korea appears more similar to China. However, China has many



(a) (b)

Figure 2.14: Similarity Effect explained using Quantum Framework

other features associated with it. One has more knowledge about China than Korea, while judging Sim(China, Korea). Therefore it does not appear as similar as Sim(Korea, China). Similarity between two objects is a function of distance between points in a multidimensional space, the objects being represented by the points. Thus it should not depend upon the order in which the objects are considered. So this is another instance where human decision making does not conform to the existing methods of modelling.

Different explanations and models have been proposed for the judgement fallacies described above [11, 88, 89, 111, 112, 154, 156]. For the Quantum probabilistic explanation of asymmetry in similarity judgment, [127] propose to model the distinct features of concepts as different subspaces. So concepts with higher number of features are represented as subspaces of higher dimensionality. Consider a simplified example of a three dimensional Hilbert space where the concept China is associated with a two dimensional subspace, and Korea is associated with a one-dimensional subspace. The initial cognitive state $|\psi\rangle$ is uniformly suspended between the two subspaces. For Sim(Korea, China), it is first projected onto the subspace for Korea and then onto the subspace for China(Figure 2.14.a) The order of projections is reversed for Sim(China, Korea). As can be seen(Figure 2.14.b), the final projection(Projection 2) is larger in the case of Sim(Korea, China). The geometrical reason behind this is that for the Sim(Korea, China) case, the last projection is to a higher dimensional subspace, which preserves a larger portion of the vector than a projection to a lower dimensional subspace. This also intuitively explains the fact that since China has more features than Korea, it is easier to think of those features which are similar to Korea(form of government, etc.), when evaluating Sim(Korea, China).

2.2.4 Summary and Discussion

In this section I discussed the foundations of QT including its axioms of defining and calculating probabilities, fundamental quantum phenomena, and the features which differentiate it from the classical probability theory. I also discussed some applications of QT outside Physics from the field of Quantum Cognition and from the point of view of certain cognitive biases which can be potentially modelled using QT. Decision making is at the heart of Information Retrieval, as the most important task of an IR system is to retrieve documents which a user judges relevant. QT thus holds potential to be applied to IR to not only model existing IR methods, but also to investigate and model irrational user behaviour. In the following section I survey the work done in developing IR models inspired from QT.

2.3 Quantum-Inspired Information Retrieval

The application of Quantum Theory to Information Retrieval(IR) can be broadly divided into four aspects. First is the Quantum-inspired representation of entities like documents, queries, etc. in IR [131]. Related to the representational aspect and often overlapping with it in our review is that of Ranking in IR. The way the documents and queries are represented often determines the method of ranking documents. Third aspect to Quantum-inspired IR is that of User interactions, including relevance feedback, query expansion, and user's cognitive modelling. Lastly, another area where Quantum modelling has been extensively applied is that of Concept Combinations, where the composition of concepts and words has been investigated as being non-separable composite Quantum states.

2.3.1 Representation

It was Keith van Rijsbergen who first thought of using the Quantum framework in Information Retrieval with his book, *The Geometry of Information Retrieval* [131]. It was out of a need to develop a formal theory unifying different IR models, namely logic, vector space and probabilistic models. It also sought to explore a formal description of user interactions and the abstraction of the concept of a document in IR. Hence one finds that the user is at centre of most of the Quantum-inspired models and user interactions permeate all the representation and ranking methodologies which I will discuss in the rest of this chapter.

A representation of document is usually related to the text it contains, but a document is in general a more abstract entity. To quote [131], "it is a set of ideas, a set of concepts, a story, etc." A document is defined as an abstract object that encapsulates answers to all possible queries. This is similar to a state vector in Quantum Theory, which encodes information about all possible outcomes of measurement. The user interaction with an IR system is considered akin to measurement in Quantum Theory and the abstract document materializes to the Information Need of the user upon interaction. The Hilbert space representation of the Quantum framework is utilized to represent documents and queries in IR. It might seem similar to the Vector Space Models(VSM) discussed earlier. However, instead of modelling them as vectors in a term space, they are represented as subspaces of a concept space, spanned by a set of basis vectors. Note that the documents and queries are themselves abstract and are defined in terms of the choice of basis. The same query or documents can be defined in different basis depending upon the user's point of view. The existence of multiple basis for the same state vector is the cause of abstraction of objects in a Hilbert space. This, coupled with the fact that documents and queries are not merely vectors but subspaces in a complex, infinite dimensional vector space, gives us the leverage over the classical Vector Space Model. Besides providing a theoretical modification of the representational concepts of traditional IR, [131] also shows how existing IR tasks like co-ordination level matching, feedback, clustering, etc. can be performed using the Quantum-like formulation.

Modelling queries and documents as multiple basis in IR was also investigated in [102]. Documents and queries are modelled using some semantic descriptors. However the semantic descriptors used for the same query or document may be different for different users, or different for the same user in a different time, location or need. Therefore the use of descriptors depend upon the context. Since descriptors are modelled as basis vectors in a VSM, one can extend the VSM to include multiple basis where each basis corresponds to a context. [103] provides a method to discover different contexts from data to model them as different basis, using a matrix decomposition algorithm (Cholesky's decomposition).

Developing further the Quantum paradigm, [122] advocates the use of an information need space to model user interaction and evolving information need(IN) as part of representation. Information need is represented as a state in form of a density matrix. For ambiguous needs, the state is a mixed state and if the IN is completely specified, it is a pure state. Before any user interaction, the IR system starts as a mixed state of all possible IN states. Consider the example when a user wants to order a pizza. In the beginning the IN is in a mixture of all possible states, but a query “pizza” restricts the information need space to a subspace. Further interactions like knowing the time of the day, location of the user, etc. leads to smaller subspaces. Hence the evolution of information need is captured in the geometry. The representation of documents is proposed as in Structured Information Retrieval(SIR) which breaks away from representing the whole document as a single retrieval unit and uses document fragments like sections or paragraphs in response to a user query. It has been shown in [124] that answers to queries usually correspond to document fragments and not full documents.

The specific details of building the information need spaces are given in [120]. This paper models documents as a set of INs, with each IN being a vector. Using the SIR approach, documents are divided into fragments - paragraphs, sentences, sections or the document itself. Each document is converted into a vector using traditional techniques like tf-idf. Each of these fragments can satisfy an information need. Further, spectral decomposition of this set of vectors is performed to construct the document subspace. If the set of vectors for a document is U_d, then a subspace S_d comprises the span of the eigenvectors of the matrix $\sum_{u\in U_d} uu^T$. Eigenvectors corresponding to the top k eigenvalues are considered since the low eigenvalues can be associated with noise. [123] extend this work to include representation for queries. As a document is represented as a set of pure IN vectors corresponding to different fragments of the document, a query term t is represented as a set U_t of IN vectors that correspond to document fragments containing the term t.

Consider the example of two documents D_1 and D_2 consisting of three different paragraphs each. Let $U_1 = \{v_1, v_2, v_3\}$ and $U_2 = \{v_4, v_5, v_6\}$ be the IN vectors corresponding to the documents. Taking the simpler case of a single term query, let the term occur in paragraphs corresponding to the vectors v_2, v_5, v_6. Assuming that each fragment is equally likely to be a pure IN composing the user's actual IN, the density matrix for the query is written as

$$\rho_q = \frac{1}{N_t} \sum_{\varphi\in U_t} \varphi\varphi^T \tag{2.33}$$

where $N_t = 3$ is the number of document fragments a term occurs in. Denoting the $S_d = \sum_{u\in U_d} uu^T$ as projector for a document as explained above, I can calculate the

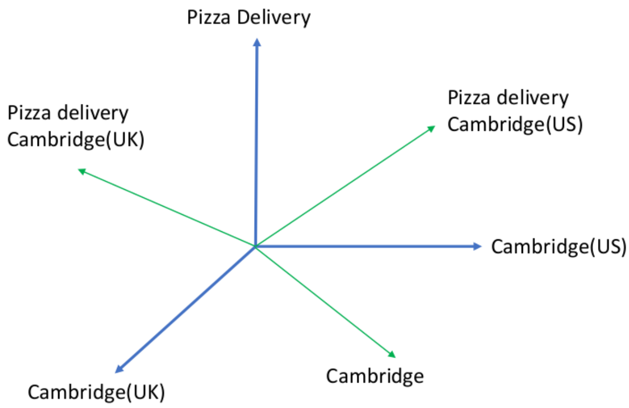


Figure 2.15: Three Dimensional Information Need Space

probability of relevance of the document for the query as

$$P(Rel|q,d) = tr(\rho_q S_d) \tag{2.34}$$

For queries with multiple terms, either a weighted mixture of density matrices for each term is considered, or in an interesting case, density matrix for a superposition of pure IN vectors is considered. Consider the three dimensional subspace of an information need space as shown in Figure 2.15. Let the vectors $\varphi_p, \varphi_{uk}, \varphi_{us}$ correspond to the INs "Pizza delivery", "Cambridge(US)" and "Cambridge(UK)". Then the IN for "Pizza delivery in Cambridge(UK)" would be represented by a superposition of φ_p and φ_{uk} vectors, as it is about both Pizza and Cambridge(UK). However the IN "Cambridge" represents classical ambiguity regarding the country and thus it is represented as a mixture of pure IN vectors φ_{uk} and φ_{us}. Thus a query "Pizza delivery in Cambridge" will be a mixture of superpositions.

This approach is extended from a single Hilbert space of information need to multiple Hilbert spaces in [119]. User IN is considered to be composed of several "aspects" which need to be addressed by the relevant documents. Each aspect is represented in a separate Hilbert space of made up IN aspect vectors for that aspect. As an example, consider the query, "What tropical storms(hurricanes and typhoons) have caused significant property damage and loss of life?". This comprises two IN aspects - tropical storms and significant damage/loss of life. So vectors for "hurricane" and "typhoons" are the IN aspect vectors for the tropical storm aspect of the query. Since each aspect vectors belong to separate Hilbert spaces, the composite system corresponding to all the IN aspects for the query is

$$\varphi_q = \varphi_1 \otimes \varphi_2 \tag{2.35}$$

where φ_1 and φ_2 are constructed in the same way as Equation 2.33. The probability of relevance of a document defined by the subspace S_d in each Hilbert space would be $P(\otimes S_d|\varphi_q) = P(S_d|\varphi_1) \times P(S_d|\varphi_2)$

The query representations for the above two approaches consider uniform weights to terms in mixtures and superpositions. Also the case of compound terms is not considered. This is dealt with in [37], which provides a sophisticated representation of query density matrices. This paper introduces a query algebra which can be used to express relationship between query terms, thus allowing for more complex representations. Several NLP techniques like Chunking and Dependency Parsing are involved to identify different IN aspects and to characterize relationship among terms within each aspect.

The concept of representing information systems as composite systems in separate Hilbert spaces is made use of in [65] for polyrepresentation of documents. A document may have different representations, from different information sources. For example, a book has a representation based on text, author profiles, reviews, rating, etc. Each representation can correspond to different aspects of the information need of the user. Assume we have two Hilbert spaces representing a collection of books, one representing the authors and another for reviews. We have two authors $|Smith\rangle$ and $|Jones\rangle$ and two types of reviews $|Good\rangle$ and $|Bad\rangle$. Then a composite system of the two Hilbert spaces will be

$$\begin{aligned} (|Smith\rangle + |Jones\rangle) \otimes (|Good\rangle + |Bad\rangle) = \\ |Smith\rangle\,|Good\rangle + |Smith\rangle\,|Bad\rangle + |Jones\rangle\,|Good\rangle + |Jones\rangle\,|Bad\rangle \end{aligned} \tag{2.36}$$

where the user is uncertain whether to read a book by James or Smith and also unaware of their ratings. However, an interesting case is that of non-separable states where a user wants a book by Smith which is rated good or wants a book by Jones which is rated as bad. The composite system of user's IN in that case is given by a non-separable state

$$|Smith\rangle\,|Good\rangle + |Jones\rangle\,|Bad\rangle \tag{2.37}$$

which reduces the uncertainty from the system point of view.

2.3.2 Ranking and Language Models

The Quantum Language Model proposed by [149] combines the Vector Space and probabilistic models of classical IR via the Hilbert space formalism. The Quantum generalization of probabilities comes in the form of representing compound terms in queries and documents as superposition events, which have no classical analogue. This generalized Quantum probability model reduces to classical in case of single terms. A document or query is represented as a sequence of projectors corresponding to each single or compound term. A document d containing words from a vocabulary of size N is represented as

$$P_d = \{\pi_i : i = 1, ..., M\} \;\; where \;\; M \leq N \tag{2.38}$$

The Hilbert space is a term space of dimension N, where each vector $|v_s\rangle$ represents a term from the vocabulary. Thus the projector for a single term is $\pi_w = |v_s\rangle\langle v_s|$. The vector for a compound term $\left|v_{s_1..s_k}\right\rangle$ is the superposition of all the vectors corresponding to the single terms

$$\left|v_{s_1..s_k}\right\rangle = \sum_{i=1}^{k} \sigma_i \left|v_{s_i}\right\rangle \tag{2.39}$$

where σ_i quantify how much the compound term represents the single term s_i. So in the same subspace, the representation of another term is created. This is not possible in traditional Vector Space Models because for every new term, single or compound when has to add a new dimension to the Vector space. Representing compound terms as superposition events solves that problem. Also, the compound term and the single terms in it are not disjoint and are related by the $\sigma_i s$. In order to construct the projectors for a document, the terms co-occurring in the document in a fixed window of size L are considered as compound terms. The language model is density matrix ρ and for a document represented by projectors $P_d = \{\pi_1, \pi_2, ..., \pi_M\}$, the language model is obtained by maximizing the following function

$$L_{P_d}(\rho) = \prod_{i=1}^{M} tr(\rho\pi_i) \tag{2.40}$$

The language model is estimated using a generalization of the E-M algorithm, called the $R\rho R$ algorithm [100].

The language model for a query ρ_q can be estimated in a similar way and the relevance of a document for a query can be calculated using a generalization of the

KL divergence method [160] called *quantum relative entropy* or *Von-Neumann(VN) divergence*. Given two language models ρ_q and ρ_d, the scoring function is

$$\Delta_{VN}(\rho_q||\rho_d) = -tr(\rho_q \log \rho_d) \tag{2.41}$$

where $tr(x)$ denotes the trace of the matrix x. The QLM performs better than baselines Language models and baseline Markov random fields for Mean Average Precision(MAP) scores for document ranking in IR.

The QLM is extended as a neural network in [188] for Question Answering Systems, a form of IR. Using word embeddings as vectors, density matrix for a sentence is constructed, for both queries and documents. The density matrix represents a mixture of semantic spaces. A joint representation of queries and documents is constructed by multiplying the density matrices for queries and documents. Then a convolution layer is applied over this joint representation followed by pooling layer, fully connected layer and a softmax layer. The binary output of the softmax layer represents probabilities of relevance and non-relevance of the answer for the question. This process is repeated for each question and answer pair and a ranking of documents based on their relevance probabilities is produced.

The Probability Ranking Principle(PRP) posits that an IR system should rank the documents for a user IN in decreasing probability of relevance. It makes the assumption that the relevance of a document to an information need does not depend on other documents. However, in real world situations, judgement of documents by a user is influenced by its previously judged documents [61]. The utility of a document may become void if the user has already obtained the same information. This 'interference' between documents can be due to information overlap between documents or a change in IN, and is accounted for in a Quantum Probability Ranking Principle [198]. It draws an analogy [105] with the Double Slit Experiment by assuming the two slits to be two documents A and B which the user judges for a query. The position x on the screen corresponds to the event that the user is satisfied by the documents A and B and decides to stop the search. If A is first document presented to the user, we have $p_{AB}(x)$ as the probability that the user stops the search at document B. In the Double slit experiment, if slit A is fixed and slit B is varied in dimensions, which is analogous to having different documents listed after document A, we get $p_{AB_i}(x)$ as the probability of stopping the search having seen document A and B_i. The problem then boils down to finding which configuration of slits AB_i exhibits maximal $p_{AB_i}(x)$.

In the classical case, if there is no interference, i.e. only one of the B_i slit is opened at a time, we have $p_{AB_i}(x) = p_A(x) + p_{B_i}(x)$.

$$\arg\max_x(p_{AB_i}(x)) = \arg\max_x(p_A(x) + p_{B_i}(x)) = \arg\max_x(p_B(x)) \tag{2.42}$$

However, in the quantum case, with all slits open, or all documents considered by the user till B_i, $p_{AB_i}(x) = p_A(x) + p_{B_i}(x) + I_{AB_i}(x)$, where $I_{AB_i}(x)$ is the interference term. Thus

$$\begin{aligned}\arg\max_x(p_{AB_i}(x)) &= \arg\max_x(p_A(x) + p_{B_i}(x) + I_{AB_i}(x)) \\ &= \arg\max_x(p_B(x) + I_{AB_i}(x))\end{aligned} \tag{2.43}$$

Hence the best choice of document to rank after A is not whose relevance probability is maximum, rather whose sum of individual relevance probability and the interference term with A is maximum. Hence, between two documents B and C, B is ranked before C iff

$$p_B(x) + I_{AB} \geq p_C(x) + I_{AC}(x) \tag{2.44}$$

This is the crux of the Quantum Probability Ranking Principle.

Recall from Equation 2.14 that the interference term depends upon the phase difference of the probability amplitudes of the two quantum systems. Thus

$$P_{AB}(x) = P_A(x) + P_B(x) + 2\sqrt{P_A(x)}\sqrt{P_B(x)}\cos(\theta_{AB}) \tag{2.45}$$

The QPRP paper [198] does not give details of how to estimate the interference term. This estimation is done in an application of the QPRP to subtopic retrieval in [197]. Subtropic retrieval is the task of providing a list of documents which covers all possible topics(IN aspects) relevant to the user IN. It advocates a more diverse ranking of documents, with minimal redundancy. Thus redundant relevant documents can be assumed to be destructively interfering(negative interference term) and the documents having exclusive information be positively interfering. This paper estimates the $\cos(\theta)$ part of the interference term as the Pearson's correlation between the term vectors of two documents. The term vectors are constructed using the BM25 scheme. The QPRP based ranking for subtopic retrieval performs better than classical approaches for subtopic retrieval.

Another method for document ranking using Quantum probabilities is discussed in

[192]. Document retrieval process is considered to be similar to a photon polarisation process. A photon has a Horizontal or Vertical polarisation which can be measured by a polarizer. There also exists superposition states of both vertical and horizontal polarizations, which is detected by a horizontal or vertical polarizer rotated at 45 degree angle.

$$|\nwarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) \tag{2.46}$$

Superposition states can be generated by passing a horizontal or vertically polarized photon through the rotated polarizer. Mathematically, the vertical and horizontal polarizers form an orthonormal basis of a two dimensional Hilbert space. The rotated polarization state form another orthonormal basis in the same Hilbert space. In the analogy, the first round of document retrieval for a query is analogous to the measurement along the vertical or horizontal basis. Then, a second round retrieval is performed to re-rank the documents by comparing all retrieved documents with the top k documents. This is analogous to passing the photons coming from a horizontal or vertical polarizer through the rotated polarizer. Or, mathematically projecting a vector represented in one basis onto the subspace generated by another rotated basis.

In the first round of retrieval, let $|\uparrow\rangle$ and $|\downarrow\rangle$ denote relevance and non-relevance of document respectively for a query. Then a document d with probability $|\alpha_d|^2$ is represented in the first round as

$$|d\rangle = \alpha_d |\uparrow\rangle + \beta_d |\downarrow\rangle \tag{2.47}$$

Taking the simple case of $k = 1$, let the topmost document in the first round of retrieval be represented as

$$|t\rangle = \alpha_t |\uparrow\rangle + \beta_t |\downarrow\rangle \tag{2.48}$$

Then, re-ranking is done by representing the document d in terms of t

$$|d\rangle = \lambda |t\rangle + \mu |\tilde{t}\rangle \tag{2.49}$$

where $\lambda = \alpha_d \alpha_t + \beta_d \beta_t$ (see appendix). The probability of relevance of document d when re-ranking is done using the top-ranked document of first round is the square of the projection of d on t, which is $|\lambda|^2$, multiplied by the probability of relevance of t, which is $|\alpha_t|^2$,

$$P(d|t) = |\lambda \alpha_t|^2 \tag{2.50}$$

when $d = t$, then $\lambda = 1$ and the probability becomes $|\alpha_t|^2$, the original probability of relevance of the top-ranked document.

The query drift problem is defined as the inferiority of results obtained on query expansion, than the original query. This is largely due to the change in underlying intent in the expanded query, from the original one. Several different solutions have been proposed for the query drift problem using pseudo relevance feedback [193].

- CombMNZ rewards documents that are ranked higher in both original retrieval list and second retrieval list by adding the relative score of a document in each of the two lists.
- Interpolation technique makes a weighted addition of relative scores in the two lists.
- The re-rank method ranks the pseudo relevance feedback based documents based on their original scores.

In [190], a document is represented in terms of relevance and non-relevance for a query

$$|d\rangle = a_d\,|q\rangle + b_d\,|\widetilde{q}\rangle \tag{2.51}$$

In terms of the expanded query q^e, the document is represented as

$$\left|d^e\right\rangle = a_d^e\left|q^e\right\rangle + b_d^e\left|\widetilde{q^e}\right\rangle \tag{2.52}$$

The existing fusion models listed directly combine the above probabilities $|a_d|^2$ and $|a_d^e|^2$. The CombMNZ reduces to

$$(\delta_q(d) + \delta_q^e(d)).(\delta_q(d)|a_d|^2 + \delta_q^e(d)|a_d^e|^2) \tag{2.53}$$

where $\delta_q(d) = 1$ if d is relevant to query q. The interpolation method becomes

$$\lambda\delta_q(d)|a_d|^2 + (1-\lambda)\delta_q^e(d)|a_d^e|^2 \quad 0 \leq \lambda \leq 1 \tag{2.54}$$

However the two probabilities $|a_d|^2$ and $|a_d^e|^2$ are under different basis and we need to write one in terms of the other. The Quantum Fusion Model(QFM) proposed in [190] does that and the final outcome combines the probabilities in the following way

$$(\delta_q(d)|a_d|^2).(\delta_q^e(d)|a_d^e|^2) \tag{2.55}$$

Thus the Quantum based model is a multiplicative model while the classical models are additive. Another slightly modified version is

$$(\delta_q(d)|a_d|^2).(\delta_q^e(d)|a_d^e|^2)^{1/\eta} \tag{2.56}$$

where a small η can make scores of different documents retrieved for q^e more separated from each other, leading to more distinctive scores. The QFM achieves better performance than the CombMNZ and interpolation methods in terms of Mean Average Precision(MAP) of retrieved documents.

The analogy to Quantum interference is also used in [148] for modelling interactions between topics. Topic modelling is used to discover hidden themes in text collections. Each topic is a probability distribution over a vocabulary and each document is a mixture of topics. Each word of a document is generated from one of the topics. The probability of a term w in a document model θ_d with k topics is given as

$$p(w|\theta_d) = \sum_k p(w|z=k,\phi)p(z=k|\theta_d) = \sum_k \theta_{dk}\phi_{kw} \tag{2.57}$$

where $z \in \{1,...,K\}$ is the topic index, $w \in \{1,...,N\}$ is a word from the vocabulary, $\theta_d = (\theta_{d1},...,\theta_{dk})$ are the topic proportions for the document d and ϕ is a $N \times K$ matrix giving distribution of topics over terms.

pConsider the case of two topics – 'war' and 'oil'. The term 'Iraq' is present in both topics. Now if a document contains both topics, still the probability of term 'Iraq' in the document is less than the the maximum of its probability in either of the topics.

$$\begin{aligned} p(w=Iraq|\theta_d) &= p(Iraq|war)*p(war|\theta_d)+p(Iraq|oil)p*(oil|\theta_d) \\ p(Iraq|\theta_d) &\leq \max(p(Iraq|war),p(Iraq|oil)) \end{aligned} \tag{2.58}$$

However, the probability of the term 'Iraq' occurring in the document should be significantly more given it contains topic 'war' and 'oil'. Current topic models do not consider the interference or relation between two topics when generating a word. They assume the topics to be independent. To capture topic dependence via Quantum probabilities, [148] assume a Hilbert space where each dimension corresponds to a word from the vocabulary. Then, each topic is a vector in this Hilbert space z_k which is a superposition of vectors corresponding to the terms. Thus we have

$$|z_k\rangle = \sum_w z_{kw}|e_w\rangle = \sum_w \sqrt{\phi_{kw}}e^{i\varphi_{kw}}|e_w\rangle \tag{2.59}$$

where $\sqrt{\phi_{kw}}$ is the complex amplitude for the topic $|z_k\rangle$ in state $|e_w\rangle$ and $(\sqrt{\phi_{kw}})^2 = p(w|z=k,\phi)$. A document can be represented as a superposition of topic states, with the coefficients being the proportion of topic in the document.

$$|d\rangle = \frac{1}{Z_d}(\sum_k \sqrt{\theta_{dk}}\,|z_k\rangle) \tag{2.60}$$

where Z_d is a normalization constant. The projection of a document vector onto a word vector is given as

$$d_w = \langle e_w|d\rangle \propto \sum_k \sqrt{\theta_{dk}\phi_{kw}}\,e^{i\varphi_{kw}} \tag{2.61}$$

And the probability of a term in the document is given by

$$\begin{aligned} p(e_w^+ e_w) &= |\langle e_w|d\rangle|^2 \propto |\sum_k \sqrt{\theta_{dk}\phi_{kw}}\,e^{i\varphi_{kw}}|^2 \\ &= \sum_k \theta_{dk}\phi_{kw} + 2\sum_{i<j}\sqrt{\theta_{di}\theta_{dj}}\sqrt{\phi_{iw}\phi_{jw}}\cos(\varphi_{iw}-\varphi_{jw}) \end{aligned} \tag{2.62}$$

This equation defines an interference topic model. The first component corresponds to the classical topic model given in 2.57 and the second is the interference term which boosts or penalizes the probability for term w in the final document model depending on the phase differences $\varphi_{iw}-\varphi_{jw}$. If a pair of topics is in phase for a given term then $\varphi_{iw}-\varphi_{jw}=0$ and $cos(\varphi_{iw}-\varphi_{jw})=1$, which increases the probability of seeing the word w in the document. Also for the phase difference of $\frac{\pi}{2}$, the interference term vanishes and the classical topic model is recovered. In their experiment, [148] estimate the interference term using a similarity measure between the topic distributions, such as the cosine similarity. The topic model helps in relevance ranking in IR by provided a better match for queries and documents, beyond the term level. This Quantum inspired topic model is applied to retrieval tasks like the TREC newswire corpora and performs better than the classical topic model.

2.3.3 User Interactions

The concept of user interactions in IR has many aspects, ranging from the cognitive level of interaction to understanding the user IN by reformulation and expansion of queries to building a user profile based on its previous interactions. In section 2.3.1, it was mentioned about the work in [102, 103] which use multiple basis of a Hilbert space to model different user contexts. This work is further extended in [104, 106] to combine different user interaction and contextual features for Implicit Relevance Feedback(IRF).

Their model uses interaction features like document display time, document saving, document bookmarking, webpage scrolling, webpage depth and document access frequency to build a user interest profile. Each of these features is represented by a basis vector. Documents are matched against a user profile by projecting a document vector unto the subspace spanned by the basis vectors for the user profile, and the larger the projection, the more the document is relevant to the profile. The features described above are calculated for each document which the user has interacted with and a document-features correlation matrix is formed. Singular Value Decomposition(SVD) is performed to get the eigenvectors, which form the basis for a user profile.

In [66, 121], a general framework for query reformulation using Quantum probabilities is described. The queries are represented as density matrices in a term space and query reformulation updates the query density matrix, which can be used to detect change in user IN in a search session.

The Query drift problem presented in the previous subsection is approached using user's search history in [187]. A document is represented as a superposition of query vectors for current query and for a latent query defined by the user's query history.

$$|d\rangle = a_d\,|q_c\rangle + b_d\,|q_h\rangle \tag{2.63}$$

q_h denotes the user IN which the user has in its mind, based on historical context, but has not been expressed into words. A document, in the superposition state of being relevant to both the current (q_c) and latent query (q_h), is then evaluated in terms of an expanded query. This is similar to the double slit experiment analogy with the two slits representing q_c and q_h and the detector screen representing the evaluation of this document in terms of the expanded query. Thus the document relevance with respect to the queries q_c and q_h interfere with each other. If $|q_e\rangle$ represents the vector for the expanded query and $|d\rangle = a_d\,|q_c\rangle + b_d\,|q_h\rangle$, then the projection of document onto the expanded query vector is

$$\begin{aligned} d \to q_e &= |\,\langle q_e|d\rangle\,|^2 \\ &= |a_d\,\langle q_e|q_c\rangle + b_d\,\langle q_e|q_h\rangle\,|^2 \\ &= |\,\langle q_c|d\rangle\,\langle q_e|q_c\rangle + \langle q_h|d\rangle\,\langle q_e|q_h\rangle\,|^2 \\ &= |\,\langle q_c|d\rangle\,\langle q_e|q_c\rangle\,|^2 + |\,\langle q_h|d\rangle\,\langle q_e|q_h\rangle\,|^2 + 2\,\langle q_c|d\rangle\,\langle q_e|q_c\rangle\,\langle q_h|d\rangle\,\langle q_e|q_h\rangle\cos\theta \end{aligned} \tag{2.64}$$

where θ is the phase between the two paths $d \to q_c \to q_e$ and $d \to q_h \to q_e$. We get interference between the two paths, because the actual path is superposed, $d \to (q_c \& q_h) \to q_e$ i.e. the first round retrieval is assumed to be using both the current and the latent query at the same time. This method of query expansion using user's previous interactions, is termed as the Quantum Query Expansion(QQE) approach for session search. It gives better results than the QFM discussed in the previous subsection, over the NDCG evaluation measure.

This approach of using user's historical queries into context is also used in [92] for a Contextual Quantum Language Model(CQLM), extending the QLM discussed in the previous subsection. The density matrix representing the language models is constructed both for the current query and for the previous historical queries in the session and both are combined to give the CQLM.

$$\rho_{CQLM} = \xi \times \rho_c + (1-\xi) \times \rho_h \tag{2.65}$$

where $\xi \in [0,1]$ combines the two language models and determines the extent of the impact of previous information on the CQLM. The construction of ρ_h is done by combining all the ρ_{h_i} of the previous queries in the session. The historical queries in the session which are similar to the current query are given more weight. Hence

$$\rho_h = \sum_{i=1}^{n-1} \gamma_i \times \rho_{h_i} \tag{2.66}$$

where γ_i is the similarity between current query q_c and previous query q_i. The similarity is calculated by representing each query as a TF-IDF vector, derived from the concatenation of all of its result documents.

The CQLM does not capture the evolution of user's information need. To model this, the same paper proposes an Adaptive CQLM(ACQLM). The basic idea is to decompose the current query into three parts - the common part, the added part and the removed part, with respect to the previous queries in the session. For example, if $q_n = abd$, $q_{n-1} = abc$, then ab is the common part, d is the added part and c is the removed part. The common part reflects user's search theme for the session. The removed and added parts reflect the change in IN. The ACQLM adjusts the QLM in such a way, as to assign relatively higher probability to the terms(or composite terms) of the common and added parts. Thus the ACQLM builds upon the CQLM by incorporating query change signals in a structured and intuitive way, moving the QLM into the right direction.

Important work has been carried out from the user cognitive aspect of IR, drawing parallels from Quantum Theory and using the Quantum framework to model and explain some of the aspects. An early work [189] investigate the interference in relevance judgement of a topic caused by another topic. Consider the topics "Brave Heart" (William Wallace's nickname and the name for his film biography) and "William Wallace" and a biographical article about William Wallace. Both topics are relevant to the article. Consider another topic about "William Wallace's wife". In a user study, it was found out that when the topics "Brave Heart" and "William Wallace" were displayed together for the article, 93% users chose to judge the article as relevant to "William Wallace" and only 14% chose it as being relevant to the topic "Brave Heart". However, when "Brave Heart" was displayed together with "William Wallace's wife", 89% of the users judged "Brave Heart" as relevant to the article and 5% judged "William Wallace's wife" to be relevant. There were experiments conducted with different topics and articles and such type of context effects were found, where the presence of one topic or document influences the relevance judgement of another topic or document. In the first case, "William Wallace" is highly relevant to the article and it sets a high comparison baseline which effects the judgement for the topic "Brave Heart" and results in low probability of relevance. However, in comparison with "William Wallace's wife", it appears more relevant to the users. For a Quantum probabilistic explanation of this result, I regard "William Wallace" and "William Wallace's Wife" as two different contexts for the topic "Brave Heart". Each context is described by a basis. So a document or topic d can be represented in the context basis as

$$|d\rangle = a_1 |q_1\rangle + a_2 |\bar{q_1}\rangle \tag{2.67}$$

Where $|\bar{q_1}\rangle$ represents the absence of context q_1. Representing a query q in the same basis as $|q\rangle = b_1 |q_1\rangle + b_2 |\bar{q_1}\rangle$, I calculate the relevance of the document d for query q as

$$\begin{aligned} P(d|q) &= |\langle q|d\rangle|^2 \\ &= (a_1 b_1 + a_2 b_2) * (a_1 b_1 + a_2 b_2) \\ &= a_1^2 b_1^2 + a_2^2 b_2^2 + 2 a_1 b_1 a_2 b_2 \cos\theta \end{aligned} \tag{2.68}$$

where, we assumed that the probability amplitudes are complex quantities and θ represents the phase term. The third term which is the interference term can be positive or negative depending upon the phase differences. For some contexts, the interference term is negative and the relevance of the same document for the query can be low, which explains why "Brave Heart" is judged less relevant when seen in the context of "William Wallace". There is a negative interference term which lowers the probability of relevance for the given query/article.

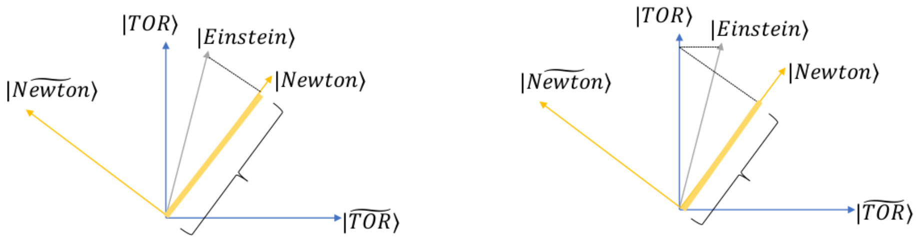


Figure 2.16: On viewing document about Theory of Relativity, the judgement of topic Newton is lower for the query Einstein

Another work which explores the influence of context in document relevance judgement is [171]. This work specifically investigates the presence of Order Effects in relevance judgement of documents. In the experiment, users are shown a pair of documents for a query and the relevance judgement by the user for a document is affected by the order in which the document is presented. For example, for the query "Albert Einstein" users are shown documents about "Issac Newton" and "Theory of Relativity". The relevance probability of "Issac Newton" is lower when it is shown after "Theory of Relativity" (called a comparative context) than when it is shown first (non-comparative context). In simple terms, having seen a more relevant document first, users judgement about a particular document may change. This can be explained as an Order Effect due to incompatibility between the topics, as shown in Figure 2.16. The paper also tested the Quantum Question Order inequality [175], which is an inequality for testing incompatibility in decision making systems.

2.3.4 Concept Combinations and Natural Language Processing

The Quantum probabilistic framework has also been investigated in literature relevant to IR, to model the combinations of words and their associations. How the combination of words give rise to meanings disconnected from the individual words is particularly useful in IR to understand user's information need from the textual queries. The work on quantum modelling of concept combinations also suggests the role of cognition in giving rise to meaning of combined words.

The principle of semantic compositionality [118] states that the meaning of a whole (syntactic entity) is a function only of the meaning of its (syntactic) parts, together with the manner in which these parts were combined. In compositional semantics, higher order semantic structures are constructed by combining the semantics of its constituent parts. Concept combinations can be broadly classified into two types [74]. First there are combinations that have intersective semantics. For example, the meaning of the composite term "Black Cat" can be understood as the intersection of black colored objects and cats. This type of semantics is compositional, as the semantics of "Black Cat" can be determined in terms of its individual components. However, concept combinations are not always intersective. For the term "astronaut's pen", there is no object in the intersection of sets of objects astronaut and pen. Allowing for a fuzzy notion of intersection leads to a problem called overextension. As an example, when users are asked to judge the word "guppy" as relating to a "pet" or a "fish", or a "pet fish", it is associated more with "pet fish" than pet or fish. This is similar to the conjunction fallacy, where an intersective component is given a higher weight than the individual components. If the weight of a member to a class is represented in terms of probability, then according to the classical axioms of probability

$$P(pet \wedge fish) \leq \min\{P(pet), P(fish)\} \tag{2.69}$$

But the probability of intersection is found to be greater than that of the individual components in the overextension effect. To explain the overextension of membership weights, consider the concepts "pet" and "fish" are represented as orthogonal vectors of a basis of a two dimensional Hilbert space. Then the intersection between them is defined as a superposition state:

$$\frac{1}{\sqrt{2}}|pet\rangle + \frac{1}{\sqrt{2}}|fish\rangle$$

Then the probability that another word w belongs to the superposition of the concepts is given by

$$\frac{1}{2}|\langle w|pet\rangle|^2 + \frac{1}{2}|\langle w|fish\rangle|^2 + I$$

where I denotes the interference term in the expression and is responsible for the overextension effect.

The first attempt to model concept combinations using Quantum Theory was in [69]. Since then, a lot of work has been done in developing quantum models of concept combinations [4, 28]. The underlying idea is that the meaning of a concept is determined by the context in which it occurs. It changes with change in context, which makes it analogous to the state of a quantum particle which is influenced by the measurement context.

suit
suit 1.00000
lawsuit 0.868791
suits 0.807798
plaintiff 0.717156
sued 0.706158
plantiffs 0.697506
suing 0.674661
lawsuits 0.664649
damages 0.660513
filed 0.655072
behalf 0.650374
appeal 0.608732

Table 2.1: Words similar with suit

Another advantage of Quantum Theory in Natural Language Processing(NLP) comes in the area of word sense disambiguation. [178] uses vector negation as a technique for word disambiguation which is further explained in [179] using the concept of superposition of orthogonal vectors. For any word embedding model of non-orthogonal vectors, addition or subtraction of two word vectors does not provide any semantic information about them. For example, if a and b are two vectors, then $a-b$ does not remove from a the part of b. Instead it may rather remove some latent information from a.

For example, $suits - lawsuit$ vector is not close to the vector for $dress$. As a solution, the negation of vectors is defined not just a simple subtraction but rather

$$a\ NOT\ b = a - \frac{a.b}{|b|^2} b \tag{2.70}$$

which makes the vector $a\ NOT\ b$ orthogonal to b, thus removing any component of b from a, as shown in Table 2.2. This way, the legal sense of suit is separated and we get the cloth sense of suit.

The Quantum framework inherently supports this type of vector negation. The vector for a word can be constructed as a superposition of all the possible senses. For example, assuming two senses for the word suit, we have

$$|suit\rangle = a\,|cloth\rangle + b\,|legal\rangle \tag{2.71}$$

In quantum theory, the vectors in a superposition are orthogonal, meaning that the word suit can be only used in one of these two senses at a time. Vector negation is modeled

suit	suit NOT lawsuit
suit 1.00000	pants 0.810573
lawsuit 0.868791	shirt 0.807780
suits 0.807798	jacket 0.795674
plaintiff 0.717156	silk 0.781623
sued 0.706158	dress 0.778841
plantiffs 0.697506	trousers 0.771312
suing 0.674661	sweater 0.765677
lawsuits 0.664649	wearing 0.764283
damages 0.660513	satin 0.761539
filed 0.655072	plaid 0.755880
behalf 0.650374	lace 0.755510
appeal 0.608732	worn 0.755260

Table 2.2: Separating legal part of suit

naturally as

$$|cloth\rangle = \frac{1}{a}(|suit\rangle - b\,|legal\rangle) \tag{2.72}$$

The first connections between Quantum Theory and semantic spaces were established in [6]. In [25], one such connection is presented using the Hyperspace Analogue to Language(HAL) model [31, 97]. For a vocabulary of N words, the HAL algorithm constructs an $N \times N$ matrix by moving a window of length l over a text corpus, thus capturing word co-occurrences within the window. Each element of the matrix is a measure of word co-occurrence and in one way, word similarity. Each window is considered as a semantic space and approximates the context or the sense associated with the word. The semantic space for a word is computed in terms of the sum of semantic spaces. If there are y windows around the word w and x of them deal with a particular context i, then the semantic space S_i occurs with probability $p_i = \frac{x}{y}$ and the semantic space for word w can be written in terms of context semantic spaces as

$$S_w = \sum_{i=1}^{m} p_i S_i \tag{2.73}$$

This formula is same as that of a mixed density matrix written as a mixture of density matrices of pure states. Thus the context of words can be considered as pure states. HAL is also used in [77, 78] to model word correlations like Quantum correlations of non-separable states.

2.3.5 Summary and Discussion

van Rijsbergen's seminal work introduces a way to look at Quantum probabilities as an extension of the classical probabilities and the similarities between the abstract properties of quantum particles and documents. The work shows how the Quantum framework can produce the same modelling as the boolean, geometric and probabilistic models. It gives only some intuitive ideas of how to incorporate the advantages of Quantum Theory to improve IR systems.

Research inspired from van Rijsbergen's ideas implement ad-hoc IR systems by considering information need space as Hilbert space and introducing ideas of superposition for ambiguous queries. These representations provide a good starting point in the field of Quantum probabilistic IR, but they fail to outperform the state-of-art methods in IR. The proposed polyrepresentation method is not yet applied to any dataset. The Quantum Language Model is the most promising of them all and intends to solve a crucial problem in NLP and IR - of representing compound terms in relation to the individual terms. Superposition principle is made use of and a quantum algorithm to build a language model is applied. It performs better than baseline models like tf-idf and BM25. But it is only applicable as a unigram model and thus cannot be applied for more complicated n-gram modelling. Herein lies the scope of improvement - to come up with an n-gram QLM and even extend it to generative models and Recurrent Neural Network Architectures.

The QPRP is an important milestone in Quantum probabilistic IR as it approaches and combines Quantum Theory and IR from an axiomatic point of view. However the problem of quantifying the interference term remains and document similarity approaches applied do not make use of the Quantum advantage, even in the topic model experiment. One needs to devise a way to subscribe complex phases to documents and then calculate the interference terms.

The Query fusion and Query expansion approaches make use of superposition and interference phenomena, however it is difficult to get an intuitive explanation of how the quantum phenomena are coming into effect and providing the advantage over classical methods. The Contextual QLM and ACQLM are promising applications of the QLM to incorporate user interactions, however they are outperformed by the state-of-art machine learning based methods.

The cognitive experiments on order effects in document judgement provides a good insight into why Quantum Probability is useful in modelling human decision making. However as such, these experiments are only on small user collected samples and not on real world search data. Also they do not provide a way to make use of these order effect information to improve the IR system effectiveness.

The work on concept combination and NLP also clearly show the usefulness of Quantum Theory over classical probability theory. However they are outperformed by the current state-of-art models in NLP like the word-embeddings. One reason could be that these concept combination and NLP approaches inspired from Quantum Theory do not yet incorporate the use of complex numbers. The use of complex numbers to represent phases leads to the non-linear interference terms and can provide an advantage to any model.

Thus to enumerate the gaps in Quantum IR literature:

1. Quantum inspired models for ad-hoc IR do not scale up for real world datasets.

2. Quantum Language Model is a unigram language model, cannot model complex text sequences very well. Although it unifies vector space and probabilistic models, the vector space has a very large dimension, of the size of the vocabulary. There is a need to find a lower dimensional latent space and also to extend the Quantum language model for n-gram modelling.

3. The algorithms making use of the interference phenomena, for example the QPRP applications, use the interference term as an additional term, estimating it using other means. The interference term should be a part of the Quantum probability calculation, arising out of the Born rule of taking square of amplitudes. Using complex amplitudes is also a challenging task.

4. The work investigating contextual effects like Order Effects in relevance judgement use few samples of user collected data. User studies have the advantage of collecting whatever data one wants. However, it is not known whether the Quantum like phenomena found in these studies be replicated for real world search data. Therefore, one needs to devise methods to detect such quantum-like phenomena in real world user data.

5. Finally, and most importantly, most of the mathematical models used in Quantum IR do not empirically justify the need for using QT. This makes it difficult to explain the quantumness of the models and whether any improvements over baseline models are due to quantum phenomena or something else.

Chapter 3

Investigation and Modelling of Quantum-like phenomena in Query Logs

As mentioned in the literature review, there have been some user studies [22, 171, 189] which have used QT to model user behaviour in IR. However, it is not known whether such phenomena can exist in larger, standard IR datasets which are a better reflection of the information seeking behaviour of real-world users. As opposed to the data collected in lab-based on crowdsourced user studies with specifically designed experiments, standard IR datasets generally have a particular format. They typically consist of a textual query, a list of retrieved documents for the query and a proxy for estimating the relevance of the documents to a user or a group of users. For example, the dataset may contain whether the document was clicked and the duration for which the user read it (dwell time), or an integer value corresponding to different relevance grades (from irrelevant to highly relevant). The datasets can be a standardised collection like Cranfield collection, TREC, etc. or could be query logs of a web search engine (e.g. Yahoo, Bing, etc.). Such datasets are popularly used by IR researchers and serve as a way of benchmarking different IR systems and provide a way of comparing different models or systems. If it can be shown that such datasets contain quantum-like phenomena, then a) it means that there are certain aspects of user behaviour in IR which cannot be modelled using the current algorithms based on classical logic and probability and thus there will always be a limit to their efficiency and b) There is a scope of further improvement of these algorithms by integrating constructs from QT with them to build quantum-inspired IR models. Hence, the first investigations in this thesis are about detecting quantum-like phenomena in large, standard IR datasets. Specifically, I focus on multidimensional relevance judgements. This is because different dimensions of relevance of document have similarity with properties of quantum systems like spin. For electrons, the property of spin is two valued (up or down spin) but measuring spin along different axes give a different value. Hence spin is always dependent on the measurement context. Similarly, the relevance of a document to a user can be different depending on the dimension considered to judge it. A document may be topically relevant but not reliable. Furthermore, measuring spin along one axis disturbs the configuration of spin along another axis such that spins cannot be jointly measured along two such axes. This second property of incompatibility is what distinguishes quantum systems from classical. While document judgement possesses the first property of multiple measurement contexts (aka relevance dimensions), in this chapter I hypothesise that judgements made along different dimensions can be incompatible. In other words, judgement of a particular relevance dimension can change a previous dimensional judgement. For example, a document may be considered reliable by a user but if on judging understandability the user finds it difficult to understand, he or she might change the judgement of reliability too. Therefore, I formulate the following research question:

- **RQ: Can standard IR datasets or query logs provide evidence of incompatibility between judgement of different dimensions of relevance?**

In this chapter I discuss two experiments. Both of them share the common approach of modelling multidimensional document relevance in a Hilbert Space. For this, I utilise some existing query-document features to extract numerical estimates for different relevance dimensions and represent a document in a Hilbert Space. Then, I seek evidence of incompatibility between relevance dimensions and use it to perform an experiment to predict order effects in document judgement. In the second experiment, I combine the Hilbert space representation and the collapse postulate of QM to propose a re-ranking algorithm based on user feedback in session search.

3.1 Hilbert Space Representation of Multidimensional Relevance

To provide a recap, as discussed in the literature review, multidimensional relevance is referred to the existence of different factors other than the topical query-document match, which are considered by users to judge the relevance of a document. The extended multi-dimensional user relevance model [182, 191] proposed in [93] defines seven dimensions of relevance namely "Novelty", "Reliability", "Scope", "Topicality" and "Understandability", "Habit" and "Interest". Figure 3.1 revisits the definitions of these relevance dimensions.

3.1.1 Theoretical Construction and Intuition of Hilbert Space

Consider a real-valued two dimensional Hilbert Space. Relevance with respect to a dimension (e.g. Reliability) is a vector in this Hilbert space. Non-relevance with respect to the same dimension is an orthogonal vector to it. Further, vectors are denoted as kets, following the Dirac's notation. For example, the vectors for relevance and non-relevance with respect to novelty are denoted as $|novelty\rangle$ and $\left|\widetilde{novelty}\right\rangle$. Figure 3.4(a) shows the construction of a two-dimensional Hilbert Space for a relevance dimension.
Next, the user's perception of a document with respect to a dimension of relevance is also modelled as a vector in this Hilbert space. This vector is a superposition of relevance and non-relevance vectors with respect to a dimension, e.g., $|d\rangle = \alpha\,|novelty\rangle + \beta\left|\widetilde{novelty}\right\rangle$. The coefficient $|\alpha|^2$ is the weight (i.e., probability of relevance) the user assigns to document d in term of novelty, and $|\alpha|^2 + |\beta|^2 = 1$. I will talk about how to calculate these coefficients in the next section. Figure 3.4(b) shows the modelling of user's cognitive state for document d with respect to the Novelty dimension.

Depending on a user's preference of relevance dimensions for a particular query, the user will judge the same document differently. A document might be of interest to the user but may not be novel while the user is looking for latest documents about the query. This phenomena can be modelled in the same Hilbert space by having different basis for different dimensions of relevance. The same document d can be written in terms of another set of basis vectors corresponding to another dimension of relevance. For example:

$$
\begin{aligned}
|d_1\rangle &= \alpha_{11}\,|novelty\rangle + \beta_{11}\left|\widetilde{novelty}\right\rangle \\
&= \alpha_{12}\,|habit\rangle + \beta_{12}\left|\widetilde{habit}\right\rangle \\
&= \alpha_{13}\,|topic\rangle + \beta_{13}\left|\widetilde{topic}\right\rangle \\
&. \\
&. \\
&= \alpha_{17}\,|scope\rangle + \beta_{17}\,|\widetilde{scope}\rangle
\end{aligned}
\tag{3.1}
$$

and so on in all seven basis. Figure 3.4(c) shows the construction of such a Hilbert space showing two basis for simplicity.

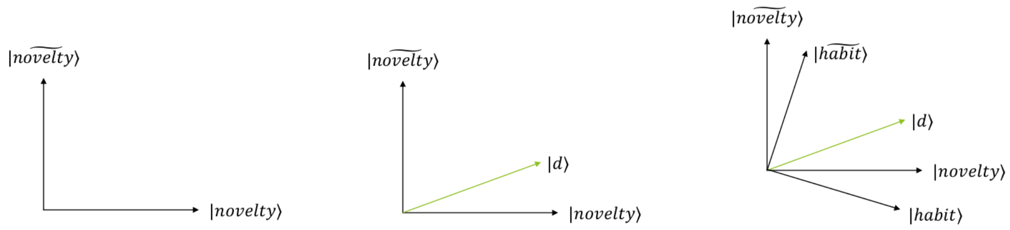
I have represented the user's cognitive state with respect to a single document in different basis corresponding to different dimensions of relevance. Similarly, one can do that for all the documents retrieved for a query. Each document will be represented in a separate Hilbert space.

The user's cognitive state for a document d is an abstract vector, because the vector has different coefficients in different basis. It does not have a definite set of coefficients and a particular representation of the vector comes into picture only when we talk of a particular relevance dimension. This is similar to the concept of the state vector in Quantum theory which contains all the information about a quantum system, yet is an abstract entity and has different representations of the same system. We get to see a particular representation of a system depending on how we measure it. A document may look highly relevant in one basis, if it has a high weight in that basis and the user makes judgement from the perspective of that relevance dimension. However, the relevance can change if the user considers a different basis (a different perspective of looking at the document).

3.1.2 Extracting coefficients for Hilbert Space construction

This subsection explains how to build an actual Hilbert Space for a given query log dataset. Algorithm 1 shows the full process of extracting the coefficients of a document vector in different basis.

We need a dataset of query logs where the minimum requirement is that for each query logged, there should be a list of retrieved documents and a measure of relevance



(a) (b) (c)

Figure 3.1: Hilbert Space representing User's perception of Document

Table 3.1: Seven dimensions of relevance

Relevance Dimensions	
Topicality	The extent to which the retrieved document is related to the topic of the current query.
Reliability	The degree to which the content of the document is true, accurate and believable. Determined by the reliability of source.
Understandability	Extent to which the contents are readable. Vocabulary, complexity of sentences, layout of pages, etc. taken into consideration.
Interest	Topics from user's past searches.
Habit	Focus on behavioural preference of users, e.g. always using certain websites for particular tasks.
Scope	Whether both breadth and depth of the document are suitable to the Information Need
Novelty	Whether the document contains information which is new to the user, or the document itself is newly created

for each of the document against the query. It can be either direct labels classifying relevance, partial relevance, non-relevance, etc. Or it can be the amount of time a user spends in viewing a document. In the latter case, there is this concept of Satisfied-Click (SAT-click) where a document viewed for at least a specific amount of time is considered as relevant to the user (usually 30 seconds). Next, we need to assign a score to each query-document-relevance dimension triplet. Such a score should reflect the relevance of the document for the query with respect to the dimension and be convertible to or interpreted as a probability. These scores or an appropriate function of these scores will form the coefficients of the document vector in the respect basis (remember from the previous section that a document vector in a Hilbert space is represented in different basis, one for each dimension).

Algorithm 1 works as follows. The arguments for the main procedure are a query, a set of retrieved documents for it and a list of relevance dimensions (Step 1). Assume that there is a set of features for each relevance dimension which are extracted from each query-document pair, per dimension (Step 3). As discussed above, the features alone cannot be interpreted as probabilities. So I feed the features for each dimension into a ranking algorithm which outputs a relevance score for each document for each dimension (Step 4). In other words, each document will have seven scores assigned to it for a query, corresponding to relevance with respect to each of the seven dimensions. These seven scores assigned to a document are normalised using the min-max normalisation technique across all the documents for the query (Step 5). This converts them into a number between 0 and 1 which can be interpreted as a probability of relevance of the document with respect to a dimension. The square root of the normalised score for each dimension forms the coefficient of superposition of the relevance vector for the respective dimension (Steps 7, 8). Square root is taken because in the quantum framework, probability is the square of the coefficient and not the coefficient itself. For example, for a query q, let $d_1,d_2,...,d_n$ be the ranking order corresponding to the "Interest" dimension. Let relevance scores be $\lambda_{11},\lambda_{21},...,\lambda_{n1}$ respectively. Here λ_{ij} represent the score of document d_i ($i \in \{1,2,...,N\}$) for dimension j ($j \in \{1,..,7\}$). The vector for document d_1 in the Interest basis is constructed as:

$$|d_1\rangle = \alpha_{11}\,|interest\rangle + \beta_{11}\left|\widetilde{interest}\right\rangle \tag{3.2}$$

where $\alpha_{11} = \sqrt{\frac{\lambda_{11}-min(\lambda_{i1})}{max(\lambda_{i1})-min(\lambda_{i1})}}$, where $max(\lambda_{i1})$ is the maximum value among $\lambda_{11},\lambda_{21},...,\lambda_{n1}$. Note α_{ij} is the coefficient of the document vector for d_i in the basis corresponding to dimension j. Similarly, the second document is represented in another Hilbert space for Interest and Reliability dimensions as :

$$\begin{aligned} |d_2\rangle &= \alpha_{21}\,|interest\rangle + \beta_{21}\left|\widetilde{interest}\right\rangle \\ &= \alpha_{22}\,|rel\rangle + \beta_{22}\left|\widetilde{rel}\right\rangle \end{aligned} \tag{3.3}$$

Algorithm 1 Extracting vector coefficients from query log data for Hilbert space

procedure EXTRACTCOEFFICIENTS(rel, $docsALL$, $query$)
    for all r in rel **do** ▷ rel - list of 7 dimensions
        $features[r] \leftarrow getFeatures(docsALL, r, query)$ ▷ Extract features from all retrieved docs for a given query
        $scores[r][d] \leftarrow reRank(docsALL, features[r])$ ▷ re-rank based on each dim and get score
        $normScores[r][d] \leftarrow normaliseScores(scores[r][d])$
        for all d in docsALL **do**
            $\alpha[d][r] \leftarrow sqrt(normScores[d][r])$ ▷ construct vectors
            $\beta[d][r] \leftarrow 1 - |\alpha[d][r]|^2$

3.2 Experiment 1 - Session Search Re-ranking using Hilbert Space

3.2.1 Hypothesis

The hypothesis for the experiment is that in a particular search session or search task, there is a particular relevance dimension or a combination of relevance dimensions which the user considers for judging documents. For example, if the user wants to get a visa to a country, he or she would prefer documents which are more reliable (Reliability) for this task, but when looking to book flights to that country, the user might go to his or her preferred websites (Habit). Therefore, for next few queries of the session, "Habit" dimension becomes more important. On the other hand, the consideration of a particular dimension may only happens when the user interacts with the document. For example, initially a user has no dimensional preference but on reading a document which is poorly written (low Understandability), the user might give more consideration to the Understandability dimension for judging future documents. Thus, the importance given to relevance dimensions might change as the session progresses or tasks switch. By capturing the importance assigned to each dimension for a query, I can model the dimensional importance and use it to improve the ranking for the subsequent queries. There is a quantum analogue of this process - in terms of superposition principle and collapse postulate. Before a document is judged, it is not possible to know which dimension will be preferred by a user. All possibilities exist. On interacting with the document, the user's cognitive state collapses into a particular dimension or into a combination of dimensions (weak measurement).

3.2.2 Capturing User's Cognitive State during Information Interaction

In the previous section, I have discussed how to construct a Hilbert Space for a given query, documents, and relevance dimensions. Note that a collection of Hilbert spaces corresponding to different documents for a query is represented by the $N \times 7$ matrix α where N is the number of documents in the collection and 7 is the number of relevance dimensions in my model. Although an element of this matrix $\alpha[d][r]$ represents relevance of the document d with respect to the dimension r, I can easily calculate non-relevance because these are mutually exclusive (probabilities add to one as I am only considering binary relevance here).
Suppose such a space exists and from the query logs one obtains a list of SAT-clicked documents or relevance documents for a query in a session. Algorithm 2 shows how to use the available data to capture user's cognitive state with respect to dimensional importance. The input to the main procedure is the list of relevance dimensions, list of relevant or SAT-clicked documents for the query and the Hilbert spaces matrix α. Let d be one of the relevant documents for the query. The coefficients of superposition for $|d\rangle$ vector in a basis corresponding to a dimension r are - $\alpha[d][r]$ and $\sqrt{1-\alpha[d][r]^2}$. They represent the relevance and non-relevance of document d with respect to the dimension r. In QT, the probability of relevance of d with respect to r is calculated by taking the projection of $|d\rangle$ onto the respective basis by taking square of its inner product with the relevance vector of that basis. For example, $|\langle novelty|d\rangle|^2$, $|\langle reliability|d\rangle|^2$, etc. In the algorithm they are accessed by squaring $\alpha[d][r]$. (Step 6 in algorithm 2). Let $w_{d_1},...,w_{d_7}$ be the projections obtained. *I interpret them as weights assigned by the user to each relevance dimension while judging that document for a given query.* Where there are more than one SAT-clicked or relevant documents for a query (in that case the loop in step 5 will run more than once), I average over the projection scores for each dimension (steps 5-8). The procedure returns a list of seven weights which correspond to the average of the importance assigned by the user to the different relevance dimensions for a query.

Thus, for a given query in a search session, I have quantitatively captured the user's cognitive state. It is in terms of the user's preference for each dimension and is the average relevance score for that dimension over all SAT-clicked or relevant documents of the query. These weights are used to re-rank documents for the next query in the session, as explained in the next section.

Algorithm 2 Capturing weights given to relevance dimensions

procedure CAPTUREWEIGHTS(rel, $docsSAT$, α)
  for all r in rel **do** ▷ rel - list of 7 dimensions
    $totalWeight \leftarrow 0$
    $avgWeight[r] \leftarrow 0$ ▷ Variable to store user's weight to each dimension for a query
    for all d in docsSAT **do**
      $w_{dr} \leftarrow |\alpha[d][r]|^2$ ▷ Take projections ($|\langle r|d\rangle|^2$)
      $totalWeight \leftarrow totalWeight + w_{dr}$
    $avgWeight[r] \leftarrow totalWeight/|docsSAT|$ ▷ Only SAT clicks considered
  return $avgWeight$ ▷ Returns user's weight for each dimension for a query

3.2.3 Experiment and Analysis

I use the same datasets as used in [93], which were given to me on request by the first author of the paper. The first dataset is a query log of the Bing search engine and the second one is the combined session tracks of TREC 2013 and TREC 2014. While the Bing query logs contain information about each query in a session, the TREC dataset only contains the data about the last query for each session. The relevance criteria for Bing logs is SAT-clicks (more than 30 seconds spent by user on a document) and for TREC data I consider relevance grades of 1 and above to correspond to relevant documents (relevance grades are -2,0,1,2,3,4). In the previous subsection, I showed how to capture user's dimensional preference for a query in the form of weights. The essential part of that process involves re-ranking the original list of retrieved documents using the Learning to Rank algorithm to generate relevance scores. This is implemented using the open source library RankLib (`https://sourceforge.net/p/lemur/wiki/RankLib/`) with default settings. Using these scores, the Hilbert space is constructed (algorithm 1) and the weights assigned to each dimension for a query are calculated (algorithm 2). I now use these weights for the next query in the session, to take a weighted combination of the relevance scores of all seven dimensions for each document of the next query. Thus, for the new query, a new relevance score for each document is created based on the weighted dimensional preference for the previous query. I re-rank the documents according to these new scores and perform evaluation. I use the NDCG metric for evaluation and compare the values with those obtained in [93].

I also performed an initial analysis of the data to support my hypothesis that some combination of relevance dimensions are preferred by the user in a search session. For some randomly sampled 4837 sessions of the Bing query logs, I found that in 3910

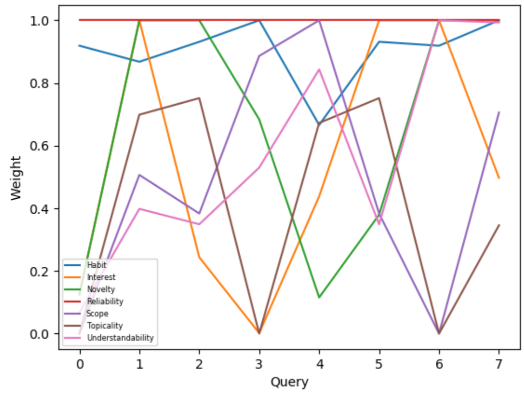


Figure 3.2: Random session for Bing data showing dynamics of dimensional preference

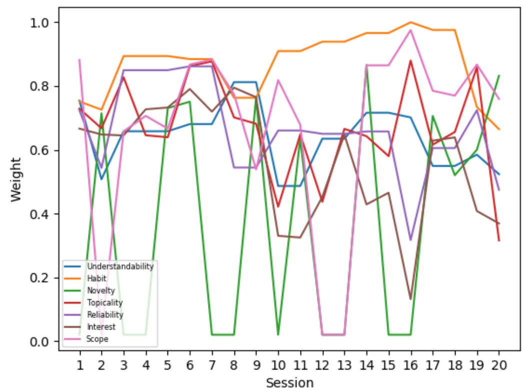


Figure 3.3: Random session for TREC data showing dynamics of dimensional preference

or 80.84 percent of the sessions, one of the top three dimensions for the first query of the session remains in the top three for all the queries of the session. This supports the theoretical model of a weak superposition or a partial collapse where among all possibilities of relevance dimensions, the user's cognitive state reduces to a subset which may evolve in a session. Figure 3.2 is the snapshot of one such session showing that the "Reliability" remains the top dimension throughout. Figure 3.3 shows 20 consecutive sessions for TREC data.

3.2.4 Results and Discussion

I summarize the evaluation results for Bing query logs in Table 3.2 and for the TREC session track 2013 and 2014 in Table 3.3. In the re-ranking algorithm proposed in [93], the best results are obtained by ranking according to the 'Reliability' dimension for Bing query log data and the 'Interest' dimension for TREC Session track 2013 and 2014 dataset. Therefore the algorithm suffers from a limitation that the system cannot know according to which dimension to re-rank the results in order to optimise the NDCG metric. For both the datasets, my proposed algorithm gives marginally better results. However, the advantage lies in the fact that this algorithm will generalise to different datasets. The system need not re-rank documents according to any particular dimension, but instead use the weighted combination. The weights themselves reflect user's dynamic preference for dimensions which is ascertained based on their current and past interactions.
It is to be noted that TREC data contains information about the last query of each session, and not all the queries. Thus the weighted approach uses the captured weights of the last query of a session to re-rank the documents for the last query of the next session. Improvement over the best result (corresponding to Interest) means that the weighted combination method for ranking works across sessions as well. This indicates that dimensional preference is not only dependent upon the task, but user might have an intrinsic preference for some dimensions as well. Also note that the "Topicality" scores correspond to a traditional baseline ranking model as I use tf-idf and BM25 as features for the "Topicality" dimension.

Dimension	NDCG@1	NDCG@5	NDCG@10	NDCG@ALL
Habit	0.3772	0.5958	0.6533	0.6645
Interest	0.4574	0.6178	0.6844	0.6955
Novelty	0.4110	0.6025	0.6688	0.6783
Reliability	**0.6457**	**0.7687**	**0.8038**	**0.8110**
Scope	0.2501	0.4692	0.56156	0.5726
Topicality	0.2001	0.4486	0.5352	0.5482
Understandability	0.2782	0.4968	0.5867	0.5971
Weighted Combination	**0.6552**	**0.7814**	**0.8127**	**0.8189**

Table 3.2: Bing logs evaluation

I have thus shown that capturing user's weights for relevance dimensions and ranking based on the combination of these weights leads to a better performance than using only

Dimension	NDCG@1	NDCG@5	NDCG@10	NDCG@ALL
Habit	0.0989	0.1406	0.1418	0.1592
Interest	**0.1981**	**0.2126**	**0.2242**	**0.1831**
Novelty	0.0966	0.1180	0.1316	0.1557
Reliability	0.1120	0.1333	0.1431	0.1614
Scope	0.1318	0.1526	0.1647	0.1671
Topicality	0.1459	0.1520	0.1887	0.1701
Understandability	0.1653	0.1913	0.1878	0.1764
Weighted Combination	**0.2364**	**0.2663**	**0.2729**	**0.1944**

Table 3.3: TREC data evaluation

one of the dimensions. The need for a Hilbert space is not explicit in this experiment. However, it is inspired by the fact that some relevance dimensions are incompatible for some documents. A document may not have high relevance weights for both "Novelty" and "Habit" dimensions at the same time. The more relevant it is in the "Novelty" dimension, the less relevant it will be in the "Habit" dimension. This is similar to the Uncertainty Principle in QT. I therefore model each relevance dimension as a different basis. For some documents, the basis might coincide, but in general there is incompatibility between relevance dimensions which leads to interference and order effects [32, 171]. For example, a user may find a document less reliable due to its source, but when the user considers the Topicality dimension and reads it, it might remove the doubts about the reliability. Thus "Topicality" interferes with "Reliability" in relevance judgement. Such order effects were investigated in [23] through user studies. I intend to investigate such cognitive phenomena in real world data through the quantum framework. The methodology reported in this experiment to construct a Hilbert space of document judgements is essential for investigating different quantum-like phenomena in the real-world query log data. I will use it to investigate the psychological phenomena of order effects in query log data, which corresponds to incompatibility and interference in QT. Order effects as the cognitive manifestation of incompatibility has been discussed in the literature review chapter.

3.3 Experiment 2 - Test of Order Effects in Relevance judgement with Query Logs

3.3.1 Introduction

In [22], order effect between different relevance dimensions was investigated via a user study. Participants were asked to judge pairs of relevance dimensions in two different orders. It was a between subjects design where one group judged a pair of dimensions in one order and another group in the other order. Hence, relevance probabilities could be calculated based on different orders of relevance dimensions, which could be tested whether they are statistically significant from each other. Such user study based approach, on the one hand, directly captures users' responses in a controlled environment, while on the other hand, is limited in terms of scalability and reflection of natural and real-world IR settings. In this experiment, I investigate the order effects in multi-dimensional relevance judgement with a real-world query log dataset. Specifically, the Bing query log dataset from the previous experiment is used.

3.3.2 Dimensional Profile

First, let us discuss the concept of a 'Dimensional Profile' as introduced in this experiment. Consider that for a query, each document is represented as a seven dimensional vector, where each component of the vector corresponds to the importance of the corresponding relevance dimension to the document. This vector forms the Dimensional Profile of the document. [1]

The similarity between two documents based on their Dimensional Profiles can be measured based on the relative difference between the values of each dimension for the two documents. For example, if the Dimensional Profile of document d_1 is given by the vector $[\alpha_{11},...,\alpha_{17}]$ and that of document d_2 given by $[\alpha_{21},...,\alpha_{27}]$, I get their relative differences as the vector $[\frac{|\alpha_{21}-\alpha_{11}|}{max(\alpha_{21},\alpha_{11})},....,\frac{|\alpha_{27}-\alpha_{17}|}{max(\alpha_{27},\alpha_{17})}]$. I then specify a matching criteria where each value in the difference array should be within that criteria. For example, matching criteria of value 0.05 means that the differences in corresponding scores for the relevance dimensions between the two documents are all within 5%. A matching criteria of 0 means both the documents have exact same scores for all the seven relevance

[1] Note that the usage of the word 'dimension' in 'seven dimensional vector' is in the geometric sense, i.e, a vector space of seven dimensions. The mention of 'dimensions' in 'relevance dimensions' is in the non-technical sense, similar to features, aspects, properties, or judgement criteria.

dimensions. The reason why the matching criterion applies to each dimension and not an aggregate over all dimensions because in this case I need to define similarity between documents in terms of relevance dimensions. The alternative criterion can be e.g. that the average of the values in the difference vector be less than the threshold value. But this can cause a pair of documents to be marked similar if they have no difference between scores of six dimensions but a large difference between the scores of a particular dimension, such that the average comes to be within the threshold. This can make the difference between the two documents with respect to that dimension perceivable to the user. These scores are calculated from the query log data in the same manner as in the previous experiment. Algorithm 1 is used to calculate the features and utilise the learning to rank algorithm to generate relevance scores with respect to each relevance dimension. Note that a row in $\alpha[d][r]$ serves as a dimensional vector for a give document d.

3.3.3 Experiment and Results

With query log data, there is only one relevance decision per document. Hence it is not possible to test judgements based on different orders of relevance dimensions. To do so, I have developed the following protocol:

First, I identify the subset of queries where the first two documents in the ranked list have a similar Dimensional Profile (based on some matching criterion as a threshold). This subset is named **similarFirstTwo (SFT)**. Next, from this subset I find out those queries where second document is SAT-clicked. This subset is named as **similarFirstTwoSecondClicked (SFTSC)**. Now, if the second document is clicked and its dimensional profile is similar to the first, rationally one would expect the user to have SAT-clicked the first document as well. Therefore I am curious to find whether there are cases where users do not SAT-click the first document but do so for the second document. Hence, among the queries of the set **SFTSC**, I find the subset of queries where the first retrieved document is not SAT-clicked. Indeed, such queries are found and this subset is labelled as **similarFirstTwoFirstSkipped (SFTFS)**.

Out of the total 152941 queries in the query log dataset used in this study, I found 170 queries where the top two documents have the same Dimensional Profile (Matching criteria of 0). Of these 170, in 25 queries I had only second document SAT clicked and not the first one. Although **SFT** does not form a significant fraction of the total queries analysed, they represent the set of those queries which can potentially exhibit user's quantum-like behaviour. The subset **SFTFS** forms 14.71% of the subset **SFT**, which

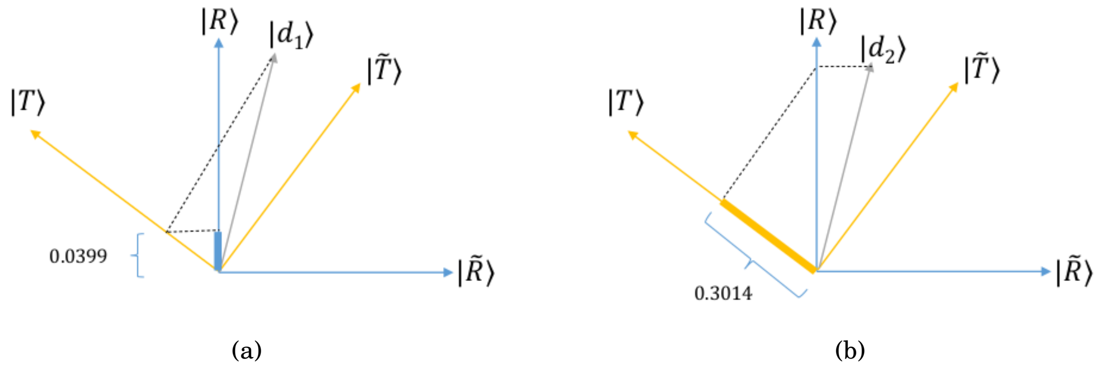


Figure 3.4: Different orders of relevance dimensions produce different judgements

is a significant fraction. Table 3.4 summarizes the results with different criteria of Dimensional Matching.

Matching Criteria	SFT	SFTSC	SFTFS	SFTFS percent(of SFT)
10%	309	44	40	12.94
5%	238	30	27	11.34
0%	**170**	**27**	**25**	**14.71**

Table 3.4: Analysis of Bing query log dataset

3.3.4 Quantum Cognitive Explanation of Observed Phenomenon

Having obtained results as shown in Table 3.4, I take one query from set $SFTFS$ and utilise Hilbert spaces of the first two documents to drive predictions about the user behaviour. Table 3.5 shows the Dimensional Profiles of the top two ranked documents for the query. The seven relevance dimensions of Habit, Interest, Novelty, Reliability, Scope, Topicality, Understandability are labelled as **H, I, N, R, S, T, and U** respectively. Here one can see that for both documents, the scores of all dimensions are exactly the same and the second document is SAT-clicked. I hypothesise that the dimension with the highest score is the preferred dimension for the query, which is Reliability in this case. Also, one can see that there are a few dimensions with very low scores. Let us construct a Hilbert space for Document 1 representing the basis for Reliability and Topicality. So we have:

Document rank	H	I	N	R	S	T	U
1	0.3040	0.1251	0.0000	0.9438	0.1250	0.1250	0.5619
2	0.3040	0.1251	0.0000	0.9438	0.1250	0.1250	0.5619

Table 3.5: Random Query Analysis

$$|d_1\rangle = 0.9715\,|Reliability\rangle + 0.2370\left|\widetilde{Reliability}\right\rangle \tag{3.4}$$

$$= 0.3535\,|Topicality\rangle + 0.9354\left|\widetilde{Topicality}\right\rangle \tag{3.5}$$

where, $0.9715 = \sqrt{0.9438}$ and so on. I take the Reliability basis as the standard basis. Representing Topicality basis in the standard Reliability basis, I get (Appendix A):

$$|Topicality\rangle = 0.5651\,|Reliability\rangle + 0.8250\left|\widetilde{Reliability}\right\rangle \tag{3.6}$$

Suppose that while judging Document 1, the user first considers the Topicality dimension and then considers the Reliability dimension (denoted here as Topicality -> Reliability). So initially the user's cognitive state is along the vector $|d_1\rangle$ and it changes to $|T\rangle$ and $|R\rangle$ as it evolves because of examining the document. Let's see what happens when we project from d_1->T->R in the Hilbert space as shown in Figure 3.4(a). The probability of this sequence is calculated as $|\langle T|d_1\rangle|^2|\langle R|T\rangle|^2 = 0.3535^2 * 0.5651^2 = 0.0399$. If the user reverses the order of relevance dimensions considered while judging document d_2, I get d_2->R->T = $|\langle R|d_2\rangle|^2|\langle T|R\rangle|^2 = 0.9715^2 * 0.5651^2 = 0.3014$, which is 7.5 times larger (Figure 3.4(b)). The joint probabilities are different for the two orders. There is a significant order effect predicted if one assumes a quantum-like structure employed in human decision-making (i.e. uses the quantum framework for predicting probabilities). Now since the Hilbert space for both Document 1 and Document 2 are same, I get these results if the user follows the order d_1->T->R for Document 1 and d_2->R->T for Document 2.

It is well accepted in IR that topicality is the biggest predictor of relevance. In fact, if a document has high weights for all other dimensions but is not topical at all, it is highly unlikely to be judged relevant. I have empirically shown this in Chapter 6. Now both documents here have a low score for topicality but a very high score for reliability. For the first document, the user considers topicality as the first judgement criteria and finds it less topical. Even though a subsequent consideration of reliability carries a very high score independently, the overall probability is very low as the low topicality also influences user's perception of reliability. For the second document, because reliability is considered first and it has a very high score, it influences user's perception about the topicality of the document. Hence the document appears more topical than it would when topicality is considered before reliability or the reliability was very low. Cognitively, this behaviour can be explained in terms of either an anchor bias where the first judgement influences subsequent judgements, or the Halo effect, which is also a cognitive bias. In the Halo effect, positive impressions about one attribute of an entity positively influences one's opinions or feelings about another attribute. So finding a highly reliable document interferes with the user's cognitive state so that he or she considers the document to be more topical than before. In this simple case, if the user does not consider other dimensions, the final inference of relevance would be higher for Document 2. This offers one plausible explanation why Document 2 ends up being SAT-clicked and not Document 1.

However, one important question to ask here is what causes the user to use two different orders for the two documents. To explain this, I suspect an Attraction effect as a type of Context effect [171]. Since the user considers Reliability as the last dimension when judging Document 1, there is a memory bias leading to an considering Reliability to judge the second document.

In the quantum framework, relevance dimensions are not objective properties of the document, which are judged independently of each other. Rather they are highly interdependent and contextual. As discussed earlier in the thesis, the quantum framework neatly embeds such contextual effects. Classical approach to model these would involve different heuristics for different types of effects.

3.4 Conclusion

The experiments reported in this chapter were the initial attempt to investigate the overarching question of the thesis - is there evidence of quantumness in document judgements. The research sub-question for this chapter was whether query logs provide us with such evidence. In the first experiment, I showed how to construct a Hilbert space from query log data. The Hilbert space is the building block of the quantum framework. It represents document judgements along different relevance dimensions and allows for the modelling of incompatible judgement perspectives should they exist in the data. In the re-ranking algorithm developed, I make use of the superposition principle and the collapse postulate. The user's cognitive state, where initially all relevance dimensions can be potentially considered, collapses or partially collapses to one or more dimensions on interacting with the documents and queries. The limitation of this consideration is that it is difficult to show the quantum advantage in this. It could very well be shown by a classical model where the users' consideration of a combination of relevance dimensions be governed by some hidden variables. Indeed, it is difficult to separate a superposition state from a classical mixture state in QT without performing sequential measurements. If one performs sequential measurements and they are incompatible, one would be able to see the difference between classical and quantum systems. Therefore in the second experiment, I investigate order effects which in QT are a result of sequential, incompatible measurements.

In the second experiment, a Hilbert space representation and in general incompatibility between relevance dimensions is assumed. A particular scenario is considered which shows unexpected user behaviour. Quantum probability calculations demonstrate an order effect between relevance dimensions which is proposed as a possible explanation of the unexpected user behaviour. However, there can be many different explanations of the behaviour, including randomness and noise as the number of queries showing such effects forms a small fraction. As such, evidence of superposition and incompatibility from the above two experiments is not substantial. What is needed is a more principled approach to proving quantumness in user behaviour data in IR. Therefore I turn to the methodology used in Quantum Physics to provide a more mathematical formulation of quantumness of a system, as discussed in two experiments in the next chapter.

CHAPTER 4

USING BELL-TYPE INEQUALITIES TO TEST FOR QUANTUMNESS OF USER RELEVANCE JUDGEMENT DATA

In the two experiments in the previous chapter, I have captured user's judgement of a document in different bases of a two dimensional Hilbert space. The two dimensions correspond to potential judgements of relevance and non-relevance of the document with respect to a query. An initial state vector induces a probability distribution onto the two potential judgements by way of the square of the inner product. Thus, I am assuming what in QT is called a superposition state. In such a state a document cannot be categorised into being relevant or non-relevant to a query before judged by a user. Before the interaction, one can only define a probability distribution which corresponds to the proportion of users who would judge the document relevant and non-relevant respectively. The different bases correspond to the different perspectives of judging relevance for a given document. Based on which relevance dimension is considered, the same document will have different probabilities of relevance and non-relevance. This is analogous to the measurement of electron spin which is either up or down in direction, but depends upon which axis it is measured in. Electrons with spin up along the Z-axis may have both up and down components along the X-axis. So a document may have a high probability of relevance based on the Topicality dimension but a different distribution along the Reliability dimension. This model of information interaction with different bases corresponding to different superposition states, predicts order effects should the users consider (sequentially) different judgement perspectives.

However, this analogy alone is not sufficient to warrant the need for a radical departure from traditional user state representations to a quantum-inspired model. One needs to formally prove the quantumness of the user judgements data (usually by proving the insufficiency of traditional approaches). In this chapter I discuss two experiments which apply certain type of inequalities from QT known as Bell-type inequalities to test for quantumness of multidimensional relevance judgement. The research question answered in this chapter is:

RQ 2: How to verify quantumness of IR data using no-go theorems of Quantum Theory?

The inequalities are constructed in such a way that they are obeyed by classical systems. Data generated by a quantum system violates these inequalities and introduces the need for a non-classical model of the system. I divide the research question into two sub-research questions for this chapter:

Sub-RQ1: How can we formulate relevance judgement tasks in terms of Bell-type inequalities?

Sub-RQ2: How to design relevance judgement tasks which fully exploit the hypothesised quantumness such that Bell-type inequalities are violated.

The common concept behind both the experiments in this chapter is the idea of contextuality in QT. Contextuality is regarded as the most fundamental difference between quantum and classical systems. The next section discusses the ideas behind contextuality which also clarifies my motivations behind the Hilbert space models and assumption of superposition in IR as discussed above.

4.1 Contextuality

Consider a system with properties p, a_1, a_2, b_1, b_2. Measurement outcomes of these properties are represented by random variables P, A_1, A_2, B_1, B_2. Let p be measured along with a_1 and a_2 in one experimental condition and measured along with b_1 and b_2 in another experimental condition, such that one can form joint probability distributions $Pr(P, A_1, A_2)$ and $Pr(P, B_1, B_2)$ respectively for the two conditions. Contextuality is defined as the impossibility of assigning a joint probability distribution to all the variables - $Pr(P, A_1, A_2, B_1, B_2)$ such that the marginal distributions obtained from it (e.g. $Pr(P, A_1, A_2)$) agree with those obtained experimentally. However, the condition is that the variables measured within each context should be compatible with each other, i.e., their joint probability distributions should be well-defined. In the traditional approach to probabilities, also referred to as the Kolmogorovian approach, we find that joint probability of events is always defined. However, there is a scenario where two separate order of events will lead to different joint probabilities. Although $P(A, B) = P(B, A)$ is assumed to be true in Kolmogorovian probability theory, the same cannot be said to be true always and in such a case the joint probability is said to be undefined and the two events called incompatible. So, in the definition of contextuality, P should be compatible with A_1, A_2 and also with B_1, B_2 and thus can be jointly measured. On the other hand, A_1, A_2 are incompatible with B_1, B_2 and therefore these two define different experimental conditions or contexts. Compatibility of events or properties also mean that the measurement of one property does not disturb or influence the outcome of the other property.

The impossibility of assigning the said joint probability distribution compatible with the marginal distributions is related to the fact that we do not measure all the variables P, A_1, A_2, B_1, B_2 together. One assumes that the value of the property p is pre-defined and remains the same whether it is measured alongside A_1, A_2, or measured in context of B_1, B_2. This assumption is what leads to the contradiction. The context of the measurement influences the value of measurement of a property, even though the properties measured along p within the context have no influence on its outcome (are compatible). Thus there is a latent or implicit influence which is different from all influences seen in real world measurement scenarios and is not a causal influence. Contextuality can also be defined as the impossibility of assigning a definite value to a measurement, independent of which other measurement it is performed along with. In other words, there do not exist pre-defined values to measurement of a property, and any such potential hidden values have to depend upon the context where they are probed or measured. This points to the fact that properties of contextual systems are inherently in-deterministic. The variance in responses obtained on measuring a property of a contextual system is not due to the ignorance of certain 'hidden variables', but rather is a fundamental feature of the nature of contextual systems. This is an argument against realism. Realism means that every property of a system has pre-defined values and measurement on the system merely reveals the value of the property. For example, our weight has a fixed value prior to measurement and the measurement device simply reveals that value to us. Therefore, it remains the same irrespective of whether it is measured along with our height or along with our waist size.

Contextuality has been originally discovered in Quantum Physics, where the first debate about realism was sparked by Einstein, Podolsky and Rosen in 1935 [60]. Einstein's argument was of a world where realism existed even for properties of microscopic particles. The fact that Quantum Mechanics predicted the contrary, showed that Quantum Mechanics as a theory could not provide a complete description of our world. He suspected that the departure from realism in Quantum Mechanics appears due to presence of certain hidden variables which we don't know about yet. A more complete theory could describe the hidden variables associated with the microscopic systems, allowing us to predict with certainty the outcome of measurements. It was in the 1960s that Kochen and Spekker and John Bell provided mathematical proofs showing that Quantum Mechanics was incompatible with hidden variable theories. In particular, Kochen and Spekker [85] provided geometric proofs that non-contextual assignment of values to orthogonal projectors forming a context leads to contradictions. Bell [17] showed that an inequality formed using assumptions of hidden variables is violated by quantum systems in certain cases. A more simplified and commonly used version of the Bell inequality was given by [41], known as the CHSH inequality. The CHSH inequality is given by Equation 4.1 for two systems A and B where properties A_1 and A_2 can be measured in system A and B_1 and B_2 can be measured in system B. A_i and B_i can take values only in $\{\pm 1\}$. At each measurement instance, one property A_1 or A_2 is measured of system A and simultaneously, one out of B_1 or B_2 is measured for system B. The two systems A and B are separated by large distances such that they cannot effect each other physically. Assuming a non-contextual or classical assignment of pre-defined values to A_i and B_i, such a system should obey:

$$|E(A_1B_1)+E(A_1B_2)+E(A_2B_1)-E(A_2B_2)| \leq 2 \tag{4.1}$$

where $E(x)$ stands for the expectation value of x. The intuition behind the inequality is that a variable, say A_1 assumes a particular value out of ± 1 independently of whether it is measured along with B_1 or B_2. Thus even though one can measure either A_1 and B_1 or A_1 and B_2 at a time (or A_2 and B_1 or A_2 and B_2), the expression $A_1B_1+A_1B_2+A_2B_1-A_2B_2 = \pm 2$ holds, and on averaging over repeated measurements we get the CHSH inequality in 4.1. However it is shown that an entangled system of spin-half particles (like electrons) violate this inequality for certain measurement of spins. The violation can be attributed to the presence of contextuality in the system. The value of A_1 is influenced by which measurement of system B - B_1 or B_2 is performed along with it, even though the two systems are separated by large distances to have any physical signalling effects.

4.2 Experiment 1 - Testing Violation of CHSH Inequality in Document Judgements

To recap, the CHSH inequality is given by Equation (4.2) for two systems A and B where observables A_1 and A_2 can be measured in system A and B_1 and B_2 can be measured in system B. A_i and B_i can take values only in $\{\pm 1\}$. It is assumed that the observables have pre-existing values which are not influenced by any other measurement.

$$|\langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - \langle A_2 B_2\rangle| \leq 2 \tag{4.2}$$

The CHSH inequality is violated in Quantum Mechanics using a special composite state of two systems, called the Bell state [108], which has the following form:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{4.3}$$

where $|0\rangle$ and $|1\rangle$ represent the standard basis for the two systems. Initially, both the systems are in a superposed state. The two outcomes, i.e., corresponding to the $|0\rangle$ and $|1\rangle$ vectors can be obtained with equal probabilities. However, on measuring one system, if one obtains the outcome corresponding to the basis vector $|0\rangle$, the state of the composite system collapses to $|00\rangle$. Now it is known for certain that the outcome of the second system also corresponds to $|0\rangle$. This is true even if the two systems are spatially separated - the measurement on one system reveals the state of the other, instantaneously.

Violation of Bell inequalities by such entangled states prove the impossibility of the existence of a joint probability distribution for the two systems. It rules out the concept of "Local Realism" of the classical world, which is the assumption made while deriving the Bell inequalities. Local stands for the fact that measurement of one system does not influence that of a spatially separated system. Realism assumes that values of physical properties of systems have definite values and exist independent of observation [108].

There have been several works which have investigated violation of Bell inequalities in macroscopic and cognitive systems [3, 8, 29]. This experiment also investigates the

Bell inequalities for violation by user's composite state for judgement of two documents. After describing the methodology used to quantify the seven relevance dimensions, I describe equivalent Bell inequalities for the document states. Subsequently I give details of the experimental settings used to form the composite system of documents.

4.2.1 CHSH Inequality for Documents

In previous experiments, I have calculated the relevance probabilities of a document for different dimensions. A Hilbert space is constructed for each document, consisting of seven different basis, representing each dimension of relevance. Two or more such documents can be considered as a composite system by taking a tensor product of the document Hilbert spaces. If $|d_1\rangle$ and $|d_2\rangle$ are the state vectors of two documents, one can represent the tensor product as $|d_1\rangle \otimes |d_2\rangle$. Figure 4.1 shows the geometrical representation of two such Hilbert spaces. Here $|R\rangle_{hab}$ represents Relevance in the Habit basis, or in IR terms, relevance of document d with respect to the Habit dimension. Similarly, $|\widetilde{R}\rangle_{hab}$ represents irrelevance in the Habit basis.

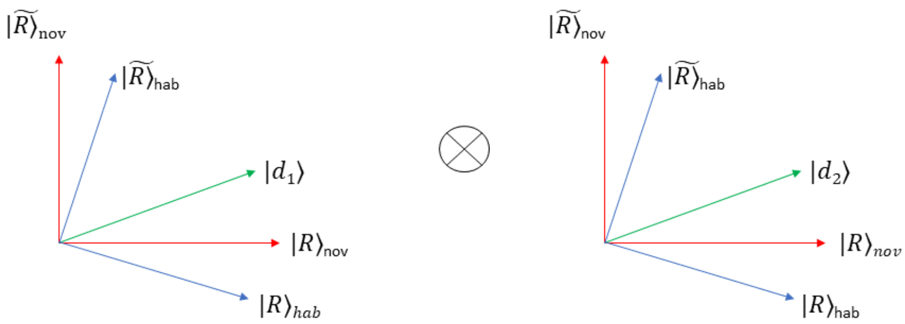


Figure 4.1: Tensor product space for two documents

In the CHSH inequality, we have observables A_1 and A_2 for a system taking values in ± 1. For a document d_1, we have observables corresponding to the different relevance dimensions. Taking the case of two relevance dimensions, Habit and Novelty, we have observables R_{hab} and R_{nov} which take values in ± 1. Where $R_{hab} = +1$ corresponds to a projection on the basis vector $|R\rangle_{hab}$, $R_{hab} = -1$ corresponds to the projection on its orthogonal basis vector $|\widetilde{R}\rangle_{hab}$.

Taking two documents as a composite system, we can write the CHSH inequality in the following way:

$$|\langle R_{hab1}R_{hab2}\rangle + \langle R_{hab1}R_{nov2}\rangle + \langle R_{nov1}R_{hab2}\rangle - \langle R_{nov1}R_{nov2}\rangle| \leq 2 \tag{4.4}$$

Where the subscripts 1 and 2 denote that the observables belong to document 1 and document 2 respectively. Using the fact that $\langle AB\rangle = 1 * P(AB = 1) + (-1) * P(AB = -1)$ and $P(AB = 1) + P(AB = -1) = 1$, we can convert the above inequality into its probability form as:

$$\begin{aligned} 1 \quad \leq \quad & P(R_{hab1}R_{hab2} = 1) + P(R_{hab1}R_{nov2} = 1) + \\ & P(R_{nov1}R_{hab2} = 1) + P(R_{nov1}R_{nov2} = -1) \quad \leq \quad 3 \end{aligned} \tag{4.5}$$

Assuming $P(AB) = P(A)P(B)$, we get:

$$\begin{aligned} 1 \quad \leq \quad & P(R_{hab1} = 1)P(R_{hab2} = 1) + P(R_{hab1} = -1)P(R_{hab2} = -1) + \\ & P(R_{hab1} = 1)P(R_{nov2} = 1) + P(R_{hab1} = -1)P(R_{nov2} = -1) + \\ & P(R_{nov1} = 1)P(R_{hab2} = 1) + P(R_{nov1} = -1)P(R_{hab2} = -1) + \\ & P(R_{nov1} = 1)P(R_{nov2} = -1) + P(R_{nov1} = -1)P(R_{nov2} = 1) \quad \leq \quad 3 \end{aligned} \tag{4.6}$$

As mentioned above, $R_{hab} = +1$ corresponds to the basis vector $|R_{hab}\rangle$ and therefore $P(R_{hab1} = 1)$ corresponds to the probability that document d_1 is relevant with respect to the $Habit$ dimension of relevance. Therefore one can calculate these probabilities as projections in the Hilbert space:

$$\begin{aligned} P(R_{hab1} = 1) \quad &= |\langle R_{hab}|d_1\rangle|^2 \\ P(R_{hab1} = -1) &= |\langle \widetilde{R_{hab}}|d_1\rangle|^2 \\ P(R_{nov1} = 1) \quad &= |\langle R_{nov}|d_1\rangle|^2 \\ P(R_{nov1} = -1) &= |\langle \widetilde{R_{nov}}|d_1\rangle|^2 \end{aligned} \tag{4.7}$$

and similarly for document d_2.

4.2.2 CHSH Inequality for documents using the Trace Method

Another way to define the CHSH inequality for documents is by directly calculating the expectation values using the trace rule. According to this rule, expectation value of an observable A on a state $|d\rangle$ is given by

$$\langle A\rangle = tr(A\rho) \tag{4.8}$$

where the quantity $\rho = |d\rangle\langle d|$ is the density matrix of the state $|d\rangle$

Let the two documents be represented in the standard basis as follows:

$$|D_1\rangle = a_1|H\rangle_1 + b_1\left|\widetilde{H}\right\rangle_1 \tag{4.9}$$
$$|D_2\rangle = a_2|H\rangle_2 + b_2\left|\widetilde{H}\right\rangle_2$$

where $|H\rangle_{1,2} = \begin{pmatrix}1\\0\end{pmatrix}$ and $\left|\widetilde{H}\right\rangle_{1,2} = \begin{pmatrix}0\\1\end{pmatrix}$

Hence, the the density matrix for a document $|D\rangle$ can be written as:

$$|D\rangle = \begin{pmatrix}a\\b\end{pmatrix} \qquad |D\rangle\langle D| = \begin{pmatrix}a^2 & ab\\ ab & b^2\end{pmatrix} \tag{4.10}$$

The document representations in another basis are as follows:

$$|D_1\rangle = c_1|N\rangle_1 + d_1\left|\widetilde{N}\right\rangle_1 \tag{4.11}$$
$$|D_2\rangle = c_2|N\rangle_2 + d_2\left|\widetilde{N}\right\rangle_2$$

H and N are basically relevance with respect to two relevance dimensions, say Habit and Novelty. One can write the N basis in terms of the H basis(see appendix) as:

$$|N\rangle_1 = (a_1c_1 + b_1d_1)|H\rangle_1 + (b_1c_1 - a_1d_1)\left|\widetilde{H}\right\rangle_1 \tag{4.12}$$
$$\left|\widetilde{N}\right\rangle_1 = (a_1d_1 - b_1c_1)|H\rangle_1 + (a_1c_1 + b_1d_1)\left|\widetilde{H}\right\rangle_1$$

and similarly for the second document.

Thus I get the vector representations for basis states $|N\rangle_1$ and $\left|\widetilde{N}\right\rangle_1$ as:

$$|N\rangle_1 = \begin{pmatrix}a_1c_1 + b_1d_1\\ b_1c_1 - a_1d_1\end{pmatrix} \qquad \left|\widetilde{N}\right\rangle_1 = \begin{pmatrix}a_1d_1 - b_1c_1\\ a_1c_1 + b_1d_1\end{pmatrix} \tag{4.13}$$

Now the observable $\boldsymbol{H}$ and $\boldsymbol{N}$ are defined as:

$$\boldsymbol{H} = |H\rangle\langle H| - \left|\widetilde{H}\right\rangle\left\langle\widetilde{H}\right| \tag{4.14}$$
$$\boldsymbol{N} = |H\rangle\langle N| - \left|\widetilde{N}\right\rangle\left\langle\widetilde{N}\right|$$

where $|H\rangle\langle H|$ and $|\widetilde{H}\rangle\langle\widetilde{H}|$ are the projection operators for standard basis vectors with eigen values 1 and -1 respectively. This is the spectral decomposition of the observables. We get $\boldsymbol{H} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The matrix for observable $\boldsymbol{N}$ is obtained in terms of the amplitudes a, b, c and d. Now the CHSH inequality for the observables $\boldsymbol{H}$ and $\boldsymbol{N}$ acting on the two documents can be written as:

$$|\langle \boldsymbol{H}_1\boldsymbol{H}_2\rangle + \langle \boldsymbol{H}_1\boldsymbol{N}_2\rangle + \langle \boldsymbol{N}_1\boldsymbol{H}_2\rangle - \langle \boldsymbol{N}_1\boldsymbol{N}_2\rangle| \leq 2 \tag{4.15}$$

Here $\boldsymbol{H}_1\boldsymbol{H}_2$ denotes the fact that I measure the observable $\boldsymbol{H}$ on both the documents. In the language of tensor products,

$$\boldsymbol{H}_1 \otimes \boldsymbol{N}_2 |D_1\rangle \otimes |D_2\rangle = \boldsymbol{H}_1 |D_1\rangle \otimes \boldsymbol{N}_2 |D_2\rangle \tag{4.16}$$

And,

$$\begin{aligned} \langle \boldsymbol{H}_1\boldsymbol{N}_2\rangle &= \langle D_1 \otimes D_2 | \boldsymbol{H}_1 \otimes \boldsymbol{N}_2 | D_1 \otimes D_2\rangle \\ &= \langle D_1 | \boldsymbol{H}_1 | D_1\rangle \langle D_2 | \boldsymbol{N}_2 | D_2\rangle \\ &= tr(\boldsymbol{H}_1 |D_1\rangle\langle D_1|) \times tr(\boldsymbol{N}_2 |D_2\rangle\langle D_2|) \end{aligned} \tag{4.17}$$

In this way I can directly calculate the expectation values in Equation (4.15). As a sample calculation, $tr(\boldsymbol{H}_1 |D_1\rangle\langle D_1|) = tr\left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} a_1^2 & a_1b_1 \\ a_1b_1 & b_1^2 \end{pmatrix}\right) = a_1^2 - b_1^2$, where a_1^2 and b_1^2 are the probabilities of relevance and non-relevance respectively in the standard basis.

4.2.3 N-Settings Bell Inequality

The CHSH inequality refers to two two-dimensional systems where each system has two measurement settings (or two measurement bases). However this can be generalised for systems with multiple settings or bases [71]

$$\sum_{j=1}^{n}\Big(\sum_{k=1}^{n+1-j} E(A_jB_k) \quad - \sum_{k=n+2-j}^{n} E(AjB_k)\Big) \leq \Big[\frac{n^2+1}{2}\Big] \tag{4.18}$$

where $[x]$ denotes the largest integer smaller or equal to x.

For seven relevance dimensions, $n = 7$ and the bound is 25. One can convert Equation (4.18) into its probability form as done in Section 4.2.1, or use the trace rule to directly calculate the expectation values as done in Section 4.2.2

4.2.4 Experiment and Results

Having obtained an equivalent representation of Bell inequalities in the above sections, I proceed to substitute the values in the inequalities and test for violation using relevance scores as calculated in the previous experiments. For each query, a user judges several documents to be relevant or non-relevant according to the Information Need. I investigate the correlations between these documents, with each document having multiple decision perspectives, using the Bell Inequalities. The following types of document pairs are considered to test for quantum correlations:

I) Consider those queries where only two documents are SAT clicked. Out of 55617 queries in our dataset, 1702 queries had exactly two SAT clicked documents. I consider a composite system of these two documents and measure (judge the relevance) along different bases (relevance dimensions) corresponding to each of the Bell inequalities described in subsections 4.2.1, 4.2.2 and 4.2.3

II) Consider those queries for which we have at least one SAT clicked document. Out of 55617 queries in our dataset, 52936 queries have at least one SAT clicked document. I then consider a composite system of this SAT clicked document with all the unclicked documents for the query(one by one) and measure(judge the relevance) along different basis(relevance dimensions) corresponding to each of the Bell inequalities described in subsections 4.2.1, 4.2.2 and 4.2.3.

In both cases, no violation of the Bell inequalities for any query is found. While case (I) corresponds to correlated documents and case (II) corresponds to anti-correlated documents, it is to be noted that I am taking a composite system by taking a tensor product of two document states. This, in turn is separable back into the two document states. The reason why Quantum Mechanics violates Bell Inequalities is due to the existence of non-separable states like the Bell States. To get something similar to an entangled state, I consider another type of document pairs:

III) Consider a pair of documents which are listed together for many queries, but are always judged in a correlated manner. That is, if one document of the pair is SAT clicked, the other one is also SAT clicked for that query. And similarly both might be unclicked for another query in which they appear together. Also, I find those documents which are SAT clicked together in half of the queries they occur in, and unclicked in the other half.

This corresponds to the following Bell State:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|RR\rangle + |\tilde{R}\tilde{R}\rangle) \tag{4.19}$$

Such pairs of documents are considered to test the Bell inequalities. Out of 774 pairs of documents, no pair show the violation of the inequalities discussed above.

The composite state of the two documents described in equation(4.19) appears to be like an entangled state of the documents - knowing that one document is SAT clicked or not can tell us about the other document. However, one fundamental property of the Bell states is their rotational invariance. Representing a Bell State in any basis, one gets the same probabilities of the two possible outcomes. For example,

$$\begin{aligned} |\psi\rangle &= \frac{1}{\sqrt{2}}\left(|HH\rangle + |\tilde{H}\tilde{H}\rangle\right) \\ &= \frac{1}{\sqrt{2}}\left(|TT\rangle + |\tilde{T}\tilde{T}\rangle\right) \end{aligned} \tag{4.20}$$

where H, N and T are relevance with respect to the Habit, Topicality and Novelty basis. One can always hypothetically construct document Hilbert spaces in such a manner that the composite state is rotationally invariant, but that is not the case in the query log data, which is the target of our investigation.
As a formal test of non-separable states, I performed Schmidt decomposition [108] of the composite system of document pairs. I did not find any evidence of a separable states for any type of document pairs, as described in cases (I), (II) and (III).

4.2.5 Discussion

This experiment is successful in answering the first research question of this chapter (**Sub-RQ1**) by showing how to formulate multidimensional relevance judgement in terms of Bell inequalities. Despite the presence of incompatible measurements (relevance dimensions), the inequalities are not violated in this experiment. Hence there might exist a joint probability distribution governing user's cognitive state for a pair of documents. The experiments in which the violation of Bell inequality has been reported for cognitive systems, the users are asked to report their judgements on composite states [7]. Hence the joint probabilities can be directly estimated from the judgements. This may result in a "Conjunction Fallacy" [156] due to incompatible decision perspectives, thus violating the law of total probability by overestimating the joint probability, and therefore violating the Bell inequality. In the dataset used, there are no judgements over the document pairs. That is, the user does not judge a pair of document to be relevant with respect to some dimensions. Instead I have got the probabilities of relevance of a single document with respect to different dimensions. When I use the relevance probability of individual documents to compute the joint probabilities for a pair of documents, I am forced to assume the existence of a joint probability distribution. Thus there might be a possibility of Bell inequality violation if one can obtain data for a pair of documents.

4.3 Change of Experimental Methodology

The conclusions from the above experiment mean that IR datasets maybe insufficient to test for quantum-like phenomena. An important property of quantum-like phenomena is that they are highly contextual. Measurement itself makes up a context. So different ways of measurement, for example, different orders of measuring a system will form different contexts. In the standard IR datasets or test collections, the context is not captured at all. Every measurement, be it relevance or even other dimensions of relevance are considered as fixed, objective properties of documents. It is not possible to find out from these datasets or for that matter from query logs, what order of relevance dimensions was considered by users or what would happen to the judgement of relevance if the order of documents in the rank list is changed. Without this information, it becomes difficult to test for quantumness of the data.

Therefore, it is imperative that such contexts are captured and the only way to do it is to not use the static datasets but design one's own experiments and collect data. This way one can have different groups of users based on different judgement contexts. Contextual influences can be established with statistically significant differences in measurements among the groups. It would require a large enough sample of users in each group to test for statistical significance. For e.g., if I use categorical variables like relevance/non-relevance to measure judgements, a chi-square test of equality of proportions would require at least 20 participants in each group [72]. Having at least two groups to test for differences due to two contexts will makes it difficult to conduct laboratory based user studies due to a requirement of a large number of participants. Therefore, henceforth in my doctoral research, I decide to go for conducting experiments online and collecting data using crowdsourcing.

Also, the interpretation of the probability of relevance needs to be changed. Current probabilistic models consider a document sample space and the probability of relevance

is computed as a function of query-document statistics only. Another possible interpretation is to assume a sample space of users' relevance judgements and the probability of relevance of a document to a query is a frequentist interpretation of the proportion of users who consider the document as relevant. Probability of relevance calculated as a proportion can help in testing for statistical significance among groups.
In the next section, I describe another experiment to utilise Bell-type inequalities in detecting quantumness of relevance judgement data. However, there are two key differences from the first experiment of this chapter. Firstly, the inequality used is a modified version of the CHSH inequality, modified specifically for experiments on human decision-making. Secondly, the methodology for collecting data is different - it is collected through crowdsourcing using a custom designed experiment.

4.4 Experiment 2 - Contextuality-by-Default Theory and Relevance Judgements

Contextuality has been the subject of extensive research in the Quantum Information Science community, due to it being a useful resource for computations. There have been different frameworks developing the mathematical structure of the theory of contextuality, notably the Sheaf Theoretic formulation by Abramsky and Brandenburger [2], Operational Contextuality by Spekkens [150], Graph Theoretic approach [36] and the Contextuality-by-Default theory [52, 54–57]. However, it is being increasingly hypothesised to exist outside of Physics, e.g. in Relational Database Theory [1] and Human Decision-making [59]. For my next experiment, I make use of the Contextuality-by-Default Theory, which is used by cognitive scientists to detect contextuality in human decision-making. The next subsections describe the details and some applications of this theory.

4.4.1 Contextuality-by-Default Theory

Contextuality is a fundamental property of a system of random variables. It need not be confined only to microscopic particles of the world of quantum physics. It is rather reflected in the data generated out of a particular measurement scenario. Hence there is no restriction on its potential to be found outside of physical systems. In particular, cognitive systems share an important property with Quantum systems - they both are highly in-deterministic and subject to influences of context. Hence, it is no surprise that contextuality has been heavily investigated in cognitive science [5, 10, 26, 27, 30] However, most of the experiments have failed to show contextuality [53, 58, 59, 162]. This is because human decision making is highly susceptible to direct or explicit context effects (causal).

The CHSH inequality and all the Bell-type inequalities are constructed under the assumption of no-signalling. The two systems under measurement are located far apart from each other so that the measurement performed on one system cannot physically or directly influence the outcomes of measurement performed on another system. Hence the probability distribution of A_1 is not affected by whether it is measured along with B_1 or B_2. There are two ways to deal with the presence of signalling in cognitive systems: 1) Design the experiments in such a way that the signalling is eliminated. 2) Modify the CHSH inequality such it accounts for presence of signalling as noise. It seems it is near impossible to eliminate the signalling present in human decision making. This is primarily because of the fact that human decisions take place at a cognitive level. It is not possible to identify, let alone control for, every variable when working with a cognitive system.

The Contextuality-by-Default theory (CbD) takes the second approach. The theory posits that context influences in decision making can be a mixture of both implicit contextuality and direct influences (explicit contextuality). The first thing that it introduces is the concept of contextual identification of a random variable. A random variable R_p measuring a property p is uniquely identified not just by the property it measures but also the context in which it is measured. A random variable measuring the same property in two different contexts c_1 and c_2 is rather considered as two random variables R_p^1 and R_p^2. Hence for the CHSH inequality, the notations for the random variables measured together by Alice and Bob A_i and B_j will be written in CbD as A_i^j and B_j^i, which is read as A_i measured in the context of B_j and vice-versa. Hence A_1^1 is not the same random variable as A_1^2. This is how they are defined in this theory (hence the name 'default'). The two random variables R_p^1 and R_p^2 are stochastically unrelated - they do not possess a joint probability distribution. All variables measured under the same context possess a joint distribution. Direct influences, or signalling, is defined as the difference in the probability distributions of the random variables R_p^1 and R_p^2. The context of measurement causes this difference in distributions. This direct influence is not just present across contexts. It is of the same nature as the influence on R_p due to the property p itself. To understand this statement, suppose people are asked whether they liked a movie A in two different contexts - whether they saw it at a movie hall or on their laptops. The random variables are R_A^{hall} and R_A^{laptop} with two values 'Yes' and 'No'. Certainly, the context of a movie hall or the laptop may affect the distribution of the 'Yes' and 'No' values. This is called direct signalling because the people are conscious of the information provided by the different contexts. But this direct signalling across contexts is similar to the effect of the particular movie itself on the distribution of a particular variable (i.e. within a context). If there is a different movie asked about, the distributions within a context may also be different.

The CbD theory modifies the CHSH inequality by taking into account the direct influences. No-signalling is mathematically referred to as marginal selectivity, which is a term used to describe the fact that the probability distributions of random variables measuring the same property do not change across different contexts. For the CHSH setup, the marginal selectivity is given by the term Δ as:

$$\Delta = |E[A_1^1] - E[A_1^2]| + |E[A_2^1] - E[A_2^2]| + |E[B_1^1] - E[B_1^2]| + |E[B_2^1] - E[B_2^2]| \tag{4.21}$$

which measures the change in expectation values of the random variables A_i and B_j when measured in context of measuring respectively B_j and A_i along with them simultaneously ($i, j \in \{1,2\}$). $\Delta = 0$ represents the situation where there is no signalling/direct influence involved. This is a case of 'pure' contextuality as witnessed in Quantum Physics. However, according to the CbD theory, it is possible to have contextuality present on top of direct influences and the following modified CHSH inequality is constructed to detect contextuality mixed with direct influences:

$$\begin{aligned} &|E(A_1^1B_1^1) + E(A_1^2B_2^1) + E(A_2^1B_1^2) - E(A_2^2B_2^2)| - \Delta \leq 2 \\ &\implies |E(A_1^1B_1^1) + E(A_1^2B_2^1) + E(A_2^1B_1^2) - E(A_2^2B_2^2)| - |E(A_1^1) - E(A_1^2)| \\ &\qquad - |E(A_2^1) - E(A_2^2)| - |E(B_1^1) - E(B_1^2)| - |E(B_2^1) - E(B_2^2)| - 2 \leq 0 \end{aligned} \tag{4.22}$$

This EPR/CHSH design is based on a four dimensional quantum system. There are four contexts involved here. The CbD inequality for a general n-dimensional system is written as [90]:

$$s_{odd}(E[R_1^2R_2^1], E[R_2^3R_3^2], ..., E[R_n^1R_1^n]) - (n-2) - \Delta > 0 \tag{4.23}$$

with $\Delta = |E[R_1^2] - E[R_1^n]| + |E[R_2^1] - E[R_2^3]| + ... + |E[R_n^{n-1}] - E[R_n^1]|$.

The quantity $s_{odd}(x_1,...,x_n) = \max(\pm x_1 \pm ... \pm x_n)$, where "each $\pm$ is replaced by a + or a − and the maximum is taken over all the choices that contain an odd number of minus signs" [16]. For example, $s_{odd}(a,b) = \max(-a+b, a-b)$. I hereby refer to Equation 4.23 as the CbD inequality. For $n = 4$ and $\Delta = 0$, we get the CHSH inequality, which can be applied in cases where there is no-signalling present, as in Physics experiments. As discussed above, it is near impossible to eliminate direct influences in cognitive systems and therefore accounting for their presence in the system is essential before proceeding to test for Quantum-like effects in cognitive science. The CbD inequality, along with the CbD theory provides a mathematically sound way to do so.

4.4.2 Successful applications of CbD

The first successful application of CbD in human decision making, i.e., where contextuality is detected, was made in [39]. The main reason this experiment was successful was because it was designed in such a way so as to make the pure CHSH part (the s_{odd} value) of 4.23 ($n = 4$) attain the maximum possible value, 4. This was achieved by enforcing a particular experimental design. With the CHSH part of the inequality equal to 4, the presence or absence of contextuality depends upon the value of Δ being less than 2, which was achieved in a crowdsourcing experiment where participants were asked binary valued questions in different contexts. The questions asked were based on the story 'The Snow Queen' by Hans Christian Anderson (hence the name of the experiment was 'The Snow Queen Experiment'). Four characters from the story were considered, namely, Gerda, Troll, Snow Queen, and Old Finn Woman. Each were associated to one of four characteristics - Beautiful, Unattractive, Kind, and Evil. Each participant was asked a question from one out of four contexts. Given some information of the story line, participants were shown two character names and two characteristics and were asked to select one character from the two character choices and the corresponding characteristic consistent with the story line. If all the users match the correct pair, the pure CHSH part of 4.22 would be equal to 4. Table 4.1 shows the different contexts formed using the character and characteristic pairs. For example, one question was to choose between any one character between $Gerda$ and $Troll$ and then choose their characteristic in the story from the characteristic pair. A particular pair $Beautiful$ and $Unattractive$ formed one context, while changing the characteristic pair for the same character pair formed another context of the experiment. Table 4.2 shows the correct matches according to the story-line.

Table 4.1: Snow Queen Experiment Design

	Character Choice	Characteristic Choice
Context 1	Gerda Troll	Beautiful Unattractive
Context 2	Gerda Troll	Kind Evil
Context 3	Snow Queen Old Finn Woman	Beautiful Unattractive
Context 4	Snow Queen Old Finn Woman	Kind Evil

Table 4.2: Snow Queen Experiment Correct Matches

Character	Correct Characteristic Match
Gerda	Beautiful, Kind
Troll	Unattractive, Evil
Snow Queen	Beautiful, Evil
Old Finn Woman	Unattractive, Kind

Analogous to the Bell-type experiments, the measurement settings A_1 and A_2 represents the result of selecting a character from the two character pairs. So, A_1 corresponds to selecting a character from the first character pair with $A_1 = 1$ corresponding to $Gerda$ and $A_1 = -1$ corresponding to the choice $Troll$. Similarly A_2 corresponds to the choice between the other character pair. In the same way, the values to B_1 and B_2 are assigned based on which characteristic of the two characteristic pairs is selected. $B_1 = 1$ and $B_1 = -1$ stand for choices of $Beautiful$ and $Unattractive$ respectively from the first pair. For all correct choices of the character-characteristic pairs as listed in Table 4.2, we get probabilities in each context in the form given in Table 4.3. The modified CbD inequality for the given experiment is same as 4.22:

$$|E(A_1^1B_1^1)+E(A_1^2B_2^1)+E(A_2^1B_1^2)-E(A_2^2B_2^2)|-|E(A_1^1)-E(A_1^2)| \\ -|E(A_2^1)-E(A_2^2|-|E(B_1^1)-E(B_1^2)|-|E(B_2^1)-E(B_2^2|\leq 2 \tag{4.24}$$

Writing in terms of probabilities, we have

$$\begin{aligned} E(A_1^1B_1^1) &= 1*P(A_1^1B_1^1=1)+(-1)*P(A_1^1B_1^1=-1) \\ &= 1*(P(A_1^1=1,B_1^1=1)+P(A_1^1=-1,B_1^1=-1)) \\ &\quad -1*(P(A_1^1=1,B_1^1=-1)+P(A_1^1=-1,B_1^1=1)) \\ &= 1*(p_1+1-p_1)-1*(0+0) \\ &= 1 \end{aligned} \tag{4.25}$$

Similarly we have:

$$\begin{aligned} E(A_1^2B_2^1) &= E(A_2^1B_1^2)=1 \\ E(A_2^2B_2^2) &= -1 \end{aligned} \tag{4.26}$$

Thus one gets three perfect correlations and one anti-correlation for the different pairs of random variables measured in same contexts. Thus we can get the pure CHSH part of the modified CHSH inequality (eqn. 4.22) as $E(A_1^1B_1^1)+E(A_1^2B_2^1)+E(A_2^1B_1^2)-E(A_2^2B_2^2)=1+1+1-(-1)=4$. This gives a huge leverage over marginal selectivity violations (direct influences) and we get Δ values low enough for the CbD inequality ($n=4$, or the modified CHSH inequality (eqn. 4.22)) to be violated.

Another series of experiments which report presence of contextuality via violation of CbD inequality is reported in [16]. While the Snow Queen Experiment uses four contexts, this experiment used both three context and four context experiments. Here I give an example of a three context experiment from this paper as it forms the basis of our own experimental design. Participants were asked to choose options for different meals, breakfast (R_1), lunch (R_2) and dinner (R_3). For each meal, they had the option of a high calorie dish ($R_i=1$) or a low calorie dish ($R_i=-1$). These three meal choices formed three cyclic contexts - $\{R_1,R_2\}$, $\{R_2,R_3\}$ and $\{R_3,R_1\}$ and each participant was shown one context. For example, participants assigned context 1 were required to choose one dish for breakfast and one dish for lunch. The constraint was that both dishes could

Table 4.3: Experiment Design for Relevance Judgements

	Beautiful($B_1^1 = +1$)	Unattractive($B_1^1 = -1$)
Gerda($A_1^1 = +1$)	p_1	0
Troll($A_1^1 = -1$)	0	$1 - p_1$
	Kind($B_2^1 = +1$)	Evil($B_2^1 = -1$)
Gerda($A_1^2 = +1$)	p_2	0
Troll($A_1^2 = -1$)	0	$1 - p_2$
	Beautiful($B_1^1 = +1$)	Unattractive($B_1^1 = -1$)
Snow Queen($A_2^1 = +1$)	p_3	0
Old Fin Woman($A_2^1 = -1$)	0	$1 - p_3$
	Kind($B_2^1 = +1$)	Evil($B_2^1 = -1$)
Snow Queen($A_2^1 = +1$)	0	p_4
Old Fin Woman($A_2^1 = -1$)	$1 - p_4$	0

not be high calorie or low calorie. Both had to be different. This is how anti-correlation was enforced in the design. Note that for 3-cyclic contexts, $n = 3$ and inequality 4.23 gives $S_{odd} = 3$ for perfect anti-correlations. For this experimental design with 3-cyclic contexts, one gets the probabilities in the form as shown in Table 4.4. In the table, R_i^j means that the random variable R_i is measured along with R_j. Again, the value of Δ obtained is low enough for the violation of CbD inequality, thus proving the existence of implicit contextuality in another unique type of experiment involving human decision-making.

	$R_2^1 = 1$	$R_2^1 = -1$
$R_1^2 = 1$	0	p_1
$R_1^2 = -1$	$1 - p_1$	0

	$R_3^2 = 1$	$R_3^2 = -1$
$R_2^3 = 1$	0	p_2
$R_2^3 = -1$	$1 - p_2$	0

	$R_1^3 = 1$	$R_1^3 = -1$
$R_3^1 = 1$	0	p_3
$R_3^1 = -1$	$1 - p_3$	0

Table 4.4: Probabilities for the three cyclic contexts of the Meal experiment

4.4.3 Experiment

Thus we see that it is possible to show quantum-like nature of human decision-making via violation of the CbD inequality constructed using the Contextuality-by-Default theory. Since relevance judgement is fundamentally a decision-making cognitive process, it is possible to test for violations of CbD inequality for relevance judgements. A violation of the CbD inequality will prove that relevance judgements have an implicitly contextual nature. This would mean that relevance judgements cannot be pre-defined and they are not controlled by some combination of latent variables. Rather, they are constructed at the point of interaction with the document and the user and subject to the dynamic contexts. Also, we would require to use the quantum mathematical framework to model relevance judgements in IR.

In this section, I report an experiment in which I designed a specific relevance judgement scenario in line with the 3-cyclic context design in [16]. The relevance judgements data was collected via a crowdsourced user study. The main components were designing a survey, distributing to participants to collect data and then analysing the data to check for inequality violations. Appendix B refers to the details of the user study including the process of seeking ethical approval. The next subsections describe the various aspects of the user study.

4.4.3.1 Participants

I recruited 241 participants for the study using the online crowd-sourcing platform Prolific (`prolific.co`). Prolific is widely used by researchers to conduct experiments and post surveys to a wide variety of participants. The only pre-screening criterion for participating in the study was a cut-off of 96 percent approval rate. Approval rate for a participant in Prolific is the fraction of submitted responses approved. The participants were paid at a rate of £7.03 per hour, which amounted to £0.82 per participant for the time taken to complete the study. The questionnaire was designed using the Qualtrics platform licensed by The Open University (`https://openss.eu.qualtrics.com`). Proper consent was sought to make use of the responses provided by participants in data analysis and research publications. Participants were also informed that data protections laws are being complied with. The study was approved by The Open University's Human Research Ethics Committee with reference number HREC/3063/Uprety. The details of the ethical approval process and the forms used in the ethical approval applications are attached in Appendix B.

Query 1 Description:	**hawaiian volcano observatories** I am looking for the history of and a summary of the work performed at the Hawaiian Volcano Observatories.
Query 2 Description	**hurricane Irene flooding in manville nj** How has the flooding that resulted from hurricane Irene affected Manville, NJ?
Query 3 Description	**frank lloyd wright biography** Find biographical information for Frank Lloyd Wright.

Table 4.5: Selected queries and their descriptions

4.4.4 Design

I considered three different queries to test the presence of contextuality. My experimental design was similar to the 3-cyclic context design in [16]. I chose the three queries from TREC 2013 Webtrack dataset. The queries chosen where of type 'single' (topic numbers 228, 232 and 239). Users were presented the topic description to inform them of the underlying information need, as well as the query terms. Table 4.5 lists the queries and the topic descriptions. Three document snippets were selected for each query. Documents were paired to form three contexts and thus gave three cyclic contexts. For example, for documents D_1, D_2 and D_3, I created three contexts - $\{D_1,D_2\}$, $\{D_2,D_3\}$ and $\{D_3,D_1\}$. Note that a user was shown only one context per query because asking the user the questions in all the three contexts would enforce a joint probability distribution of the relevance of all the three documents in the user's cognitive state. Therefore, all 241 participants were randomly assigned to each context resulting in 80, 80 and 81 participants answering questions pertaining to the three contexts for each query. The documents chosen for judgements were the snippets taken from the pages of the Google search engine (`www.google.com`) for the respective queries.

The most important reason why [39] and [16] were able to demonstrate a violation of the modified CHSH inequality is because they are able to maximise the correlations between the random variables under measurement, which overpowers the direct, signalling influences. [39] enforces this by asking the users to choose one option from the character pair and one from the characteristic pair. [16] do so by restricting users to choose different types of options in their two choices (for example, one low calorie food and one high calorie food, etc.). For the three dimensional (i.e. the 3-cyclic context) scenario, this ensures that the random variables representing the responses are always anti-correlated, thus giving the maximum value of 3 to the quantity s_{odd}. In this study, I

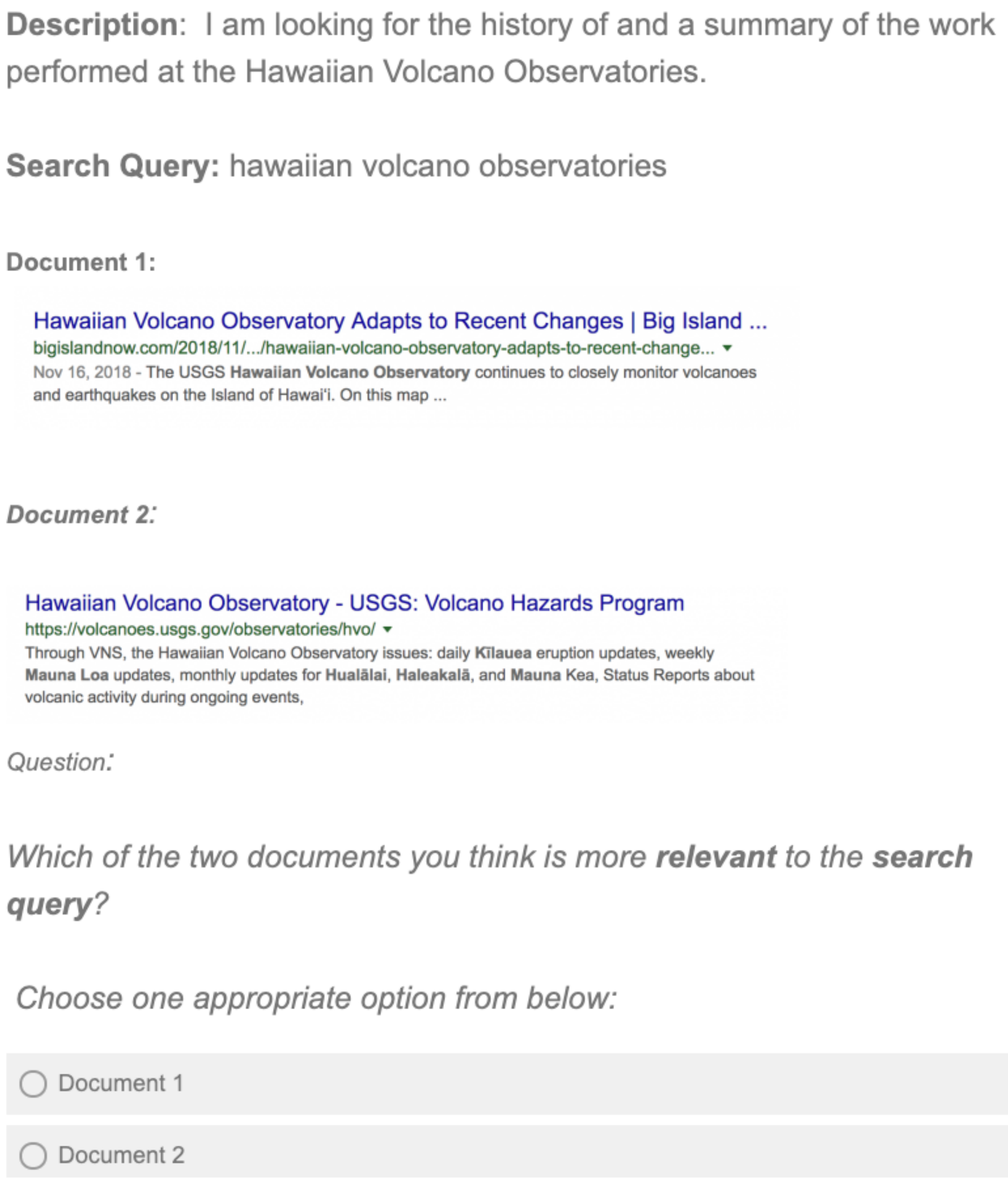


Figure 4.2: Snapshot of a study question

achieve this condition by asking users to choose one document out of the two shown that they think is more relevant for the query in question. Thus for each query, there will always be one document selected by the user (as more relevant) and the other will be automatically considered as less relevant. In this way, I ensure an anti-correlation which makes $s_{odd} = 3$. Figure 4.2 shows a snapshot from the study of a question asked to the users. Asking the users to choose the more relevant document also serves another purpose - to capture ambiguity in the decision making process. This means that the internal representation of the user's cognitive state can be considered a superposition. However, the ambiguity will be maximum when both the documents have similar relevance and it is difficult to decide which one is more relevant. It is here that I expect to capture the inherent randomness of user responses.

4.4.5 Simultaneity and Order Effects

A major condition in designing EPR/Bell-type experiment designs is that the measurement performed, in our case the questions asked, should be simultaneous. This is to negate all types of signalling which may happen when questions are asked in a sequence so that the information from one question can effect how the second question is answered. Order effects are a common phenomena in human decision making where the order of information presented has an influence on the final decision. Consider an example where, for a search query "Albert Einstein Research Areas", the topics of documents presented are, 1. Physics and 2. Theory of Relativity, in this order, and the user has to quickly assess them in this sequence. The relevance score given to the document about Physics will be much lower if the order of documents is reversed and the user comes across the document about Theory of Relativity first. A more relevant document affects the user's judgement for the second document. Such order effects in relevance judgement have been demonstrated before in [21, 164, 172].

Order effects occurs in QT as a result of incompatible measurements and interference of a measurement on the other. However, contextuality in QT is tied to the in-determinism or non-realism aspect of quantum systems and thus the Bell-type inequalities assume the measurements to be simultaneous so as to prevent any signal to impact the measurements. Order effects are a form of signalling and there one needs to ensure that they do not contribute to the violation of Bell-type inequalities.

In my experiment, by design, I ask participants to choose the more relevant document of the two presented. The cognitive analogue of measurement of a quantum system is asking a question to a user. My argument is that asking a single question about the two documents can be considered as a simultaneous measurement. However, it is discussed in [27] that the measurement analogue in cognitive science is not exactly when a user is asked a question, rather when the user consciously considers the question, in a sense taking their own measurement. Considering this, asking about which of the two documents is more relevant to the query might trigger the consideration of relevance of each of the two document sequentially in the user's mind. Now, because the document information is presented in a particular order to the user, the user might consider the same order while evaluating the relevance of the documents in his or her mind.

Therefore, I also tested for order effects in the presentation order of document snippets. For each of the three contexts for a query, half of the users were shown one order of document snippets and the other half another order. The probability of 'more relevance' or 'less relevance' answers of two documents of a context, in different orders, are shown in Table 4.6 for Query 1. Here m and l stand for the probability that the document is respectively more relevant or less relevant of the two. It is calculated by dividing the number of users who chose the particular answer (say $D_1 = m$ and (hence) $D_2 = l$) by the total number of users who were shown the question for this order of document snippets. For Order Effects to exist, the marginal probability of answering D_1 as more relevant in one order should not be different from answering $D_1 = m$ in the opposite order. Thus $P(D_1 = m)$ in the order $\{D_1, D_2\}$ is:

$$\begin{aligned} P(D_1 = m) &= P(D_1 = m, D_2 = m) + P(D_1 = m, D_2 = l) \\ &= 0 + 0.775 \\ &= 0.775 \end{aligned} \tag{4.27}$$

Similarly, in the order $\{D_2, D_1\}$ of consumption of information by the user, we have $P(D_1 = m) = 0.707$. On performing a chi-square test for equality of proportions ($\alpha = 0.05$, $n_1 = 40, n_2 = 41$), I find that this is not a statistically significant difference for Order Effects to exist. Indeed I find in this way that no order effects exist for any pair of documents in any of the three queries. Hence it is safe to approximate that the cognitive measurement performed in this EPR/Bell-type scenario is simultaneous.

4.4.6 Results

Tables 4.8, 4.9 and 4.10 report the probabilities obtained for the relevance judgements in the three contexts for queries 1, 2 and 3 respectively. Note that instead of denoting the relevance of documents as more relevant (m) or less relevant (l), I denote using $+1$ and -1. This is done just to be consistent with the notation used in literature and helps in calculation of expectation values. ± 1 merely denotes mutually exclusive values of a measurement. Given the fact that the user has to choose one document in this case, the judgement of the two documents can be considered as mutually exclusive outcomes. As discussed above, the value of Δ for contextuality to be present in our 3-cyclic system

Context 1: Order D_1, D_2 (rows D_1, columns D_2)

	m	l
m	0	0.775
l	0.225	0

Context 1: Order D_2, D_1 (rows D_2, columns D_1)

	m	l
m	0	0.293
l	0.707	0

Context 2: Order D_2, D_3 (rows D_2, columns D_3)

	m	l
m	0	0.293
l	0.707	0

Context 2: Order D_3, D_2 (rows D_3, columns D_2)

	m	l
m	0	0.625
l	0.375	0

Context 3: Order D_3, D_1 (rows D_3, columns D_1)

	m	l
m	0	0.462
l	0.538	0

Context 3: Order D_1, D_3 (rows D_1, columns D_3)

	m	l
m	0	0.525
l	0.475	0

Table 4.6: Relevance probabilities for different orders of documents in the three contexts of Query 1

should be strictly less than 2. From the first query results in Table 4.8 I calculate $\Delta = |E[D_1^2] - E[D_1^3]| + |E[D_2^1] - E[D_2^3]| + |E[D_3^2] - E[D_3^1]| = 0.9629$. Similarly, the Δ values for all three queries are listed in Table 4.7.

Query	Δ
Query 1	0.9629
Query 2	1.400
Query 3	1.7999

Table 4.7: Δ values for all the three queries

4.4.7 Discussion

This experiment answers both the sub-research questions for this chapter (**Sub-RQ1** and **Sub-RQ2**). Taking inspiration from successful psychological experiments, a document judgement scenario is designed to exploit the hypothesised quantumness arising due to ambiguity and comparative judgement (which forms a context). The results show that the direct influences due to marginal selectivity or signalling for all the three queries are low enough to allow for the violation of the CbD inequality.

	$D_2^1 = 1$	$D_2^1 = -1$
$D_1^2 = 1$	0	0.7407
$D_1^2 = -1$	0.2593	0

Context 1 (n = 81)

	$D_3^2 = 1$	$D_3^2 = -1$
$D_2^3 = 1$	0	0.3333
$D_2^3 = -1$	0.6667	0

Context 2 (n = 81)

	$D_1^3 = 1$	$D_1^3 = -1$
$D_3^1 = 1$	0	0.4683
$D_3^1 = -1$	0.5316	0

Context 3 (n = 79)

Table 4.8: Relevance probabilities for the three cyclic contexts of Query 1

	$D_2^1 = 1$	$D_2^1 = -1$
$D_1^2 = 1$	0	0.8500
$D_1^2 = -1$	0.1500	0

Context 1 (n = 80)

	$D_3^2 = 1$	$D_3^2 = -1$
$D_2^3 = 1$	0	0.2375
$D_2^3 = -1$	0.7625	0

Context 2 (n = 80)

	$D_1^3 = 1$	$D_1^3 = -1$
$D_3^1 = 1$	0	0.5309
$D_3^1 = -1$	0.4691	0

Context 3 (n = 81)

Table 4.9: Relevance probabilities for the three cyclic contexts of Query 2

	$D_2^1 = 1$	$D_2^1 = -1$
$D_1^2 = 1$	0	0.8148
$D_1^2 = -1$	0.1852	0

Context 1 (n = 81)

	$D_3^2 = 1$	$D_3^2 = -1$
$D_2^3 = 1$	0	0.7125
$D_2^3 = -1$	0.2875	0

Context 2 (n = 80)

	$D_1^3 = 1$	$D_1^3 = -1$
$D_3^1 = 1$	0	0.0500
$D_3^1 = -1$	0.9500	0

Context 3 (n = 80)

Table 4.10: Relevance probabilities for the three cyclic contexts of Query 3

It is important to emphasise the difference between contextuality mentioned in this experiment and the usage of context and context-sensitivity in other fields, especially in user modelling and personalisation. Coming back to the example of the query 'Things to do near me', it is very straightforward to say that the weather serves as a contextual influence on the relevance of a document. This type of explicit context is part of the signalling noise and is removed in the CbD theory. Quantum contextuality is about implicit influences which remain even when all such hidden/latent variables have been taken into account (as signalling). I label these types of contextual influences as implicit contextuality. It is not possible, as in QT, to give an intuitive account as to what an implicit contextual influence is. For QT, the lack of realism which contextuality proves is considered as a fundamental property of nature at the microscopic level. The same can be said of human decision-making under ambiguity. When the judgements are constructed on-line by the decision-maker, rather than pre-defined from memory, quantum-like contextuality can manifest. It can be detected in probabilistic formulations of judgement of a large number of users through specifically designed experiments.

Presence of contextuality supports my hypothesis regarding the in-deterministic and thus implicitly contextual nature of relevance itself. One cannot pre-assign values of relevance to a document based on arbitrary parameters. In other words, un-judged documents do not have any relevance up until the point at which relevance is judged by an individual. The uncertainty provided by the broad nature of the queries in the present study (comparable to real-world queries) opens the way for decisions to be based not only on the defining information from the documents themselves, but from a variety of other factors including parallel judgements made about other documents in the results. In reality, rather than deciding "is this document relevant to me", one may ask themselves, "is this document the most relevant", or "is this document relevant enough to be worth reading before some of the other documents in the results". By asking comparative questions such as this, one is saving cognitive resources by considering multiple questions simultaneously. The same types of comparative questions were asked in this experiment. Comparative judgements about documents that have a less than certain answer regarding their relevance to a query are common in everyday searching experience, and my finding that relevance judgements made under these conditions show may implicit contextuality apply to these real-world cases.

4.4.7.1 Remark: Arguments against Contextuality-by-Default Theory

In recent months, the CbD theory has been challenged in that it fails to sufficiently eliminate all classical influences in the experiments and that the inequality violation may be due to these classical influences and not any quantum influence. [9] argues that CbD falls within the classical (Kolmogorovian) probability framework by allowing for different sample spaces for the different contexts. In [184], the authors point to the

fact that the direct influences in CbD theory are defined as a difference in expected values of a variable measured in two different contexts, i.e., $E(R^a) - E(R^b)$ (see Equation 4.21 above). However, it is possible to show that direct influences can exist but still the average value of the above difference comes out to be zero. [12] argue that signalling can be of two types - signalling with communication and signalling without communication. Here, communication means any type of information transfer which can influence a quantum/cognitive state. The definition of signalling used in CbD theory is restricted to only signalling with communication. That is, it is assumed in CbD theory that there are no direct influences when there is no communication of information. Hence it eliminates direct influences by eliminating noise from communication. However, [12] provide a dummy scenario where signalling can take place with communication, in pre-defined correlations which violate randomness.
The above works argue that it may not be possible for implicit contextuality to exist in human decision-making at all, as it is very difficult to account for, let alone eliminate signalling. It remains to be seen how the authors of CbD respond to these criticisms.

While the debate around CbD is left for the cognitive scientists to resolve, I feel that the result of violation of CbD inequality is worthwhile because CbD goes a long way in removing the direct signalling influences from the judgements. As we can see, order effects are also eliminated in the experimental design and still CbD is violated. This means that the underlying contextuality, even though not purely quantum-like, cannot be modelled using classical methods.

4.5 Conclusion

In this chapter I have hypothesised against the deterministic nature of relevance of documents. Relevance cannot be pre-determined because it is constructive in nature. Various contexts associated with user's interaction with the document influence the user's judgement of relevance. Especially those documents whose relevance to a query is not straightforward and is subject to ambiguity, multiple interpretations or previous beliefs/biases of a user. Such absence of determinism/realism is analogous to the nature of quantum mechanical systems, therefore I borrow the tools from quantum theory to empirically demonstrate in-determinism of relevance.

Two experiments carried out to test the hypothesis are reported in this chapter. They are the first attempts at using Bell-type inequalities to model IR data, although researchers have used them in the past in other types of decision-making experiments.

The first experiment fails to show any violation of Bell-type inequalities although it provides a method to formulate multidimensional relevance judgement in terms of CHSH inequality (**Sub-RQ1**). Also, I learn a few important things in the course of carrying it out. It signals a major shift in my experimental methodology from standard IR datasets and query logs to user studies and crowdsourced data collection.

The second experiment is based on recent success in detection of implicit contextuality in human decision-making. It is a crowdsourced user study asking users to judge documents for real-world queries under multiple contexts. Different documents present alongside a given document provide different judgement contexts. Results confirm existence of implicit contextuality, although I discuss some of the recent criticisms of the CbD approach.

The tools from QT used in this chapter only allow one to test for quantumness of the data. They cannot be used to model or exploit the quantumness. This is particularly necessary for IR as it is an application oriented discipline. Any theoretical discovery should be aimed at helping improve IR systems or models of users and data. In the next chapters, I intend to continue gathering evidence for quantumness of IR data using such tools from QT which can also be used to build or augment IR models.

CHAPTER 5

Complex Hilbert Space Modelling with Crowdsourced Data

In this chapter, I continue my pursuit of finding evidence for the need for a quantum probability framework over the classical one for modelling user interactions and relevance judgements in IR. However, instead of continuing with different Bell-type inequalities and related experimental designs, a new experiment design protocol is followed here. Bell-type inequalities reveal quantumness in based on the the users' external interaction behaviours, and now I go one step further to investigate the complex representations of user's internal cognitive states behind such interaction behaviours. The methodology of custom designed user studies and crowdsourced data collection is the same as in the second half of the previous chapter. This new protocol not only allows one to test for quantumness in the data, but can also provide quantum based predictive modelling of data.

The Stern-Gerlach (S-G) experiment, which is described in the first section is based on the spin properties of electrons (or the polarisation of photons) which are dependent on the direction of measurement (which sets up a context). As discussed in the previous chapters, the spin property is analogous to multidimensional relevance and hence the focus of this chapter is back to multidimensional relevance. The striking aspect of the S-G experiment is the effect of the measurement of one property of a quantum system on another property. Hence my focus is on the effect of a relevance dimension on other dimensions.

In the Stratified model of relevance [138], the different dimensions (referred in that paper as manifestations) of relevance are considered as different layers interacting with each other. Each of these interacting layers include considerations or inferences of relevance. Hence my aim in this chapter is to formulate this interaction between the different relevance dimensions analogous to the interaction between different spin measurements of quantum systems. In particular I study the effect of consideration of one relevance dimension on another.

Two experiments are reported in this chapter. Both the experiments are designed based on the S-G experiment, but use different protocols to test for quantumness. It is important to note that quantum systems do not always reveal quantum-like data. There might be some measurements of a quantum system which can be modelled using classical theories. For example, sequential measurement of two compatible properties will always follow classical probability laws, and so will be the pattern obtained in the double slit experiment if there are detectors at each slit. Deviation from classical physics was only observed in specific experiments like the double slit experiment without detectors as discussed earlier, the S-G experiment, EPR/Bell-type violation, etc. Therefore, it is possible that human decision-making may reveal quantum-like data in only certain situations.

The main hypothesis of the quantum cognition research programme is that human cognitive states bear similarity with quantum systems and these quantum-like effects can be seen if the cognitive states are probed in a particular way. The claim is not that human cognitive system functions exactly as a quantum system. Rather, some aspects of it are similar to certain aspects of quantum systems, and these cannot be modelled using classical probability based models. Human cognition is unarguably much more complicated than both quantum and classical systems and it would require probabilistic models even more generalised than quantum models to fully describe it.

Extending from my hypothesis in the previous chapter, in this chapter I focus on relevance inference which takes place for each relevance dimension considered by the user. The final decision of relevance is a fusion of all the individual inferences. Thus, my hypothesis for this chapter is that *relevance inference at each dimension does not happen independently*, like a pre-defined value being read out of the internal cognitive state. It is rather constructed at the point of information interaction and thus influenced by the other dimensions considered by the user previously, which serve as a context for the inference of relevance for the current dimension. It is straightforward to see that consideration of an individual relevance criterion can affect the final judgement of relevance. For example, inferring that readability of a document is low may lead to a lower probability of judging it as relevant. Nevertheless, does consideration of readability as low also provide a context that affects the subsequent judgement about credibility of the document?

The research question answered in this chapter is:

- **RQ 3:** How to adapt existing experiments from quantum theory to study dynamic interactions between relevance dimensions so as to reveal quantum-like nature of user cognitive states?

This chapter is divided into two sections, each of which report two different experiments based on the S-G experiment protocol. The details of the S-G experiment are reported in the first section. The implications of my findings to IR are also discussed within the sections.

5.1 Experiment 1 - Stern-Gerlach Experiment for Multidimensional Relevance Judgements

5.1.1 S-G Experiment Setup

The Stern-Gerlach experiment (S-G) was one of the first experiments to show the necessity of a radical departure of modelling microscopic data from existing formalisms [135].

Consider a quantum system, say an electron. Here I focus on a particular property of an electron called the spin. A spin of an electron has two possible values - up or down (positive or negative). The spin is a magnetic property and it is possible to measure it by subjecting an electron beam towards the two poles of a magnet placed in a particular orientation. Those electrons which deflect towards the North pole of the magnet can be attributed, say, a positive spin and those who are deflected towards the opposite pole are said to have been in the negative spin state.

Now consider the series of experiments as shown in Figure 5.1. In the first setup 5.1(a), the negative spin electrons (S_z^-) coming out from the Z-axis apparatus are blocked and the spin positive electrons (S_z^+) are made to pass along the Z-axis apparatus once again. As expected, the output from the second Z-axis apparatus are all S_z^+ electrons. However, if instead of the second Z-axis apparatus, one puts magnets along the X-axis, it is seen that half of the S_z^+ electrons deflect to the negative pole of the magnet (S_x^-) and half deflect towards the positive pole of the magnet kept along the X-axis (S_x^+). Thus, the positive or negative spin of an electron is not independent of the choice of measurement axis. Some of the electrons deflected towards the positive pole when measured along Z-axis are also getting deflected along the negative pole when measured along the X-axis. Things get weirder in setup 5.1(c). A third Z- apparatus placed in the line of the S_x^+ electron beams shows presence of two beams - for the S_z^+ and S_z^- spin states. This is despite that fact that S_z^- was blocked after the first apparatus. It can be said that the measurement of S_x^+ component by the apparatus along the X-axis influences (in this case, completely destroys) any previous information about S_z^+ and S_z^- (i.e. the fact that we had all electrons in positive spin with respect to Z-axis and no S_z^- components).

In order to understand these results more clearly, a model of these electron spins is constructed using the bra-ket notation [135] to represent vectors. Any complex valued vector A is represented as a ket - $|A\rangle$ and the complex conjugate of A is a bra vector - $\langle A|$. The inner product of two vectors A and B is calculated by taking the product of the bra of one vector and the ket of another - $\langle B|A\rangle$. The norm of a vector is written as $|\langle A|A\rangle|^{1/2}$. As we saw, the S_z^+ electrons split equally into two directions when subject to magnets along X-axis, the S_z^+ state of an electron is represented as a linear combination of the states S_x^+ and S_x^-:

$$|S_z^+\rangle = \frac{1}{\sqrt{2}}|S_x^+\rangle + \frac{1}{\sqrt{2}}|S_x^-\rangle \tag{5.1}$$

where the coefficients of state vectors $|S_x^+\rangle$ and $|S_x^-\rangle$ are called probability amplitudes and the square of these coefficients give the probability of finding an electron in a particular state. Here, an electron in $|S_z^+\rangle$ state is said to be in both $|S_x^+\rangle$ and $|S_x^-\rangle$ states at the same time, a concept called Superposition. Similar experiment with S_z^- leads us to:

$$|S_z^-\rangle = \frac{1}{\sqrt{2}}|S_x^+\rangle - \frac{1}{\sqrt{2}}|S_x^-\rangle \tag{5.2}$$

It is worth noting that the vectors $|S_i^+\rangle$ and $|S_i^-\rangle$ are orthogonal and therefore we have $\langle S_i^+|S_i^-\rangle = 0$. Using this property, one can also express the X-axis spin states in terms of Z-axis spins:

$$\begin{aligned} |S_x^+\rangle &= \frac{1}{\sqrt{2}}|S_z^+\rangle + \frac{1}{\sqrt{2}}|S_z^-\rangle \\ |S_x^-\rangle &= \frac{1}{\sqrt{2}}|S_z^+\rangle - \frac{1}{\sqrt{2}}|S_z^-\rangle \end{aligned} \tag{5.3}$$

This explains the observation that the S_x^+ component from the second apparatus had both the S_z^+ and S_z^- components. On careful examination of Equations 5.1,5.2 and 5.3, one can see that a definite state along Z-axis, say $|S_z^+\rangle$, is an indefinite state along the X-axis (as it is an equal superposition of both positive and negative spins). One cannot jointly determine both the Z and X component of the spin of the electron. These two properties are thus *incompatible* with each other.

To round up my explanation of the fundamentals of Quantum Theory through the S-G experiment, consider that instead of measuring along X-axis, the magnets are positioned along the Y-axis. A similar, symmetrical behaviour of electron spins is seen such that one can consider the spin along the Y-axis to be in a superposition or linear combination of positive and negative spins along the Z-axis. So, one can represent the electron spin states along the Y-axis in terms of the states along Z-axis as: $|S_y^+\rangle = \frac{1}{\sqrt{2}}|S_z^+\rangle + \frac{1}{\sqrt{2}}|S_z^-\rangle$ and similarly for spin negative state along Y-axis. However, this makes $|S_y^+\rangle = |S_x^+\rangle$, but we know that the spin state of the electron along Y-axis exists separately as they get deflected onto the magnetic poles when aligned along the Y-axis. In order to resolve this issue, Quantum Theory turns towards complex numbers. One can represent the probability amplitudes of the states as complex numbers. Thus for the spin state along the Y-axis:

$$\begin{aligned} |S_y^+\rangle &= \frac{1}{\sqrt{2}}|S_z^+\rangle + \frac{i}{\sqrt{2}}|S_z^-\rangle \\ |S_y^-\rangle &= \frac{1}{\sqrt{2}}|S_z^+\rangle - \frac{i}{\sqrt{2}}|S_z^-\rangle \end{aligned} \tag{5.4}$$

Thus a two-dimensional vector space needed to describe the two-valued spin states of an electron along three different axis must be a complex vector space.

5.1.2 S-G Experiment in the IR Context

The cognitive analogue to the S-G experiment was originally discussed in [63]. In order to draw an analogy of the electron spin states in terms of human judgements, I consider the two-valued spin measurements to be equivalent to the yes/no answers. The measurement along the different axes are equivalent to making judgements along different perspectives or dimensions. So, for relevance judgements, one can consider positive spin and negative spin outcomes to be decisions of relevance and non-relevance respectively. The different axes are the different dimensions of considering relevance. Just like spin of an electron is not an independent quantity and depends on the axis of measurement, similarly relevance of a document cannot be assumed to exist independently of choice of dimension considered. Of course, relevance is also a separate measure, a fusion of all dimensional

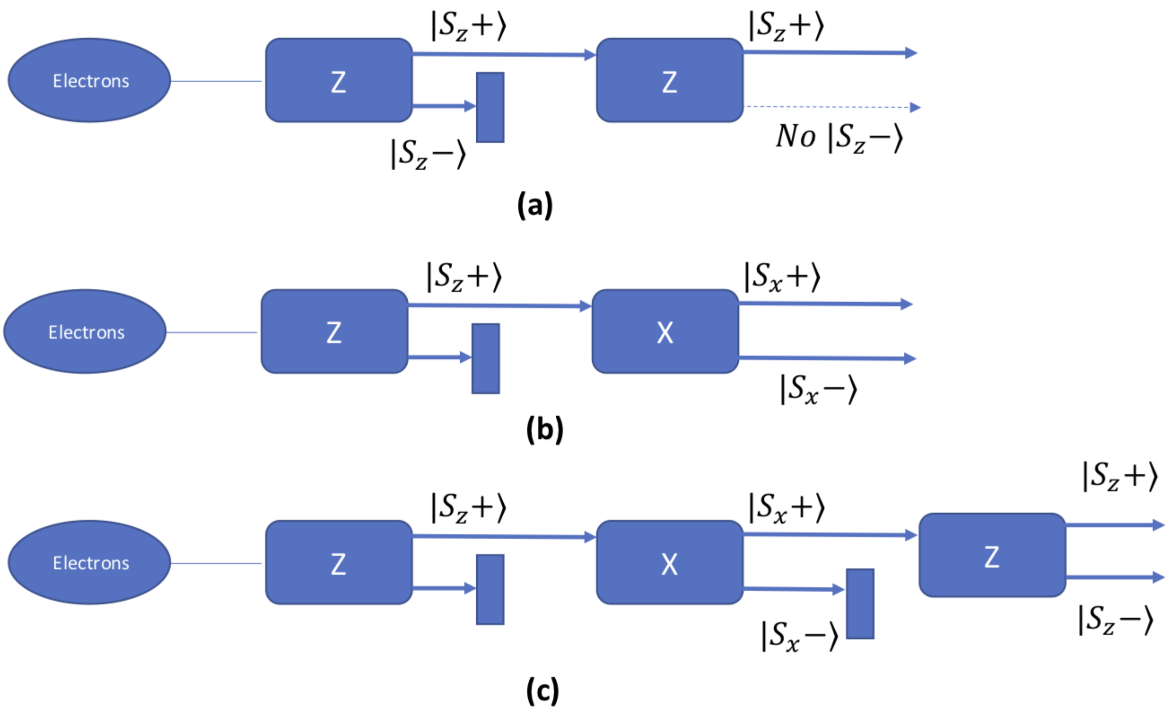


Figure 5.1: Stern-Gerlach Experiment

relevances. There is no corresponding analogy for electron spin. But in this chapter, my focus is on the different dimensions of relevance.

A document may appear relevant when considering the topicality dimension but the users may be uncertain about its Reliability or Understandability. In my experiment, I consider three dimensions. Topicality - whether the information contained in the document is related to the topic of the query, Reliability - whether the user would rely on the information obtained from the document, and Understandability - how easy is it to understand the information presented in the document. The cognitive state of a user before judging a document is represented as:

$$|S\rangle = t\,|T+\rangle + \sqrt{1-t^2}\,|T-\rangle \tag{5.5}$$

where $|T+\rangle$ represents the cognitive state of a user judging the document as topically relevant (with probability t^2) and $|T-\rangle$ represents the state of a user judging the document as topically irrelevant (with probability $1-t^2$). I could have represented the state $|S\rangle$ in terms of Reliability or Understandability states, but I choose the Topicality basis as the standard basis of representation. Similarly, one can represent the Understandability basis in terms of the Topicality basis as:

$$\begin{aligned} |U+\rangle &= u\,|T+\rangle + \sqrt{1-u^2}\,|T-\rangle \\ |U-\rangle &= \sqrt{1-u^2}\,|T+\rangle - u\,|T-\rangle \end{aligned} \tag{5.6}$$

Note that $|U-\rangle$ is constructed using the orthogonality constraint of $|U+\rangle$ and $|U-\rangle$. Here u^2 is the probability that users judge a document Understandable, given that they also consider it as Topical.

Further, in order to express Reliability dimension in terms of its interaction with the Topicality perspective, we write:

$$\begin{aligned} |R+\rangle &= r\,|T+\rangle + \sqrt{1-r^2}\,e^{i\theta_r}\,|T-\rangle \\ |R-\rangle &= \sqrt{1-r^2}\,e^{-i\theta_r}\,|T+\rangle - r\,|T-\rangle \end{aligned} \tag{5.7}$$

where, recall from Section 2, the need of using a complex probability amplitude for the third measurement basis. Thus, one needs four parameters in order to construct the Hilbert space - t, u, r and θ_r. I intend to find these parameters by asking three sequential questions of each user analogous to performing measurements along different axes of the spin for a beam of electrons.

For an initial state of the system $|S\rangle$, the probability of event $|A\rangle$ in the quantum framework is given by $P(A) = |\langle A|S\rangle|^2$ i.e., square of projection of vector $|S\rangle$ onto vector $|A\rangle$. Note that the notation $\langle A|B\rangle$ is the inner product of two vectors. The probability for event A followed by B is given as [32]:

$$P(B,A) = |\langle B|A\rangle|^2 |\langle A|S\rangle|^2 \tag{5.8}$$

which is read from right to left as projecting the initial state $|S\rangle$ to the vector for event A and then projecting this state (to which the initial state has collapsed) onto the state vector for event B. The quantum framework does not define joint probability of events A and B as, in general, $P(A,B) \neq P(B,A)$. As we can see $P(A,B) = |\langle A|B\rangle|^2 |\langle B|S\rangle|^2$, which for $\langle A|S\rangle \neq \langle B|S\rangle$ is not equal to $P(B,A)$ in Equation 5.22. Note that I use the notation $P(B,A)$ to refer to the probability of the sequence $A->B$, i.e. B takes place after event A.

5.1.3 Experimental Design

My experiment thus consists of the following steps:

1. I first prepare a user's cognitive state into one of $|T+\rangle$ or $|T-\rangle$ states by asking a user whether a document is topically related to the query or not. This consists of projecting the user's cognitive state $|S\rangle$ onto the vectors $|T+\rangle$ and $|T-\rangle$. Thus the probability of obtaining a positive response on asking the question about topicality from a user is given as $P(T+) = |\langle T+|S\rangle|^2 = t^2$. I can thus obtain the value for t.

2. a) Next, I take the users who answered yes to the topicality question and ask them about the understandability of the document. This way one can obtain the probability $P(U+, T+)$. It is represented in the Hilbert space as

   $$P(U+,T+) = |\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2 = u^2 * t^2 \tag{5.9}$$

   (Note from Equation 5.6 that $\langle U+|T+\rangle = u$). I thus obtain value of u.

   b) Instead of asking the question about understandability, if we take some of the users who respond positively to topicality and ask them about reliability of the document, one can calculate the probability $P(R+,T+) = |\langle R+|T+\rangle|^2 |\langle T+|S\rangle|^2 = r^2 * t^2$ and thus obtain the value of r.

3. Now we are left with figuring out the value of θ_r. This is done in the following way - those users who answer positively to topicality and understandability questions are asked the third question about reliability. Thus

   $$P(R+,U+,T+) = |\langle R+|U+\rangle|^2 |\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2 \tag{5.10}$$

   Note that $\langle R+|U+\rangle$ is a complex quantity and its square is calculated by multiplying it by its complex conjugate. Thus $|\langle R+|U+\rangle|^2 = \langle U+|R+\rangle \langle R+|U+\rangle$. Hence we have,

   $$\begin{aligned} \langle U+|R+\rangle &= (u\langle T+| + \sqrt{1-u^2}\langle T-|) \times (r|T+\rangle + \sqrt{1-r^2}e^{i\theta_r}|T-\rangle) \\ &= ur + \sqrt{(1-u^2)(1-r^2)}e^{i\theta_r} \\ \langle R+|U+\rangle &= |\langle U+|R+\rangle|^+ \\ &= ur + \sqrt{(1-u^2)(1-r^2)}e^{-i\theta_r} \end{aligned} \tag{5.11}$$

   Finally,

   $$\langle U+|R+\rangle\langle R+|U+\rangle = (ur)^2 + (1-u^2)(1-r^2) + 2ur\sqrt{(1-u^2)(1-r^2)}\cos\theta_r \tag{5.12}$$

   Now, we know u and r from previous steps, the probability $P(R+,U+,T+)$ obtained from the experimental data helps us to calculate the value of θ_r

5.1.4 Participants

I recruited 300 participants for the user study using the online crowd-sourcing platform Prolific (prolific.ac). The only pre-screening criterion for participating in the study was a cut-off of 96 percent approval rate. Approval rate for a participant in Prolific is the fraction of submitted responses approved. The participants were paid at a rate of £7.08 per hour. Data of 5 participants was excluded as they completed the study in much less time than the minimum duration assumed for proper responses. The questionnaire was designed using the Qualtrics platform (qualtrics.com/uk). Proper consent was sought and they were also informed that data protections laws are being complied with. The study was approved by The Open University UK's OU Human Research Ethics Committee with reference number HREC/3063/Uprety. Please refer to Appendix B for further details regarding the ethical approval process and the user study.

5.1.5 Material

Query	Information Need	Source
Radio Waves and Brain Cancer	Look for evidence that radio waves from radio towers or mobile phones affect brain cancer occurrence	TREC 2005 Robust track (310)
symptoms of mad cow disease in humans	Find information about mad cow disease symptoms in humans	TREC 2013 Web Track (236)
educational advantages of social networking sites	What are the educational benefits of social networking sites?	TREC 2014 Web Track (293)

Table 5.1: Selected queries and their descriptions

The participants were shown three queries and one document snippet for each query as it appears in popular search engines like Google and Bing. The queries and description of the information need (IN) were shown as consistent with the TREC style, as listed in Table 6.2. The document snippets were constructed manually by altering particular aspects of existing documents obtained in order to introduce both uncertainty in judging with respect to a particular dimension and also incompatibility between the dimensions.

For instance, Figure 5.2 shows the snippet for the first query. The source URL is created in a way as to create some uncertainty about the reliability of the source. In the same way, the title of the document does not explicitly reflect that it is about the topic of the query and it is also not easy for everyone to understand the information in the body of the snippet. An uncertain user might answer negatively to the topicality question (attains the definite $|T-\rangle$ state), but on being asked to consider the understandability dimension, the user might read the snippet body carefully which can influence the user to become uncertain about topicality again. Similar criteria was followed in designing the document snippets for the other two queries, shown in figures 5.3 and 5.4. The three queries chosen thus allowed us to design document snippets to exhibit these characteristics.

Note that while a document snippet might not represent the whole document, my aim here is not to correct accurate relevance labels. Rather, the manually designed document snippets help create and simulate ambiguity in judging different dimensions of relevance.

Here are some possible radiation dangers in your environment....

https://hustlebustlenews.com/here-are-some-radiation..

May 5, 2014 - A study examined the role of occupational RF/ MW-EMF exposure in the risk of meningioma.....the International Commission on Non-Ionizing Radiation Protection. Several conditional logistic regressions performed for glioma and meningioma. No significant association.....However, the slight increase in risk...merits further research....

Figure 5.2: Document for Query 1

What Chronic Wasting Disease and Mad Cow Disease can Teach Us ...

https://rightasrain.uwmedicine.org/.../what-chronic-wasting-disease.../

Apr 16, 2018 - The most well-known prion **disease in humans** is variant Creutzfeldt-Jakob disease. ... bovine spongiform encephalopathy, a.k.a. **mad cow disease**, in the ... If there is lead or mercury in the environment, pets often show **signs** ...

Figure 5.3: Document for Query 2

7 Ways that Social media is affecting us positively - Currati

https://curatti.com/social-media-positive-effects/

Feb 14, 2018 - Recently, we published the article "Is **Social Media** Really an Existential ... approach candidates through **social networking sites** like LinkedIn. ... **Social Media** Has a Lot of **Benefits** for Students and Teachers ... Did you know- 59% of schools say their students use....

Figure 5.4: Document for Query 3

Parameter	Query 1	Query 2	Query 3
$P(T+)$	0.7622	0.6736	0.8993
$P(U+,T+)$	0.4405	0.5416	0.8724
$P(R+,T+)$	0.4609	0.4857	0.5616
$P(R+,U+,T+)$	0.2587	0.4513	0.6442
$P(R+,U-,T+)$	0.1188	0.0694	0.0000
$P(U+,R+,T+)$	0.2765	0.4285	0.5410
$P(U+,R-,T+)$	0.1560	0.0857	0.2739
t^2	0.7622	0.6736	0.8993
u^2	0.5779	0.8041	0.9701
r^2	0.5462	0.7311	0.6456
θ_r	80.62 deg	56.79 deg	51.43 deg

Table 5.2: Parameter values and associated probabilities

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

Figure 5.5: Probabilities for the questions of TUR and TRU for the three queries

5.1.6 Procedure

The participants were shown the query and the document and after that asked the following questions:

1. Is the document about the topic of the search query? (T)
2. Is it easy to understand the information presented in the document snippet? (U)
3. Would you rely on the information presented in this document? (R)

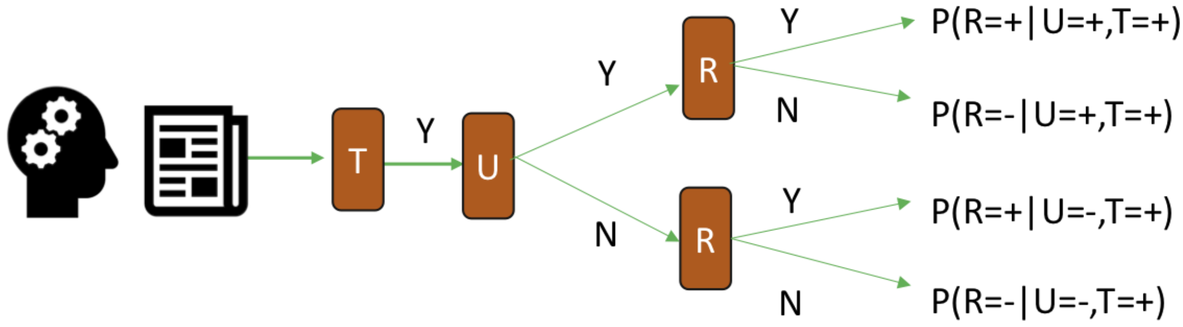


Figure 5.6: Asking three questions in TUR order

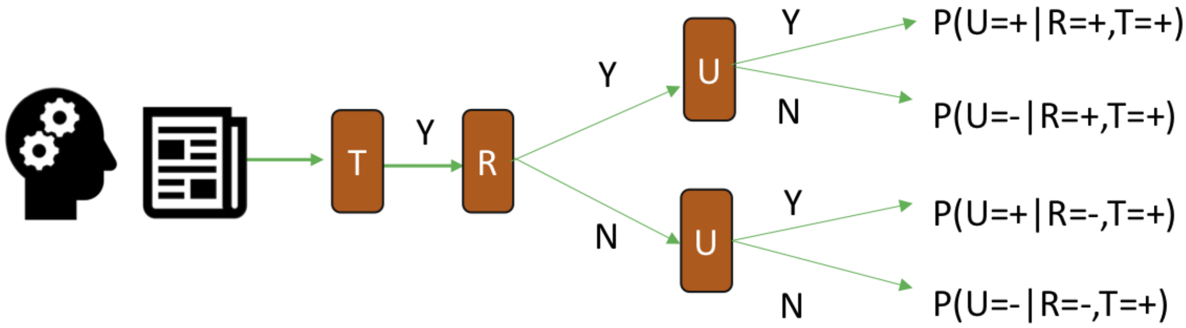


Figure 5.7: Asking three questions in TRU order

Note that a between subjects design was carried out and the participants were uniformly split into two groups - one group was asked questions in the TUR sequence (figure 5.6), which is used to calculate parameters u and θ_r, by calculating probabilities $P(U+,T+)$ and $P(R+,U+,T+)$. The other group was asked questions in the TRU sequence (figure 5.7), in order to calculate the parameter r, by calculating $P(R+,T+)$. The participants were shown the next question only after answering the current question, so that they their answers are not primed by seeing all the three questions.

5.1.7 Results and Discussion

All the probabilities obtained are listed in Figure5.5. Those probabilities required to calculate the parameter values for constructing the Hilbert space are shown in Table 5.2, along with the calculated parameter values. As we see, we get the complete two-dimensional Hilbert space involving the 3 real parameters and the complex phase θ_r. 3 questions are necessary because it implies measurement along 3 different basis which gives rise to the need for complex number representation. The existence of a superposition state signifies that initially a user does not exist in a definite state of judgement with respect to an information object. When a particular question is asked or is considered in the user's mind, e.g. about reliability, this uncertainty resolves and the user's cognitive state collapses to one of two values of relevance or non-relevance.

5.1.8 Wigner Function

I constructed a complex-valued Hilbert space to model the user cognitive state for decisions from incompatible perspectives. In doing so, I utilised the mathematical framework of quantum theory. Another method to verify the "quantumness" of the model is using a discrete Wigner function [70, 181]. It is a criterion used in quantum theory which distinguishes between quantum and classical statistics in the data. Wigner functions are quasi-probability distributions which map quantum states to a phase space. Quasi-probability distributions relax some of the axioms of the Kolmogorov probability theory, the highlight being the existence of negative probabilities in the distribution for states which do not have a classical model. The discrete Wigner function distribution is given as:

$$W = \begin{pmatrix} 1+r_x+r_z & 1-r_x+r_z \\ 1-r_x-r_z & 1+r_x-r_z \end{pmatrix} \tag{5.13}$$

where $r_x = 2 * \sqrt{t^2(1-t^2)}$ and $r_z = 2 * t^2 - 1$. The full derivation can be found in [63]. The Wigner function for the three query-document pairs is obtained as:

$$W_1 = \begin{pmatrix} 0.5939 & 0.1683 \\ -0.0939 & 0.3317 \end{pmatrix} \; W_2 = \begin{pmatrix} 0.5712 & 0.1024 \\ -0.0712 & 0.3976 \end{pmatrix} \tag{5.14}$$
$$W_3 = \begin{pmatrix} 0.6001 & 0.2992 \\ -0.1001 & 0.2008 \end{pmatrix}$$

The negative values in the Wigner function distribution is an indicator of quantum interference which shows that the statistics generated by the Stern-Gerlach type experiment are quantum statistics and thus a quantum model is needed to model such data. As discussed before, the interference effects are due to the incompatibility between decision perspectives. The decision of reliability interferes with that of understandability, for example. Thus, this experiment successfully answers the sub-research question of this chapter (**Sub-RQ**). I successfully adapt the S-G experiment to a multidimensional relevance judgement scenario and verify the quantumness of the data obtained.

5.1.9 Incompatibility

In QT, incompatibility of measurements can be represented in the form of non-commutating operators. Operators are matrices which encapsulate a measurement which can be performed on a quantum state. Measuring a property of a system generates an event. One can construct operators by first constructing their eigenvectors, also called projectors. The projector of an event A is represented by the outer product of the vector corresponding to A with itself, i.e $|A\rangle \langle A|$. In the complex-valued two-dimensional Hilbert space we constructed, there are have 3 bases, corresponding to T,R and U questions/measurements. I had assumed Topicality as the standard basis, and hence the orthogonal vectors $|T+\rangle$ and $|T-\rangle$ are given as:

$$|T+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |T-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.15}$$

Thus the projector for event T is given by:

$$|T+\rangle \langle T+| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad |T-\rangle \langle T-| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.16}$$

The two projectors form the eigen vectors of the operator for the event T with eigen values $+1$ and -1. Thus we have the operator T as :

$$\begin{aligned} \hat{T} &= |T+\rangle \langle T+| - |T-\rangle \langle T-| \\ &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{5.17}$$

Combining Equations 5.32 and 5.6, one can write the vectors for event U as:

$$|U+\rangle = \begin{pmatrix} 0.7601 \\ 0.6496 \end{pmatrix} \quad |U-\rangle = \begin{pmatrix} 0.6496 \\ -0.7601 \end{pmatrix} \tag{5.18}$$

Thus, I get the operators for U and R for query 1 as:

$$\hat{U} = \begin{pmatrix} 0.1558 & 0.9874 \\ 0.9874 & -0.1558 \end{pmatrix} \qquad \hat{R} = \begin{pmatrix} 0.0924 & 0.9955e^{i80.62} \\ 0.9955e^{-i80.62} & -0.0924 \end{pmatrix} \tag{5.19}$$

I find that all the three operators do not commute pairwise - $[\hat{T},\hat{U}] \neq 0, [\hat{T},\hat{R}] \neq 0, [\hat{R},\hat{U}] \neq 0$ for the three queries. The consequence of incompatibility is that it is not possible to form joint distributions over answers involving incompatible questions. Thus $P(T = +, R = +)$ is not defined because it will be different if order of questions are different, i.e., $P(T = +, R = +) \neq P(R = +, T = +)$.

This is where the advantage of using the Quantum framework lies. In classical probability, events always commute and thus order effects cannot be modelled. Order effects are bound to occur when the relevance dimensions are considered in different orders. From a cognitive point of view, a user is unable to be in a certain state of decision using two relevance criteria. Certainty in one relevance criterion does not imply certainty in another incompatible criterion. The method of constructing incompatible operators formally establishes and predicts order effects.

5.1.10 Interference between Dimensions

My hypothesis for the experiments in this chapter is about the effect of consideration of one dimension on the judgement with respect to another dimension. In this particular experiment, I am testing the interaction between Understandability and Reliability. I ask all users the question of Topicality first because Topicality is generally the foremost criterion of judging a document. I suspect that judgement of Reliability will be effected by whether or not Understandability has been considered. By consideration of Understandability, it is meant that the user has made an effort in comprehending the content of the document.

In Table 5.3, I compare the probability of answering 'Yes' to Reliability question after the Topicality question, with the probability obtained had Understandability been answered before. It can be seen that when users are unable to understand the document, they do not find it reliable either. Statistically significant results are reported for queries 1 and 2, shown in bold font in the Table(Chi-square two tailed test of the equality of proportions, $\alpha = 0.05$). On the other hand, although one can see that having comprehended the information better increases the probability of judging it more Reliable, the increase in probability is not statistically significant.

It is also seen that Reliability has a similar effect on Understandability as shown in Table 5.4. Those users who do not find the documents Reliable don't find it Understandable either. Intuitively one feels that Understandability should be independent of the Reliability of the document, but the data shows the dependence. A possible explanation could be that users who do not find the document Reliable do not make much effort to judge the Understandability dimension and hence the high correlation. This also supports my hypothesis for this chapter, that the inference of relevance at each dimension does not take place independent of the other relevance dimensions.

Quantum-like interference is another implication of incompatibility in decision making which is witnessed in decision data as Law of Total Probability (LTP) violation. For the participants who have answered the question about 'Understandability' first and then 'Reliability', I calculate the probability of answering 'Yes' to 'Reliability' using the law of total probability (LTP) as:

$$P_u(R+,T+) = P(R+,U+,T+) + P(R+,U-,T+) \tag{5.20}$$

The probabilities on the two sides of the above equation are calculated from the data and reported in Table 5.5. I find that $P_u(R+) \neq P(R+)$. However, none of the results are statistically significant. One of the reasons for interference effects to not be significant can be that even when users are judging Reliability without being asked about Understandability, some of them do consider it in their mind (also due to learning from judging the first query). This is equivalent to being asked the question about Understandability as a form of self-elicitation which creates a definite belief state with respect to Understandability. Therefore we do not see a statistically significant difference in the probabilities in the two situations. As such, it is difficult to segregate judgements made by only considering Reliability, from those considering Understandability before Reliability. However, when they do consider Understandability before Reliability, it does make a difference in judgement of Reliability (as discussed in the above two paragraphs).

Note that the calculation of $P_u(R+,T+)$ in the quantum framework incorporates the interference term, which is a function of the complex phase θ_r, which is able to model this interference in experimental data.

$$P_u(R+,T+) = P(R+,U+,T+) + P(R+,U-,T+) + Int(\theta_r) \tag{5.21}$$

	Q1	**Q2**	**Q3**
P(R+ | T+)	0.5462	0.7311	0.6456
P(R+ | U+,T+)	0.5872	0.8332	0.7384
P(R+ | U-,T+)	**0.3692**	0.5261	**0.0000**

Table 5.3: Effect of Understandability on Reliability

	Q1	**Q2**	**Q3**
P(U+ | T+)	0.5779	0.8040	0.9701
P(U+ | R+,T+)	0.5999	0.8822	0.9633
P(U+ | R-,T+)	**0.4074**	**0.4801**	0.8887

Table 5.4: Effect of Reliability on Understandability

Query	$P(R+,T+)$	$P_u(R+,T+)$
Query 1	0.3775	0.4609
Query 2	0.5207	0.4857
Query 3	0.6442	0.5616

Table 5.5: Interference as violation of LTP

5.1.11 Quantum Probabilities vs Classical Probabilities

In this subsection I formulate conditional probabilities of relevance judgement along one dimension given another, using classical vs. quantum frameworks. They are then compared with the same conditional probabilities obtained with the experimental data.

For an initial state of the user cognitive 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 [32]:

$$P(U+,T+) = |\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2 \tag{5.22}$$

To reiterate, 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 Equation 5.22. The conditional probabilities are given according to Luder's rule [32, 84] as:

$$\begin{aligned} P_q(U+|T+) &= P(U+,T+)/P(T+|S) \\ &= \frac{|\langle U+|T+\rangle|^2 |\langle T+|S\rangle|^2}{|\langle T+|S\rangle|^2} \\ &= |\langle U+|T+\rangle|^2 = u^2 \end{aligned} \tag{5.23}$$

Note that subscript q is added to distinguish from classical conditional probability. Then $P_q(R+|U+,T+)$ is given as (see Section 5.1.3 for derivation):

$$P_q(R+|U+,T+) = |\langle R+|U+\rangle|^2 = (ur)^2 + (1-u^2)(1-r^2) + 2ur\sqrt{(1-u^2)(1-r^2)}\cos\theta_r \tag{5.24}$$

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:

$$P(U+,R+,T+) = P(R+,U+,T+) \tag{5.25}$$

which can be written in terms of conditional probabilities as:

$$P(T+)P(R+|T+)P(U+|R+,T+) = P(T+)P(U+|T+)P(R+|U+,T+) \tag{5.26}$$

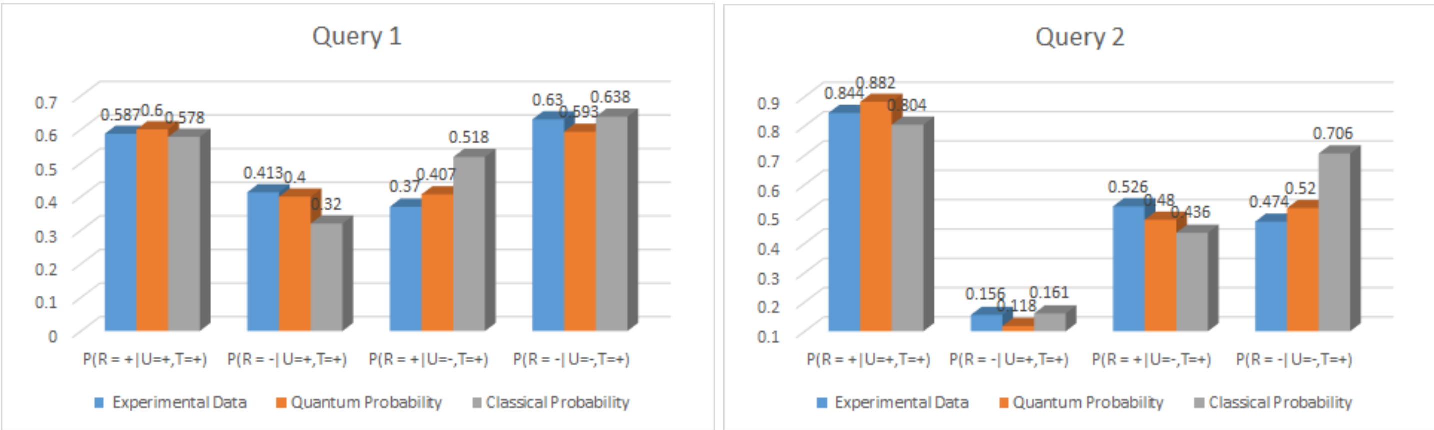
This enables calculation of conditional probabilities using the Bayes rule:

$$P(R+|U+,T+) = \frac{P(U+|R+,T+)P(R+|T+)}{P(U+|T+)} \tag{5.27}$$

Similarly, the other conditional probabilities can be obtained. Again, note that the probabilities in Equations (5.27) and (5.24) are different because of the difference in the underlying assumption of commutativity or joint probability.

Figure 5.8 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 Figure 5.5) 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). 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 the previous sections and the order effects shown in [24] suggest that this is not always the case in general. Therefore one finds Bayesian predictions deviating 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.



(a) (b)

Figure 5.8: How Quantum and Classical Probabilities compare with the experimental data for Query 1 and Query 2

5.2 Experiment - 2

Extending from the Stern-Gerlach protocol, in this experiment I give further evidence for the violation of classical probability theory in multidimensional relevance judgements. Specifically, I investigate the violation of a particular axiom of Kolmogorovian probability theory [86].

5.2.1 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 [86]:

$$0 = \delta = P(A \vee B) - P(A) - P(B) + P(A \wedge B) \tag{5.28}$$

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 (5.28) in quantum mechanics is given by the following definition [169]:

$$\mathscr{D}(U\pm,R\pm) = \Pi(U\pm \vee R\pm) - \Pi(U\pm) - \Pi(R\pm) + \Pi(U\pm \wedge R\pm) \tag{5.29}$$

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

$$\Pi(U\pm) = |U\pm\rangle\langle U\pm|, \quad \Pi(R\pm) = |R\pm\rangle\langle R\pm| \tag{5.30}$$

This quantum correction term $\mathscr{D}(U\pm,R\pm)$ is proportional to the commutator of the projection operators of $U\pm$ and $R\pm$ and can be thus obtained as (detailed proof is beyond the scope of this thesis and can be found in [169]) :

$$\mathscr{D}(U\pm,R\pm) = \left[\Pi(U\pm),\Pi(R\pm)\right](\Pi(U\pm) - \Pi(R\pm))^{-1} \tag{5.31}$$

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 Equation 5.6. 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:

$$|T+\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \quad |T-\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{5.32}$$

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

$$|U+\rangle = \begin{pmatrix}u\\\sqrt{1-u^2}\end{pmatrix}, \quad |U-\rangle = \begin{pmatrix}\sqrt{1-u^2}\\-u\end{pmatrix} \tag{5.33}$$

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 :

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

From the values of u, r and θ_r obtained in Table 5.2, these projection operators can be constructed. The quantum analogue of δ, can then be calculated from Equation (5.31). Value of δ obtained from experiment can be compared to that predicted by the classical (always zero) and quantum probability frameworks.

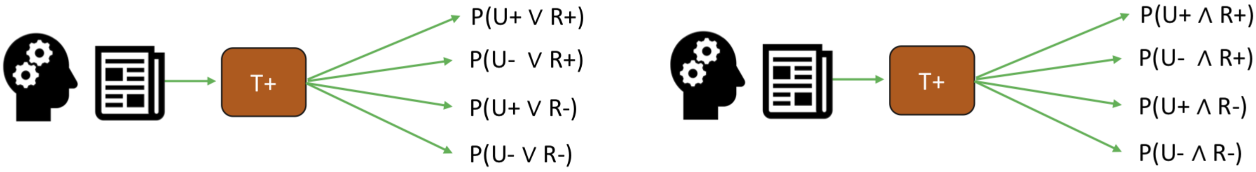
5.2.2 Experiment Methodology

The main aim of this experiment is to investigate the violation of Equation 5.28. I already have the single question probabilities from the first experiment reported in this chapter (section 5.1). I need to obtain the probabilities of conjunction and disjunction. This is done 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 I construct four pairs of statements, as listed in Figure 5.10. For the disjunction measurement, I 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, I ask the participants whether they agree with both of the questions or not. Figures 5.11 and 5.12 show the designs for the disjunction and conjunction questions for a query-document pair. This results in a total of eight such questions and I follow a between-subjects design, such that a participant is shown only one of these eight questions randomly. Note that I am able to use the probabilities from the previous experiment (5.1) because the current 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.

3. Lastly, they are randomly shown one of the eight possible conjunction or disjunction questions and asked to choose the appropriate answer.



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

Figure 5.9: Experiment design

5.2.3 Participants and Material

I 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 per hour. 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. Further details about the ethical review procedure can be found in appendix B.

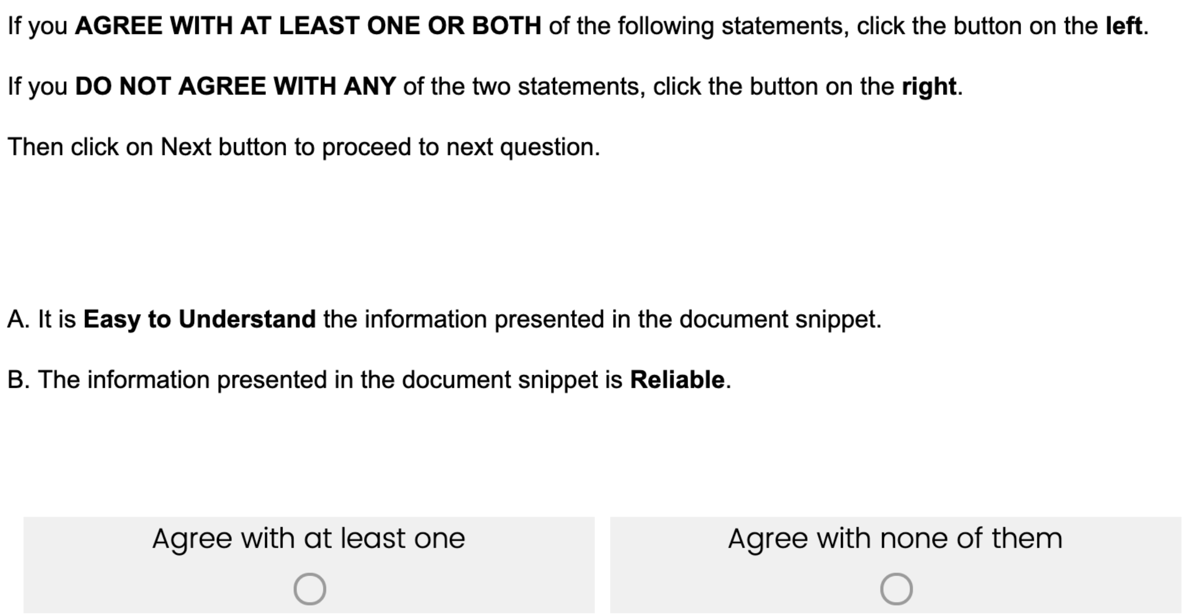
I used the same set of three query-document pairs for our experiment as used in the experiment in Section 5.1, as I have reused some of that data. Each participant was shown the three queries (and the documents) and was asked to judge the topicality of the document and one of the eight questions (so I obtain probabilities like $P(U+ \vee R+|T+)$, etc.) Thus the participants were randomly divided into eight groups for a between-subjects design.

5.2.4 Results and Discussion

The probabilities of conjunction and disjunction of the Understandability and Reliability questions are reported in Figure 5.13. In order to compute the δ reported in Equation

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.

Figure 5.10: Four pairs of statements for conjunction and disjunction questions



If you **AGREE WITH AT LEAST ONE OR BOTH** of the following statements, click the button on the **left**.

If you **DO NOT AGREE WITH ANY** of the two statements, click the button on the **right**.

Then click on Next button to proceed to next question.

A. It is **Easy to Understand** the information presented in the document snippet.

B. The information presented in the document snippet is **Reliable**.

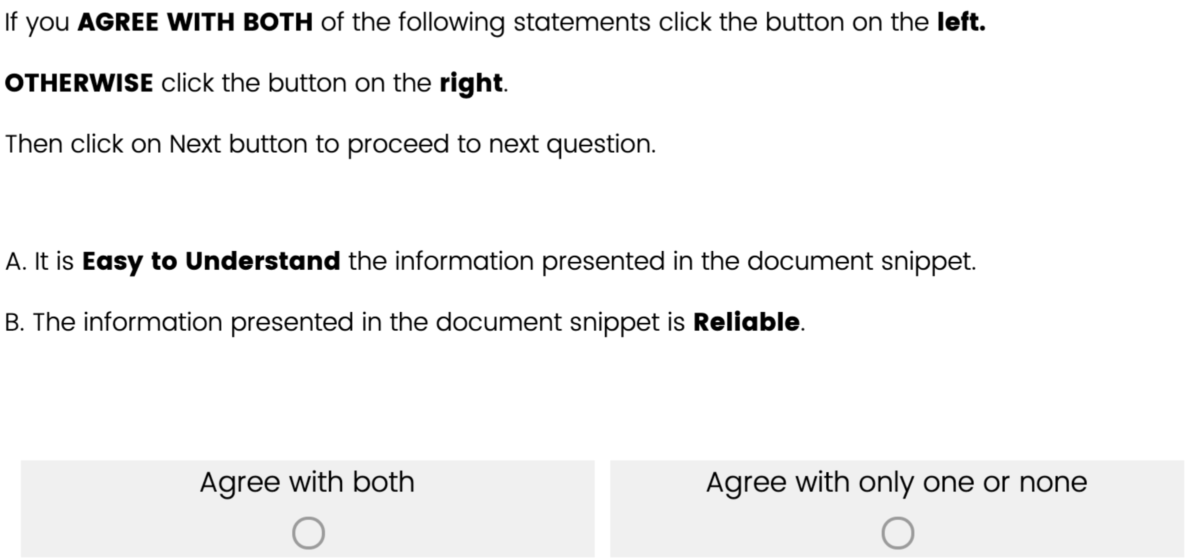
Agree with at least one

Agree with none of them

Figure 5.11: Design for disjunction question

5.28, I 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 Figure5.5). Then, I calculate $\delta = P(U\pm \vee R\pm|T+) + P(U\pm \wedge R\pm|T+) - P(R+|T+) - P(U+|T+)$. In Figure 5.13 one can see that δ is different from zero for all the three queries, although according to classical probability one would expect that δ would be zero in all cases. Equation (5.31), 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 one can see, if operators of U and R commute with each other,


If you **AGREE WITH BOTH** of the following statements click the button on the **left.**

OTHERWISE click the button on the **right**.

Then click on Next button to proceed to next question.

A. It is **Easy to Understand** the information presented in the document snippet.

B. The information presented in the document snippet is **Reliable**.

Agree with both	Agree with only one or none
○	○

Figure 5.12: Design for conjunction question

the quantum correction term in the Equation (5.31) 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 [157]. Quantum models have been previously used to explain such violation [35] where the fundamental notion of incompatibility in judgements is identified as the potential cause.

5.3 Conclusion

In this chapter I have attempted to investigate the interaction between select relevance dimensions in the form of context effects. The experiments are based on the Stern-Gerlach setup in Physics and designed in a way so as to capture these context effects. A formal model using a complex-valued Hilbert space for the user's cognitive state is constructed. My hypothesis that relevance dimensions effect each other is shown by the presence of incompatibility and interference/context effects. These are also the first experiments in quantum-inspired IR where complex numbers are used to represent the user data. Extending the S-G experiment protocol in the second experiment reported in this chapter, the quantumness of the dynamic interactions between relevance dimensions is

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

Figure 5.13: Probabilities for conjunction and disjunction questions and associated violation from Kolmogorovian probability

shown as violation of a certain classical (Kolmogorovian) probability axiom. A particular experimental design is reported which can exploit the quantum cognitive structure. The data shows violation of one of Kolmogorovian probability axioms and reproduces on of the popular findings of Conjunction Fallacy in decision-making. The implications of these findings to IR system design are discussed in chapter 7.

CHAPTER 6

QUANTUM COGNITIVE MODELLING OF RELEVANCE AND USEFULNESS

In the last chapter, I have been successful in providing evidence of quantumness in the form of quantum-like contextual interactions between relevance dimensions. As mentioned before, it is relevance and not the individual relevance dimensions which is the centre of attention in IR. While promising and interesting research is ongoing in IR, especially interactive IR, to understand, define and compute user-centric metrics like satisfaction, utility, task completion, etc., optimising for relevance is still an important part of an IR system.

Therefore, I turn my attention to relevance, in light of the findings of the previous chapters. Particularly, I want to investigate whether the dynamic interactions between the relevance dimensions have an affect on the final document judgements by the users. The sub-research question for this chapter is :

RQ 4: Do the quantum effects observed in the interaction between relevance dimensions have any effect on the final decision of document relevance?

The quantum effects talked about in RQ 4 are based on incompatibility and interference phenomenon of the quantum mathematical framework. A major focus in this chapter is to study the interference phenomenon. Interference has been discussed in the Double Slit experiment in chapter 2. Mathematically, it is defined in terms of the law of total probability (LTP) violation, another fundamental classical probability axiom. Such a violation has also been observed in human decision-making in hypothetical gambling scenarios, termed as the disjunction effect. This is discussed in the next section followed by four experiments. Experiment 1 contains two studies where I investigate disjunction effect in terms of topicality, reliability and relevance. The second experiment investigates another interference effect arising due to the categorisation of reliability before relevance judgement. The third experiment investigates differences in consideration of dimensions when judging relevance versus usefulness of documents. I transition from using document relevance as the user decision to document usefulness for the third experiment, which combines different aspects of many experiments from this chapter as well as the previous chapter.

6.1 Background

6.1.1 Disjunction Effect

The violation of law of total probability or Savage's Sure Thing principle was first studied by Tversky and Shafir [158]. They asked a group of participants (N=98) to play two rounds of a hypothetical gamble where they had an equal chance of winning $200 or losing $100. In one condition, students were informed that they had won the gamble; in another condition they were informed that they had lost the gamble and in the third condition, they were not informed about the result of the first round of gamble. 69 percent of those who had won the first round decided to play another round whereas 59 percent of those who had lost played another round. This number for those who were not told of the results of the first round was 36 percent. Since the two outcomes of winning and losing the first gamble are equally probable, the percentage for the third case should have been between 69 and 59, whereas it is significantly less.

To explain it in detail, consider the events G_2 and $\tilde{G_2}$ as playing the second gamble and not playing the second gamble respectively. W and L represents events winning and losing the first gamble respectively. Then, a player in the "unknown" condition would have to think about the events W,L and then make a choice G_2 or $\tilde{G_2}$. Let $p(W), p(L)$ be the probabilities that a person wins or loses the first gamble respectively where $p(W) = 1 - p(L)$. The total probability of the participants choosing to play the second gamble when they are not informed about the result of their first gamble is given by the law of total probability (LTP):

$$p_T(G_2) = p(W).p(G_2|W) + p(L).p(G_2|L) \\ = p(W,G_2) + p(L,G_2) \tag{6.1}$$

Now, whatever the value of $p(W)$, $p_T(G_2)$ should always lie between $p(G_2|W)$ and $p(G_2|L)$, the two known probabilities. However $p_T(G_2)$ is significantly different from $p(G_2) = 0.36$. This experiment has been conducted several times since, including in scenarios different from gambling, like the Prisoner's dilemma[129], Categorisation-Decision[151], etc. The psychological explanation if the disjunction effect is that majority of those participants who won the first round go for the second round because they feel it as an opportunity to reinvest the money they have won. Those who have lost the first round, 59 percent want to play the second round in order to recover some of their loss. However those who are not told of their result of the first round, are in an ambiguous state and decide to act in a risk-averse manner. Thus a large fraction of them decide against playing another round. But one can see that this behaviour cannot be modelled using the classical probability theory.

QT however, provides a generalisation of LTP and involves the interference term to account for the difference in probabilities [32]. The interference term is generated by considering the participants' cognitive states to be in a superposition of the win and lose outcomes of the first gamble. Figure 6.1 shows the difference between classical and quantum approaches. According to classical logic, a participant arriving at state G_2 can only do so via $G_1 - W - G_2$ or via $G_1 - L - G_2$, even when they don't know explicitly the value of the intermediate state (W or L) and it is modelled stochastically. In other words, the assumption is that participants are always in a distinct state (W or L) during their decision-making process. However, according to quantum logic, the participants who are not aware of the result of the first gamble are in an indefinite state due to ambiguity, $G_1 - G_2$. It is not a mixture or weighted average of the two definite states, but is rather a new state, a superstate, or in QT terminology the superposition state.

6.1.2 Interference of Categorisation on Decision-Making

LTP violation is also witnessed in experiments investigating effect of categorisation before decisions [151]. In these experiments, participants were shown pictures of human faces, which varied along two attributes - face width (narrow or wide) and lip thickness (narrow or wide). One group of participants was asked to categorise the person into 'good' or 'bad' and then make a decision whether to take 'attack' or 'withdraw' action. Another

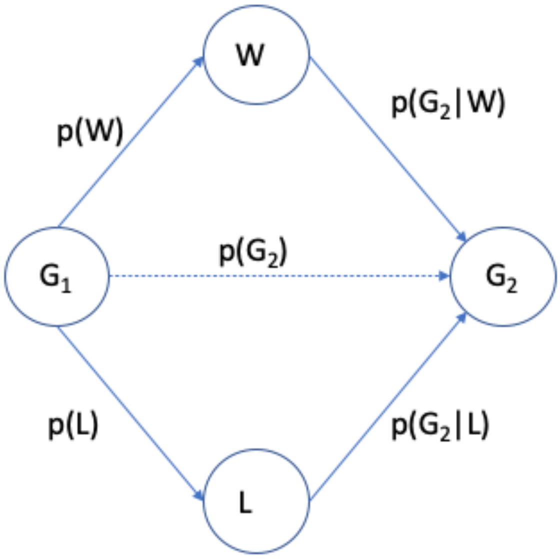


Figure 6.1: Disjunction Effect in gambling scenario

group was only asked to decide upon the action, no categorisation step in between. Again, like the disjunction effect setup, a LTP scenario can constructed here as:

$$p_T(A) = p(G).p(A|G) + p(B).p(A|B) \tag{6.2}$$

where $p(A)$ is the probability of attack and $p(G)$ and $p(B)$ are the probabilities of classifying the face as good or bad respectively. Rationally, the participants who have not undergone the categorisation step but directly decide the action, would have either decided the person to be good or bad before making the decision about their action. Hence, one would expect that the probability of attack for this group would not be significantly different from the total of probability of attack of the other group, calculated according to LTP. The experimental data however shows a statistically significant difference (only for narrow face types) between the probability predicted by LTP and the experimental probability. This difference has also been accounted for using the interference term [176]. Again, the idea is that when a categorisation has not been performed, the cognitive state of participants cannot be modelled as a weighted mixture of the distinct categorical states, but rather as a new superstate/superposition state. QT provides the tools to model such superstates.

One of the aims in the experiments in this chapter is to use similar experimental methodology as described in the experiments above to look for existence of such superstates in IR user behaviour. I first describe the common elements among the experiments - the experiment design materials - queries and documents and the nature of study and participants. Indeed, this is also similar to that used in the experiments in the previous chapter expect for some minor changes.

6.2 Material

The primary design material consists of queries, associated information needs, and document snippets. Like the experiments in the previous chapter, the document snippets were constructed manually by altering particular aspects of snippets obtained from using the queries in a web search engine (`www.google.co.uk`) to introduce both uncertainty in judging with respect to a particular dimension and also incompatibility between the dimensions. The source url, document title and snippet content were manipulated to vary credibility, topicality, and understandability.

All the queries and the associated information needs used in the experiments are listed in Table 6.2. The queries and information needs were also selected such that documents can be designed in a way as to make participants unsure about both their relevance and non-relevance. As such, the queries mostly comprised topics which did not have objective answers. The documents corresponding to these queries are listed in Appendix 8.

In each of these experiments, participants were asked questions related to the dimensions of topicality, reliability and understandability and also for judgements of relevance and usefulness. The questions asked of the participants corresponding to these concepts are listed in Table 6.1.

Concept	Question
Topicality	Is the document about the **topic** of the query or information need?
Understandability	Is it easy to **understand** the information presented in the document?
Reliability	Is the document **reliable**?
Usefulness/Utility	Is the document **useful** for the query or information need?
Relevance	Is this document **relevant** to the query or information need?

Table 6.1: Questions asked of participants about a document

Q No.	Query	Information Need
Q1	Education programs for mammographic image quality assurance	Find out about education programs available to provide training to improve mammographic image quality
Q2	swamp dwelling animals which could face genetic extinction due to interbreeding	To find information about swamp dwelling animals which could face genetic extinction due to interbreeding
Q3	statistics to aid our decision making	To find out how use of statistics can enhance our decision making process
Q4	areas of the human body which are the main targets for lead	Find which parts of the body are affected by the chemical lead
Q5	Companies policy against smoking lower maintenance costs	What evidence is there that companies which adopt a policy against smoking can lower their maintenance costs?
Q6	Radio Waves and Brain Cancer	Look for evidence that radio waves from radio towers or mobile phones affect brain cancer occurrence
Q7	symptoms of mad cow disease in humans	Find information about mad cow disease symptoms in humans
Q8	educational advantages of social networking sites	What are the educational benefits of social networking sites?
Q9	Patent applications for nonlinear neural network oscillators	Find out if there are patent applications filed in the area of nonlinear neural network oscillators.
Q10	impact of technology on democracy	Find about how the increasing spread of technology in our lives impacting democracy and the democratic institutions.
Q11	Positive effects of forest fires	Forest fires are increasing becoming bigger and more frequent in recent times. Are there any positive effects of forest fires?

Table 6.2: All queries and information needs

6.3 Participants

Similar to the previous chapter, participants for the experiments in this chapter are also recruited using the online crowd-sourcing platform Prolific (`prolific.co`). However, this time, I also pre-screen participants with English as the first language criteria (which applies to 71.76% of Prolific participants). This is because I want participants to consider the usual meaning of terms relevance, usefulness, etc. There might be a case where, e.g. the interpretation of the meaning of relevance is different for a non-native English speaker. Hence I only recruit native English speakers to have as uniform a semantic interpretation of the questions as possible. The questionnaire is designed using the Open University licensed Qualtrics platform (`openss.eu.qualtrics.com/`). Proper consent is sought and participants are also informed that data protections laws are being complied with. Details of the ethical review and user studies can be found in Appendix B.

6.4 Experiment 1 - Disjunction Effect in Relevance Judgements

6.4.1 Methodology

In this experiment, I used a similar protocol as in the disjunction effect experiment described in 6.1.1. Instead of playing two gambles, the participants answer two questions about a query and document snippet. There were two studies in this experiment - one where the questions asked are about topicality and relevance and the second in which the questions are asked about reliability and relevance. Thus the two studies or sub-experiments differ only in the first questions asked. The disjunction effect setup is achieved in the following way:

- Step 1: The participants are shown an information need, its associated query and a related document snippet.
- Step 2: They are asked a question about topicality - **Is the document about the topic of the query or the information need?** (for the second study the question asked is **Is the document reliable?**). Three options are presented of which they have to choose only one - Yes, No and Unsure.
- Step 3: All the participants in both the studies are asked the second question - **Is the document relevant to the query or information need?** and can choose only one of two options - Yes or No.

The option termed **Unsure** is to count those responses where users were uncertain about the topicality or reliability judgement. This is similar to the state of those participants of the gambling experiment who were uncertain of the results of their first gamble. The uncertainty in that scenario was enforced upon them by the experimental design of not revealing the results. In my experiment, the uncertainty is not enforced, rather the inherent uncertainty of users is captured. It can be argued that this uncertainty is also due to ignorance of all variables needed to make a certain decision about the topicality/reliability of the document snippet. Whilst some users maybe certain about the first question and answer ‘Yes’ or ‘No’.

The central idea behind both the studies is to compare the relevance judgement of those participants who are certain about the answer to the first question of topicality/reliability (and have answered it as either Yes or No) with those who are uncertain about the first question (selected the ‘Unsure’ option). This is shown in Figure 6.2 for the sub-experiment regarding topicality as the first question and in Figure 6.3 for the other experiment with reliability. The probability of answering ‘Yes’ to relevance when the first answer is ‘Unsure’ is compared with the probability of relevance as predicted by LTP based on the probabilities of users who were certain about the answer of the first question. Let $p(T+)$, $p(T-)$, $p(T_0)$ to be probabilities of answering Yes, No and Unsure respectively for topicality and $p(Rv+|T+)$, $p(Rv+|T-)$ and $p_{T_0}(Rv+)$ be the probabilities of participants judging the document relevant having judged it topical, not topical, or Unsure respectively. Then, the according to LTP, the participants who are unsure about topicality should have the following probability of judging the document relevant:

$$p_{ltp}(Rv+) = p(T+)p(Rv+|T+) + p(T-)p(Rv+|T-) \tag{6.3}$$

A similar argument follows for the other study where the first question was about reliability instead of topicality. Three different queries were shown to 89 participants for both the studies. The queries used, number of participants and the calculations of probabilities as proportions of participants are reported in tables 6.3 and 6.4. Also, the comparison between the probability of relevance for the uncertain users and that calculated by the law of total probability is reported.

6.4.2 Results and Discussion

Analysing the data obtained from the study related to disjunction effect in topicality reveal one important thing straightaway - the proportion of participants who were unsure about topicality of the documents is very less. It is around 10% for two of the queries and around 20% for the third one. This shows that there is very less ambiguity about the topicality, as it is more of an objective property of a query-document pair. Whereas, the proportion of users unsure about the reliability of documents is substantial - around 40%. It is evidently difficult to make a confident judgement of reliability of document than its topicality.

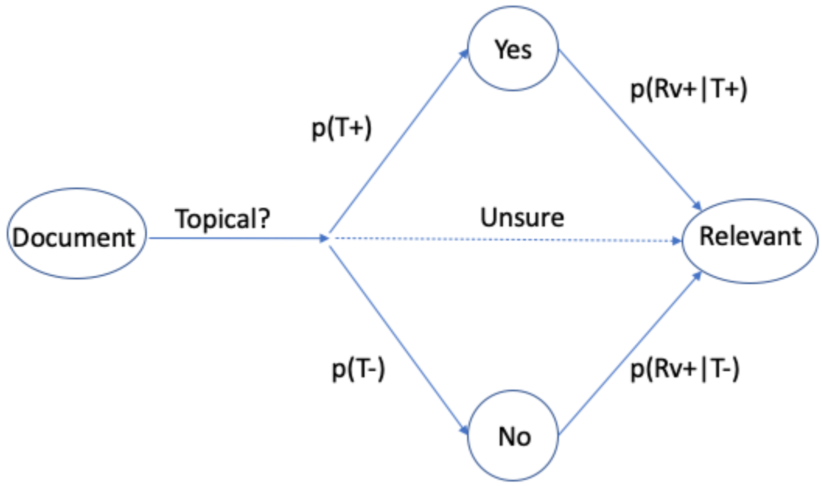


Figure 6.2: Experimental setup for disjunction effect in topicality

One can also see that there is no statistically significant deviation from the predictions of LTP. This is confirmed using a 2-sample test for equality of proportions[1] with $\alpha = 0.05$. All the probabilities of relevance of participants in the ambiguity condition behave according to classical probability. For example, $p_{T_0}(Rv+)$ is always between $p(Rv+|T+)$ and $p(Rv+|T-)$. In Q_6 the value of $p_{Rb_0}(Rv+)$ is larger than $p(Rv+|Rb+)$, but it is not significantly larger.

Thus, I do not find a significant violation of LTP in these two experiments. As described earlier in the gambling experiment, the participants who know about the results of the first gamble have different reasons of playing the second gamble. Therefore, not knowing the result makes them adopt a risk-averse approach and significantly lowers the probability of playing the second gamble. This is what causes the violation of LTP, and creates the need for the interference term in order to model. One can also say that this is what causes interference. In my current experiments, lack of violation of LTP can be simply explained as lack of any interference in the decision-making. While the uncertainty in deciding reliability of the document be quantum in nature (i.e. a superposition state), it does not interfere with the subsequent decision of relevance. Also, in QT, interference is directly related to non-commutativity or incompatibility of variables measured. I have shown in previous chapters how order effects and incompatibility manifest between the relevance dimensions. For the current experiments, one can say that topicality and relevance are compatible for the three query-documents. Similarly, reliability and relevance are also compatible variables for the three query-documents

[1]prop.test command in R

Query	n	$p(T+)$	$p(T-)$	$p(T_0)$	$p(Rv+\|T+)$	$p(Rv+\|T-)$	$p_{ltp}(Rv+)$	$p_{T_0}(Rv+)$
Q_1	89	0.538	0.462	0.101	0.465	0.324	0.400	0.444
Q_5	89	0.684	0.316	0.112	0.907	0.200	0.683	0.800
Q_9	89	0.754	0.246	0.225	0.769	0.235	0.637	0.650

Table 6.3: Results for disjunction effect with topicality

Query	n	$p(Rb+)$	$p(Rb-)$	$p(Rb_0)$	$p(Rv+\|Rb+)$	$p(Rv+\|Rb-)$	$p_{ltp}(Rv+)$	$p_{Rb_0}(Rv+)$
Q_6	89	0.788	0.212	0.416	0.927	0.636	0.865	1.000
Q_{10}	89	0.455	0.545	0.506	0.900	0.542	0.705	0.756
Q_{11}	89	0.564	0.436	0.382	0.774	0.500	0.655	0.735

Table 6.4: Results for disjunction effect with reliability

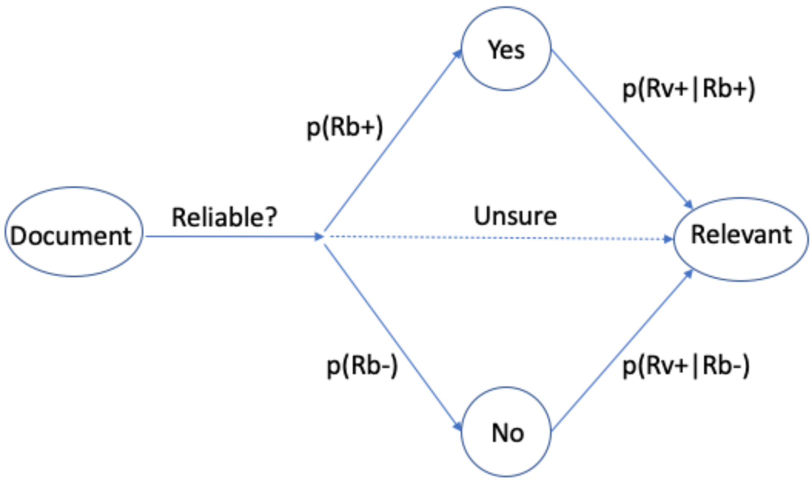


Figure 6.3: Experimental setup for disjunction effect in reliability

considered.

6.5 Experiment 2 - Interference of Reliability Categorisation on Relevance

6.5.1 Methodology

This experiment has been modelled along the lines of the interference in categorisation of faces experiment discussed before. The central idea is to ask one group of users to categorise the reliability of documents before judging the relevance and the other group is only asked to judge the relevance. Similar to the previous experiments, participants are shown an information need, a query and a custom designed document snippet. This experiment has a between subjects design with 200 participants uniformly randomly split into two groups. They are asked questions about relevance and reliability of the documents similar to the previous experiments, depending upon the group. The difference in design for the two groups is:

- **Group 1**: Participants are asked two questions about the document and query - first about the reliability of the document and then about the relevance. They have a binary choice of Yes or No for both the questions.
- **Group 2**: Participants in this group are only asked the question about relevance with a binary choice of Yes or No.

The aim is compare the prediction of marginal relevance probability calculated using LTP for Group 1 with the relevance probability obtained for Group 2.

6.5.2 Results and Discussion

The results of the experiment are tabulated in Table 6.5 along with the queries used. The condition where the group only judges the relevance of the document and not the reliability is termed as the decision-alone condition and the last column in the table lists the probability for this condition. Also note that the probability calculated through LTP is also the probability of relevance conditioned on a reliability judgement (irrespective of its value), i.e $p_{ltp}(Rv+) = p(Rv+|Rb)$. The results for three of the queries/documents (Q_2,Q_3, and Q_5) do not show any significant difference between the probability of relevance in the decision-alone condition and the probability of relevance when reliability has been judged. One can see that only for Q_4 do we get a significant difference between the probability of relevance as calculated from the experiment and that predicted using LTP, given a participant has judged or categorised its reliability. For the other three documents, the marginal probability of relevance can be predicted using LTP. For these documents, the act of judging or categorising reliability has no significant effect on the relevance judgement.

A careful analysis reveals that users' judgement of relevance is more influenced by topicality than reliability. Comparing probabilities of relevance with the probabilities of topicality and reliability, one can find that relevance probabilities are closer to probability of users finding a document topical than reliable. For example, if we consider results of the previous experiment as reported in tables 6.3 and 6.4, we find that relevance probabilities are within 20% of the probability of finding documents topically relevant. This figure is 19.70%, 16.96%, and 13.79% for queries Q_1, Q_5, and Q_9 respectively (calculate $\frac{|p(T+)-p_{T_0}(Rv+)|}{p(T+)}$). However, this measure is 28.17%, 66.15%, and 30.31% for queries Q_6, Q_{10}, and Q_1 respectively. For the current experiment reported in Table 6.5, one can see that the difference between $p(Rb+)$ and $p_{d-alone}(Rv+)$ is huge for queries Q_4 and Q_5 (0.46 vs 0.94 and 0.34 vs 0.61 respectively), but quite insignificant for queries Q_2 and Q_3 (0.61 vs 0.56 and 0.32 vs 0.36 respectively). It is possible that this similarity for Q_2 and Q_3 is merely a correlation and the probability of topical relevance is also similar to these. Indeed this speculation is found to be true in a later experiment (reported in Section 6.7. For Q_2 and Q_3, $p(T+)$ is 0.61 and 0.37 respectively (Averaging $P(T+)$ for first two rows for queries Q_2 and Q_3 respectively as topicality is the first question asked in those cases). For Q_4 and Q_5, it is 0.88 and 0.54 - closer to their probabilities of relevance. Note that all probabilities of relevance compared are that of the decision-alone condition for this experiment as this condition is devoid of the contextual influence created by judgement of reliability.

The violation for Q_4 can be explained thus: a high (94%) probability of relevance in the decision-alone condition is due to the document appearing highly topical and the participants likely considering topicality as the primary criteria to judge relevance. However, when participants are asked to judge reliability first, they also consider it when when judging relevance subsequently. As the difference between probability of reliability and probability of topicality is very large for Q_4 (0.88 vs 0.46), we see that the probability of relevance is pulled back significantly from 0.94 to 0.61. For the other queries, the pull back is not that significant because the difference between topicality and reliability is not that large. Hence following an intermediate judgement, the next question is answered from a different cognitive state. This brings us back to my hypothesis - we see another evidence that relevance judgements are influenced by dynamic contexts during the judgement process, like whether or not a user has considered reliability or not.

$n = 200$	$P(Rb+)$	$p(Rv+\|Rb+)$	$p(Rv+\|Rb-)$	$p(Rv+\|Rb)$	$p_{d-alone}(Rv+)$
Q_2	0.61	0.62	0.43	0.54	0.56
Q_3	0.32	0.59	0.29	0.39	0.36
Q_4	0.46	0.91	0.35	**0.61**	**0.94**
Q_5	0.34	0.85	0.47	0.60	0.61

Table 6.5: Interference of Reliability Categorisation on Relevance

	$P(Rb+)$	$p(T+)$	$p(Rv+\|Rb)$	$p_{d-alone}(Rv+)$
Q_2	0.61	0.61	0.54	0.56
Q_3	0.32	0.37	0.39	0.36
Q_4	0.46	0.88	**0.61**	**0.94**
Q_5	0.34	0.54	0.60	0.61

Table 6.6: Comparing Probabilities of Topicality, Reliability and Relevance

6.6 Experiment 3 - Relevance vs Usefulness Comparison

In the previous experiment, I have hypothesised that users consider topicality as the primary criterion in judging relevance than reliability. It might well be possible that participants considered relevance synonymous with topicality. Cooper [43] splits the concept of relevance into logical relevance, which is related to the topical bearing of the document, and utility - which has to do with the "ultimate usefulness of the piece of information". Hence it may be that the judgement of usefulness requires cognitive processing of a variety of other criteria and judgement of relevance is restricted to consideration of topical relevance. To gather more evidence for this, I perform a qualitative analysis of the difference between relevance and usefulness judgements in terms of the dimension considered by users. My hypothesis is that users give more consideration to dimensions other than topicality when asked to judge a document in terms of its usefulness than when judging the relevance. It must be noted that the terms usefulness and relevance are considered as in common English usage. The research question for this experiment is - Does judgement of usefulness involve consideration of a variety of dimensions than judgement of relevance?

6.6.1 Methodology

There are two user studies performed. In one study, participants judge relevance and in the second study, they judge usefulness of documents. Both studies have their design similar to the previous experiment. Thus, in each study participants are uniformly randomly split into two groups. The experimental steps for the two studies are:

- **Study 1 - Relevance**:
  - **Group 1:** Participants are asked two questions about the document and query - first about the reliability of the document and then about the relevance. They

have a binary choice of Yes or No for both the questions.

- **Group 2**: Participants in this group are only asked the question about relevance with a binary choice of Yes or No.
- **Feedback:** After the final question participants in both groups are optionally asked to write in brief (few words or sentences) about why they decided to judge relevance the way they did.

- **Study 2 - Usefulness**:

  - **Group 1:** Participants are asked two questions about the document and query - first about the reliability of the document and then about the usefulness. They have a binary choice of Yes or No for both the questions.
  - **Group 2**: Participants in this group are only asked the question about usefulness with a binary choice of Yes or No.
  - **Feedback:** After the final question participants in both groups are optionally asked to write in brief (few words or sentences) about why they decided to judge usefulness the way they did.

Although the feedback was optional, I found almost all participants elaborately discussing the reasons for their judgements. Some of the randomly selected feedbacks are listed in Table 6.7. In this experiment, I perform a qualitative analysis of the feedbacks obtained by manually classifying the feedbacks into one or more of the three dimensions of Topicality (label T), Reliability (R) and Understandability (U). Note that I also divide the feedbacks into different four types depending on the judgements of that document for the two questions. So for the first study, the judgement types are: Reliable and Relevant, Not Reliable and Relevant, Reliable and Not Relevant, Not Reliable and Not Relevant. For the second study, the judgement types are Reliable and Useful, Not Reliable and Useful, Reliable and Not Useful, Not Reliable and Not Useful. Simple heuristics are employed to decide the labels for each feedback. The classification labels are only among T, U and R, irrespective of whether the document is reliable or not reliable, relevant or not relevant, etc. This is because I want to capture consideration of these dimensions and not how they impacted the decisions. These rules are summarised thus:

- **Topicality**: Any comments on the relation between the content of the document and the query or information need.

- **Reliability**: Comments directly addressing the credibility of the source url, opinionatedness or lack of factuality of the content, scientific or lack of nature of the document.
- **Understandability**: Comments directly stating the difficulty in understanding the content due to it being too technical, poorly written.

Query	Feedback	Judgements	Classification
Q_6	The title was not misleading, the information showed referenced scientific research, the information given appeared to have a balanced viewpoint i.e. stating something is not significant but warrants further research	Reliable and Relevant	T, R
Q_6	The source isn't reliable from where it's from (hustlebustlenews doesn't seem that academic), and meningiomas, from what I've searched up, are benign. The study relating non-cancerous tumors to radio waves could skew search results for those that want to search the relation to cancer and radio waves instead.	Not reliable and Not Relevant	T, R
Q_{10}	There is no author listed, I have not heard of the site and therefore do not know its credibility. However I think this is relevant because it states a viewpoint which addresses the issue of democracy on the internet;as the statement cannot be outright disproved, its lack of credibility does not affect its relevance	Not Reliable and Relevant	T, R
Q_{10}	overly opinionated and not a trustworthy source – not trustworthy therefore not useful	Not Reliable and Not Useful	R
Q_{11}	The document did include some titbits of how fires were helpful for different types of forests and how the fires helped such as recycling nutrients	Reliable and Relevant	T
Q_{11}	It talks about facts and some outcomes of forest fires, seeming to be some positive	Reliable and Useful	T, R

Table 6.7: Random sample of feedbacks received

6.6.2 Results and Discussion

The objective in this experiment is to perform qualitative analysis to find out under what conditions are the different dimensions considered. Tables 6.8, 6.9 and 6.10 report the number of times the three dimensions of topicality (Top), reliability (Relb) and

Table 6.8: Comparing dimensions considered while judging relevance and usefulness for query Q_6

Dim	rel+/rb	rel-/rb	rel+	rel-
Top	31/32	12/14	33/33	14/16
Relb	14/32	3/14	4/33	2/16
Und	2/32	1/14	2/33	1/16

Dim	use+/rb	use-/rb	use+	use-
Top	15/19	15/22	24/26	13/19
Relb	14/19	12/22	5/26	4/19
Und	0/19	4/22	2/26	3/19

Table 6.9: Comparing dimensions considered while judging relevance and usefulness for query Q_{10}

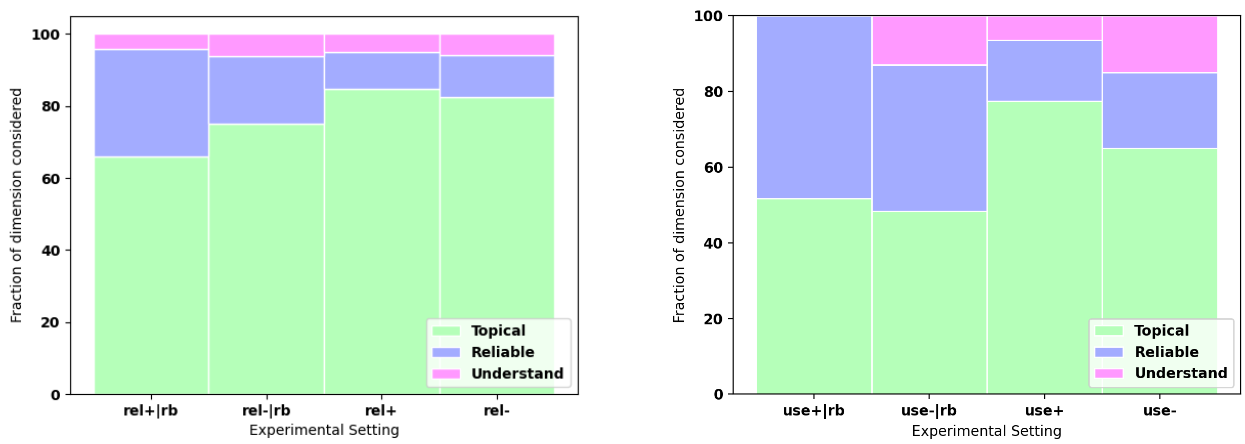
Dim	rel+/rb	rel-/rb	rel+	rel-
Top	24/27	14/15	33/33	5/5
Relb	7/29	4/15	1/33	1/5
Und	0/29	0/15	0/33	0/5

Dim	use+/rb	use-/rb	use+	use-
Top	25/28	10/16	25/25	14/17
Relb	11/28	13/16	1/25	4/17
Und	0/28	1/16	0/25	1/17

Table 6.10: Comparing dimensions considered while judging relevance and usefulness for query Q_{11}

Dim	rel+/rb	rel-/rb	rel+	rel-
Top	32/33	5/6	36/36	4/4
Relb	8/33	2/6	3/36	0/4
Und	0/33	1/6	0/36	0/4

Dim	use+/rb	use-/rb	use+	use-
Top	32/35	5/8	36/38	3/5
Relb	19/35	3/8	4/38	1/5
Und	0/35	1/8	0/38	1/5

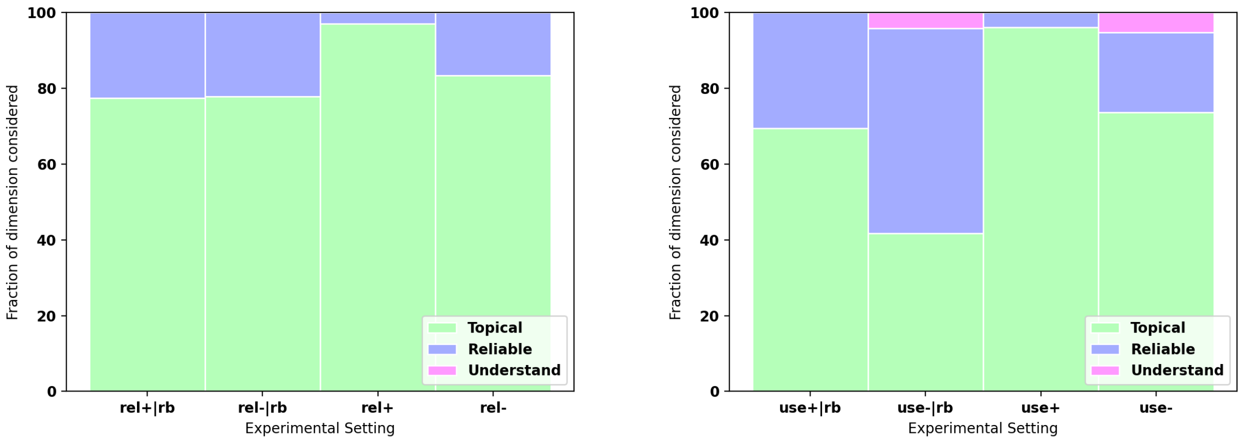


(a) Q_6 - Relevance

(b) Q_6 - Usefulness

Figure 6.4: Comparing fraction of dimensions considered for relevance and usefulness judgement of Q_6

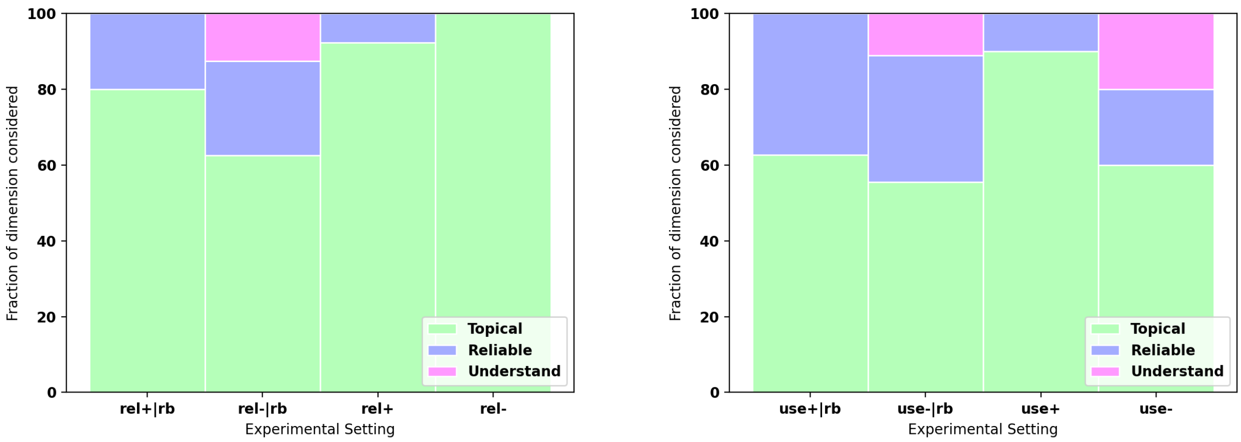
understandability (Und) are considered in four different combinations of questions asked and judgements made. Not only do I capture the relevance (useful) and non-relevance (not



(a) Q_{10} - Relevance

(b) Q_{10} - Usefulness

Figure 6.5: Comparing fraction of dimensions considered for relevance and usefulness judgement of Q_{10}



(a) Q_{11} - Relevance

(b) Q_{11} - Usefulness

Figure 6.6: Comparing fraction of dimensions considered for relevance and usefulness judgement of Q_{11}

useful) judgements, but also these judgements made after the judgement of reliability. The figures in these tables are difficult to compare for relevance and usefulness because of different number of relevance and usefulness judgements and also because of different number of responses for these conditions due to randomisation. Instead, let us refer to figures 6.4, 6.5 and 6.6 to analyse the results, which convert the absolute numbers in the tables to relative fraction so as to have a proper comparison across all settings. Each of the three figures represent a query with two sub-figures corresponding to the two studies relating to relevance and usefulness. The two sub-figures are stacked bar plots where the x-axis represent the different experimental conditions and y-axis represents the fraction of the total number of dimensions considered for the query as labelled from the feedbacks. There are two observations to be made here:

- Each chart within a figure show that participants consider more dimensions when they are judging relevance or usefulness **when the reliability question has been asked before**. It is easy to understand why this happens - participants who have been asked to judge reliability of the document are more aware of this dimension when judging the relevance or usefulness.

- Comparing the two charts within a figure, one can find that participants **consider more dimensions when judging usefulness than when judging relevance**, thus supporting my hypothesis for this experiment.

- There is also a trend in most cases where a judgement of non-relevance or non-usefulness involves consideration of more dimensions than a positive judgement. So, a negative judgement involves more cognitive processing.

Going through the feedbacks one by one, I could spot a big difference in the comments for relevance and usefulness scenarios. It seems as if while making usefulness judgements, participants become more involved and put more effort into judgement. They think about the document from a personal perspective - is it useful to me? That is why reliability becomes an important criteria. In case of relevance judgement it seems they make decision objectively, as someone on the outside, a third party - whether this document is relevant or not.

The analysis of the results of this experiment encourage me to revisit some of my previous experiments with the concept of usefulness instead of relevance. It maybe possible that the dynamic interactions between the difference relevance dimensions has a greater impact on the judgement of usefulness than on relevance, since those dimensions are considered more. So, phenomena like order effects, interference, etc. could see an impact on user judgement behaviours.

6.7 Experiment 4 - Investigating Impact of Incompatibility of Dimensions on Usefulness Judgements

6.7.1 Methodology

In this experiment I study the effect of consideration of different relevance dimensions and their permutations on judgement of document usefulness. The users are shown queries $Q_1 - Q_5$. I have one independent variable - permutation of questions about three dimensions (T,U and R) with 4 levels (corresponding to 4 conditions TUR, TRU, RUT, URT). The dependent variable is the usefulness judgement of a document (Us). I also have a control condition where a group of users are asked to only judge the usefulness. Thus the experiment procedure is the following:
For each uniform-randomly selected query in {Q_1 to Q_5}, participants are divided into following 5 groups:

1. Group 1: Asked questions T, U, R , Us
2. Group 2: Asked questions T, R, U , Us
3. Group 3: Asked questions R, U, T , Us
4. Group 4: Asked questions U, R, T, Us
5. Group 5: Asked question Us

where the questions are listed in Table 6.11. Each question is followed by a binary choice of 'Yes' or 'No'. The probabilities of answering 'Yes' or 'No' to a question are estimated as the frequencies of 'Yes' or 'No'. Each group above had 99-102 participants. This is because 102 participants were initially recruited for each condition (no two condition had the same participant to avoid learning effects), but response of some participants for some of queries in {Q_1 to Q_5} was rejected because of incompleteness. The reasoning behind choosing the above four permutations of relevance dimensions (T,U and R) is that all the three dimensions appear as the first question and the last question (before the usefulness question).

Label	Question
T	Is the document about the **topic** of the query or information need?
U	Is it easy to **understand** the information presented in the document?
R	Is the document **reliable**?
Us	Is the document **useful** for the query or information need?
Rv	Is this document **relevant** to the query or information need?

Table 6.11: Questions asked of participants about a document

Query	Order	$P(T+)$	$P(U+)$	$P(R+)$	$P(Us+)$	$P_{d-alone}(Us+)$
Q_1	TUR	0.55	0.55	0.94	0.46^a	$\mathbf{0.30^a}$
	TRU	0.51	0.53	0.92	0.42	
	RUT	0.59	0.51	0.91	0.43	
	URT	0.59	0.47	0.90	0.50^a	
Q_2	TUR	0.64	0.44	0.58	0.47	0.52
	TRU	0.58	0.51	$\mathbf{0.67^f}$	0.50	
	RUT	0.64	0.48	0.51^f	0.57	
	URT	0.52	0.49	0.50^f	0.53	
Q_3	TUR	0.39	0.50	$\mathbf{0.34^g}$	0.32	0.28
	TRU	0.34	0.46	0.32	0.27	
	RUT	0.30	0.48	0.22	0.24	
	URT	0.31	0.37	$\mathbf{0.21^g}$	0.25	
Q_4	TUR	0.88	0.45	$\mathbf{0.73^b}$	0.79	0.82
	TRU	0.88	0.48	0.61	0.83	
	RUT	0.82	0.40	0.66	0.78	
	URT	0.88	0.43	$\mathbf{0.60^b}$	0.79	
Q_5	TUR	0.56^d	$\mathbf{0.75^e}$	0.33	0.53	0.52
	TRU	0.51^d	$\mathbf{0.88^e}$	0.35	0.53	
	RUT	0.56^d	0.80	$\mathbf{0.44^c}$	0.56	
	URT	$\mathbf{0.69^d}$	0.79	$\mathbf{0.28^c}$	0.49	

Table 6.12: Comparing effect of dimensional categorisation and its order on document usefulness

6.7.2 Results and Discussion

The results obtained are listed in Table 6.12. As mentioned earlier, I have used 5 queries and asked participants questions about the 3 dimensions and document usefulness (Groups 1-4) or only the question about document usefulness (Group 5). The control group or Group 5 is also called the d-alone (decision alone) condition, as there is no categorisation of different dimensions involved. Hence the probability of participants who answered 'Yes' to the question Us in group 5 is represented as the column $P_{d-alone}(Us+)$

in Table 6.12.

One of the research question was that due to its constructive property, document judgement will differ if preceded by a categorisation decision. In this experiment the categorisation condition corresponds to the Groups 1,2,3 and 4. Here, the participants are first asked to categorise the document based on the different dimensions (T,U and R) and then make a judgement about the usefulness of the document (Us+). Note that all probability comparisons in this experiment use a 2-sample test for equality of proportions ($\alpha = 0.05$). In Group 5, which is also called the decision-alone condition, the participants are only asked to judge the usefulness of the document. We can see that for the first query-document pair (Q_1), the probability of usefulness in the decision-alone condition (0.30) is significantly lower than usefulness judgements in the categorisation conditions TUR (0.46) and URT (0.50). Other queries do not show any statistically significant difference between the probabilities of usefulness in the decision-alone condition and when it is preceded by judgement of the three dimensions.

However, statistically significant differences in judgement probabilities for the individual dimensions in different orders are found (in other words, Order effects). For a query, within a column, the statistically significantly different probabilities have the same superscript. When only two of them are different from each other, both of them are in bold. When more than one probability is significantly different from a given probability, this given probability only is marked in bold. All the others share the same superscript.

This again goes on to show the dynamic interactions between relevance dimensions. For example, for Q_5, we have the probability of the document being reliable as 0.44 when reliability is the first question asked and it is 0.28 when it is asked after Understandability. Consideration of understandability of the document significantly lowers the probability of finding it reliable. However, the aim of this experiment is to find out whether these dynamic interactions have any effect on the final decision of usefulness. As mentioned before, usefulness is used here instead of reliability because it involves consideration of a larger variety of dimensions. For all the queries, the probability of usefulness ($P(Us+)$) is not impacted when order of consideration of dimensions is changed (and even though order effects are found).

On performing basic regression analysis, it can be seen that topicality judgement still has a large bearing on the final decision of usefulness. One can see from Figure 6.7 the linear relationship between topicality and usefulness (bottom right hand corner plot with topicality on x-axis and usefulness on y-axis). Note that the figure is symmetric along the diagonal, all the three plots comparing usefulness and different dimensions can be found on the bottom row.

Table 6.13: Regression table using all three variables

	Dependent variable:
	use
top	1.016^{***}
	(0.063)
und	0.025
	(0.071)
rel	−0.070
	(0.048)
Constant	−0.058
	(0.056)
Observations	20
R^2	0.950
Adjusted R^2	0.941
Residual Std. Error	0.043 (df = 16)
F Statistic	101.118^{***} (df = 3; 16)
Note:	*p<0.1; **p<0.05; ***p<0.01

A multiple linear regression analysis using all the three dimensions gives R^2 (adjusted) = 0.941 (table 6.13 - note that the quantity in bracket below the intercept is the standard error). This means that these three dimensions can explain 94.10% of the variation in the usefulness probabilities $P(Us+)$. However, if one considers topicality alone, then adjusted $R^2 = 0.9380$ (table 6.14). That is, topicality alone explains 93.80% of the variation in $P(Us+)$.

Thus, to conclude, the interference of consideration of one or more dimensions on the usefulness judgement can be said to exist in case of some query-documents. Moreover, when more than one dimensions are considered, the order of consideration of dimensions can have an impact on the judgement of those dimensions. But these order effects do not have any significant impact on the final decision of document usefulness. Since in IR, the final user judgement of document relevance or usefulness matters more than the judgement of individual dimensions, one can say that it does not matter in what order

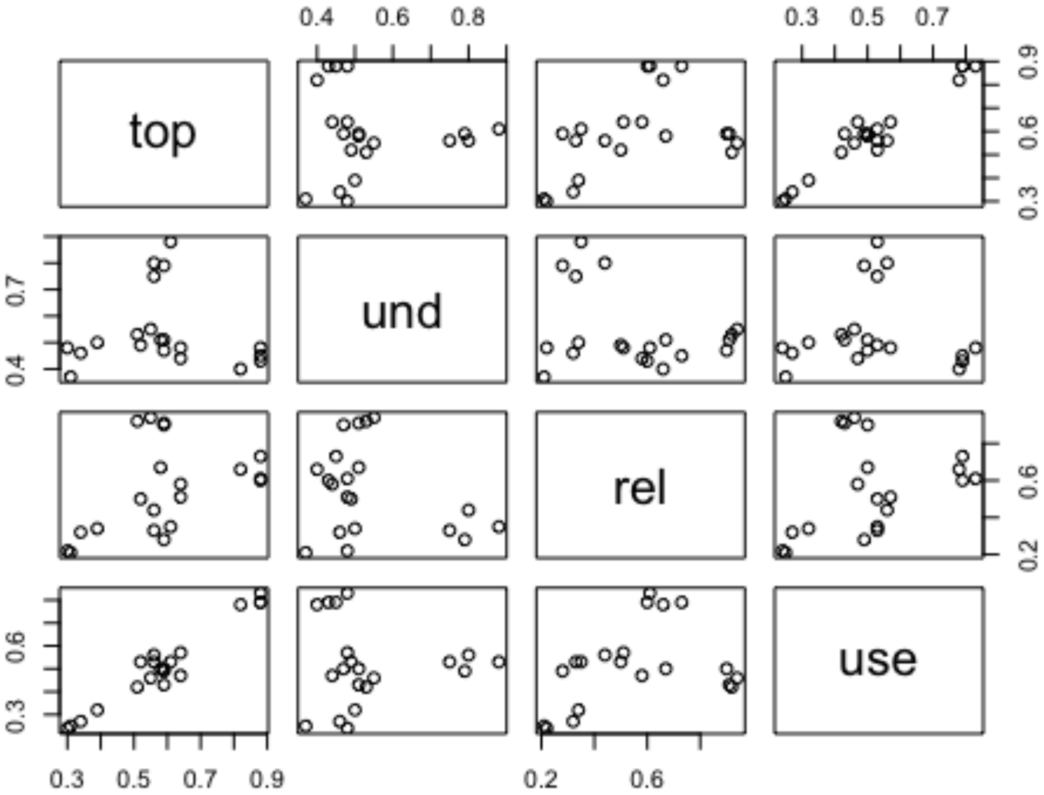


Figure 6.7: Multiple Regression plot of Usefulness against all three dimensions

these dimensions are considered or judged.

6.8 Conclusion

Continuing from previous chapters, I seek to gather more evidence of quantumness in document judgement data but with a focus on relevance/usefulness decisions. The first two experiments are analogous to experiments in cognitive science which were successfully modelled using QT. I adapt them into an IR scenario. While I don't find any evidence of a disjunction effect in IR for the small sample of queries I considered, the second experiment shows the interference effect of reliability classification on relevance judgement for one query. More experiments needs to be carried out in order to study this interference effect of classification. It can be investigated in a much larger and diverse sample of queries and documents.

Table 6.14: Regression table using only topicality

	Dependent variable:
	use
top	0.972^{***}
	(0.057)
Constant	-0.058
	(0.035)
Observations	20
R^2	0.941
Adjusted R^2	0.938
Residual Std. Error	0.044 (df = 18)
F Statistic	288.333^{***} (df = 1; 18)
Note:	*p<0.1; **p<0.05; ***p<0.01

The third experiment qualitatively analyses the difference between the number relevance dimensions considered when judging relevance versus when judging usefulness of documents. Users are asked to describe their reasons for making the decisions and their statements are manually classified by me into one or more of the three dimensions of Topicality, Reliability and Understandability. It is clear from the feedbacks received that participants consider reliability and understandability of the documents more when making a decision about their usefulness than relevance.

The final experiment seeks to design elements from all the experiments in this and the previous chapter. Using usefulness judgements, it seeks to investigate 1) interference effect of consideration of three dimensions on usefulness judgement and 2) effect of consideration of the three dimensions in different orders on usefulness. The results show interference effect of classification of three dimensions for one out of five queries but no impact of order effects on usefulness judgements. Preliminary analysis show that usefulness judgement is predominantly affected by topicality of the documents and that may be the reason that incompatibility between the other two dimensions have no effect on usefulness judgement. There is still a lot of scope for experimentation in this area. Apart from investigating a larger sample of different types of queries, one could also try different dimensions other than the three used in these experiments.

Chapter 7

Conclusion and Implications to IR

7.1 Summary

This thesis has investigated quantum-like phenomena in user's document judgements. Having introduced the motivation and need for such research, I briefly summarise the relevant literature in the field of Quantum-inspired Information Retrieval along with basic introduction to the Quantum mathematical framework. Following this are several experiments divided into four main chapters. The main research question which is the central theme of investigation throughout the thesis is: Can we find evidence of quantum-like phenomena in user cognitive behaviour in IR? On the other hand, the concept of multidimensional relevance in IR has been the focus of my experiments, as it bears similarity with measurement of different properties of quantum systems.

In the initial experiments, I look for quantumness in standard IR datasets such as TREC and other query log data. In Chapter 3, multidimensional relevance is represented in a Hilbert space. A re-ranking algorithm based on the Hilbert space and the collapse postulate of QT is constructed. The Hilbert space is also used to model order effects in query log data to verify its quantumness.

In Chapter 4, quantumness is formulated in terms of the concept of contextuality in Quantum Physics. Bell-type inequalities are used in QT to differentiate quantum and classical systems. In the first experiment in this chapter, I use the CHSH inequality as it is from QT and design a document judgement scenario within the query logs so as to reformulate the CHSH inequality with relevance judgements as variables. The

CHSH inequality is not violated and the possible reasons are discussed. In the second experiment of Chapter 4, the Contextuality-by-default inequality is used, which is a modified CHSH inequality suited for cognitive experiments. Successful violation of this inequality is seen for a crowdsourced user study. I also discuss some recent criticisms of the Contextuality-by-Default theory.

In Chapter 5, continuing with the new methodology of crowdsourced user studies, I adapt an existing experiment in Quantum Physics - the Stern Gerlach experiment in a multidimensional relevance judgement scenario. It is another protocol to study interference and incompatibility of judgements and makes use of complex numbers to construct representation models for user judgements. This also helps in comparing relevance predictions using quantum and Bayesian probabilities. In the second experiment of this chapter, the complex-valued Hilbert space model of the first experiment is utilised to design a user study to test for violation of one of the axioms of the classical probability theory. There is a clear violation of these axioms which can predicted using quantum models.

While Chapter 5 is successful in giving evidence of quantum-like phenomena of interference, incompatibility, etc. in user judgements, it only takes into account judgement of different relevance dimensions of topicality, reliability, etc. In Chapter 6, the focus is on relevance itself and also a related concept of usefulness. Different user studies based on certain quantum cognitive experiments are performed to find the impact of quantumness of multidimensional judgements on the final user decisions of relevance and usefulness. It is seen that while topicality is compatible with relevance, reliability is seen to interfere with relevance and usefulness for a small sample of query-documents. Another important observation is that while judgement of different dimensions are subject to order effects, it has no bearing on the final judgement of document usefulness or the utility of documents.

7.2 Key Findings and Limitations

The main aim of the research reported in this thesis is to find evidence of quantum-like phenomena in user judgements in IR. This has been broken down into four research questions. Below, I report the key findings and limitations of the work performed with respect to these four research questions:

RQ 1: Can standard IR datasets or query logs provide evidence of incompatibility between judgement of different dimensions of relevance?

Key Findings: This question is answered in Chapter 3. Based on the two experiments in that chapter, it is clear that it is very difficult to find evidence of quantum-like phenomena from standard IR datasets and query log data. This is because these types of datasets lack rich contextual information. For example, relevance judgements are recorded in a single context - a document is judged only by a single user or it is judged independently of other documents, or the relevance dimensions considered are not captured. Lack of context and judgement dynamics makes it very difficult to directly find evidence of quantumness such as incompatibility or interference. Instead, I design a specific scenario to model incompatibility as order effects. Also, a re-ranking algorithm based on the concept of multiple incompatible basis and collapse of indefinite cognitive state to a definite state of judgement is constructed, which performs better on the NDCG metric than some baselines. A key contribution of these experiments is a method for constructing a Hilbert space from query log data. It combines the representation of document judgements along different relevance dimensions into a single vector space and allows for the modelling of incompatible judgement perspectives should they exist in the data. Hilbert space forms the foundation of the quantum framework and can be further used for performing different experiments to verify quantumness of data and to build quantum probabilistic models.

Limitations: One can always use different variety of datasets in these experiments. For constructing the Hilbert space, user's preference for relevance dimensions is modelled using the scores of Learning to Rank algorithm based on some predefined features. It can be argued whether that is a true representation of user's judgements. Also, lack of complex-valued vectors in the Hilbert space do not give it much advantage over classical vector space models. In the second experiment, it is difficult to establish order effects due to lack of relevance judgements in different orders or any other sequential context. The evidence of order effects presented is found in a very tiny fraction of the total number of queries in the logs and could very well be due to randomness, other noise or even be some other phenomena.

RQ 2: How to verify quantumness of IR data using no-go theorems of Quantum Theory?

Key Findings: In Chapter 4, I successfully formulate Bell-type inequalities using relevant judgement scenarios. The first experiment utilises query log data and the CHSH inequality. It fails to show any violation, the reason being the lack of relevance judgement data over a pair of documents. This problem is overcome in the second experiment, where a modified Bell-type inequality (CbD inequality) shows violation on a specifically designed relevance judgement scenario. Violation of CbD inequality in relevance judgements of documents reveals the in-deterministic nature of relevance, which comes to the fore when users are faced with ambiguity in decision-making. Although the behaviour seems to be intuitive - violation of CbD inequality shows that classical probability theory lacks in sufficiently capturing the contextuality in relevance judgements. Evidence of contextuality does not mean that the judgement behaviour is new or different. It simply means that models based on the quantum mathematical framework are needed to capture the consequences of contextuality in sequential decision-making - incompatibility, interference, etc. These cannot be accurately modelled using classical methods.

Limitations: One of the limitations of crowdsourced user studies is the quality of the data received. The sample of queries and documents considered is also very small. It can be also be argued as to what extent do these studies represent real world scenarios. I have also mentioned in Chapter 4 some of the recent criticisms against CbD theory as in it fails to eliminate all possible classical influences.

RQ 3: How to adapt existing experiments from quantum theory to study dynamic interactions between relevance dimensions so as to reveal quantum-like nature of user cognitive states?

Key Findings: The experiments in Chapter 5 as an answer to **RQ 3** show that document judgements show quantum-like nature when probed in a particular way. This means that internal cognitive states are quantum-like and are revealed during judgements under ambiguity. In many other cases, there may be no quantum-like phenomena exhibited and it may be sufficient to model the user interactions with classical methods. Both the violation of Kolmogorovian probability axiom and presence of negative probabilities in the Wigner function verify this quantumness. A major implication of the findings of these experiments to multidimensional relevance is that it establishes the dynamic, contextual nature of multidimensional relevance. One can see that these dimensions are not an objective property of documents. They are rather subjective and contextual. A user may have a different judgement of understandability depending upon whether they consider it reliable or not. In fact, whether or not reliability is considered before judging understandability also impacts the understandability judgement (therefore order effects occur).

Limitations: Some of the limitations of the experiments used for answering **RQ 3** are same as those above for **RQ 2**, i.e. small sample of queries and documents, noise in collecting crowdsourced data and lack of real world user interaction scenario. Apart from this, the frequentist interpretation of probability of judgements creates limitations for construction of Hilbert space models constructed. In order to construct a Hilbert space for each document, one requires to collect multidimensional judgements of a large number of users. Doing it for a large number of query-document pairs would require tremendous resources. Hence it is not scalable to real world IR modelling. An alternative approach needs to be devised, which can approximate the Hilbert spaces for each document using significantly lesser amount of data points.

RQ 4: Do the quantum effects observed in the interaction between relevance dimensions have any effect on the final decision of document relevance?

Key Findings: Chapter 6 which answers **RQ 4** has mostly negative results. Firstly, we see that topicality is compatible with relevance and usefulness. It doesn't cause any interference. Reliability and relevance are found to be incompatible for only one of the queries-documents. Whilst order effects are observed between relevance dimensions and their dynamic and contextual nature is again established, it has no bearing on the usefulness judgements. This is also a significant finding. In one way, it means that designers of IR algorithms do not need to worry about the impact the quantumness arising out of the interaction between relevance dimensions on relevance or usefulness judgements. It only needs to be taken into account when the judgement of any of the relevance dimensions are to be modelled or predicted, e.g. in predicting reliability of documents.

Limitations: Although negative results are obtained for most the experiments in Chapter 6, one needs to keep in mind, again, the small set of queries and documents used. It could be possible to obtain different results for other samples of query-documents. Also, it is possible that lack of positive results may be due to noise in crowdsourced data and replication of these experiments in different circumstances produce significant violation of classical probability.

7.3 Future Work and Implications to IR

7.3.1 Non-classical IR Models

The experiments reported in Chapter 5 and previous studies on order effects in IR primarily show that consideration of a particular relevance dimension has an effect on judgement of a subsequent dimension. This dynamic interaction warrants a non-classical mathematical framework for building probabilistic models of judgements. An interesting experiment proposed for the future is to combine the the re-ranking algorithm developed in Chapter 3 with the complex-valued Hilbert space in Chapter 5. The phase of the complex number can capture latent information about cognitive states like ambiguity and phenomenon like incompatibility. Complex-valued vector space models have been fused with learning algorithms in the area of natural language processing [94, 170]. Future work could incorporate complex numbers in current state-of-the-art IR models.

Quantum probabilistic models can replace Bayesian models used in IR algorithms for ranking and evaluation. For example, in [116], 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 [115, 196]. 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 call 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.

7.3.2 Multidimensional Interference and Information Literacy

The role of an IR system is to provide the user with relevant information which helps the user accomplish some task or make a well-informed decision. Therefore, it is important that users are able to reconcile the different dimensions of relevance in a way which minimises ambiguity and enables them to select the most relevant information for their need. 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. E.g., 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. IR systems can help reduce uncertainty in judging information objects by providing the users extra information about different relevance dimensions. In another example, a news retrieval system can provide, along with each article, a score for dimensions like credibility, readability (understandability), factuality, opinionated-ness, etc. These scores of Information Nutrition [68] can be calculated based on information object content, or/and be collected through user provided data. This can help users reduce uncertainty in judging the information or information object, and, more importantly make them consider the optimised sequence of dimension in order to make the best possible judgement. However, one needs to be careful in considering these scores to be fixed properties of the documents.

7.3.3 Dimensional Preferences to Improve User Modelling

The order of consideration of relevance dimensions by a user can be ascertained and documents be ranked in that order. For example, for the query 'Game of Thrones news', a popular Television series, a cautious user might look for highly reliable articles but a more adventurous user might consider articles which appear 'Interesting', talking about different conspiracy theories or spoilers about the TV show, the credibility of such articles being a secondary criterion. The IR system can present documents according to the dimensional preference of the user if it is able to profile the user appropriately. On other hand, the order of preference of relevance criteria might be largely independent of the user but depend upon the type of information need, e.g., for 'Visa to US' queries, users may always prefer reliability of the source as the first criterion to judge (along with topicality, of course). A related approach has been proposed in [49, 50] where, taking inspiration from the Multi-Criteria Decision Making (MCDM) approach in the area of Decision Theory, a prioritised aggregation operator is defined for different types of users. My work informs this approach by exploring the dynamic interaction between the relevance dimensions themselves.

7.3.4 Relevance Assessment and Measurements

This work can also inform the design of relevance assessment collection procedure in many ways. Firstly, the annotators could be given a list of relevance dimensions and asked to rate the relevance of a document along each of these dimensions, apart from a final relevance judgement. This will help us capture more context surrounding the relevance assessment of a document. As we saw in chapters 5 and 6, whether or not users consider particular relevance dimensions has an effect on their final relevance judgements. Hence, when relevance judgements of two users differ, consideration (or lack of) of different relevance dimensions can be a factor behind it. Fixing the relevance dimensions considered can help in eliminating this noise. Also, instead of having a fixed list, they can choose the order of dimensions. This method can reveal whether users prefer a particular order and whether having a fixed order or a random order of dimensions considered for relevance judgement lead to the same or different final judgement.

Secondly, as seen in a smaller scale in this thesis, usefulness judgements can replace relevance judgements for the entire collection and IR systems based on usefulness and relevance can be compared for user satisfaction or task completion metrics.

An important finding of this thesis is that certain properties of documents, e.g. reliability and understandability dimensions are not objective properties and can be influenced by interference with each other. For example, if the reliability of a document is different when understandability is considered low, than when it is considered high or not considered at all, then in what way can we pre-define reliability of the document? It becomes a dynamically changing property. Perhaps, these dimensions can be better quantified probabilistically and with more knowledge of the user context. For example, in addition to usual factors, reliability of a document can also be conditioned on user's understandability using background knowledge of user's expertise, past interactions, etc.

7.3.5 Incorporating Document Interference in Ranking Algorithms

As discussed, in Chapter 6, we fail to see relevance or usefulness of documents affected by change in order of consideration of dimensions. However, one needs to study the effect of order of documents themselves on relevance judgements. Order effects in document judgements has been investigated from a quantum cognitive point of view [173], where the relevance of a document is different when another document or documents have been judged before it. This could be further investigated using larger crowdsourced user studies. In some state-of-the-art ranking models, the scoring functions take into account all other documents in the collection, e.g., score of a document d_1 in a collection of document $d_1,...,d_n$ is calculated as a multivariate group function $f(d_1|d_2,d_3,..,d_n)$. However, this scoring function is indifferent to the order of documents, i.e., $f(d_1|d_2,d_3,..,d_n) = f(d_1|d_3,d_n,..,d_2)$. Should order effects, exist and relevance be impacted by not only by the set of other documents but also their order, the scoring function should be modified to account for these effects. Again, a quantum-inspired model, with non-commutative features would be better equipped.

7.3.6 Detecting and Eliminating Order Effects

While the above subsection discusses the effect of different order of documents on relevance judgements, the order of documents may also have an impact on the cognitive aspects of information consumption, opinion formation, etc. For example, suppose that news documents related to a controversial topic (e.g. Brexit) be divided into two categories - one favouring the argument and the other against the argument. An order effect can affect opinion formation of a neutral user if documents of one category are shown before the documents of the other category. Hence it can be used in a negative way to manipulate or persuade the users. Another example is the case of item reviews in an e-commerce system. Listing positive reviews of an item before negative reviews can lead to users forming a positive opinion of the item. Or, grouping the comments of a particular sentiment or stance below a social media post can influence a reader's interpretation of the post. There are two areas where IR systems can be improved - one is to be able to predict the presence of order effects over a sequence of information objects like documents or text snippets. That is, to predict that different orders of presentation of information can lead to a different judgements decision or opinions. Supervised learning is one way in which such algorithms to detect order effects can be developed. Training examples of information objects in different orders can be collected via user studies along with the label corresponding to whether an order effect occurred or not (Order effect will be said to occur when different order of presentation leads to different decisions). Some of the hand-crafted features can be the difference in topicality or sentiment scores between document pairs, etc. There needs to be studies carried out to understand what features of information objects cause order effects. Another area in IR where one can work is figuring out the appropriate ranking such that order effects are eliminated, without significantly reducing the overall ranking metric. There might be naive ways to eliminate bias due to ordering e.g. alternating item reviews of different sentiments. However for some other type of IR systems, like search engines, it might not lead to an optimal performance.

7.3.7 Exploiting Order Effects - The Case of Nudging

Order effects need not be eliminated always. Sometimes, they might be used to the IR system stakeholder's advantage. An interesting case is to use them to persuade or 'nudge' users into making a particular decision. For example, in a movie recommendation scenario, assume that we have four movies to be recommended to users and the IR system ranks them as M_1, M_2, M_3, and M_4, in the decreasing order of relevance score for the user. Generally, most of the IR systems will follow this order when recommending the movies to the user. However, if it is desired that the user selects movie M_1, then order effects can be utilised to further increase the probability of user finding it relevant. As mentioned in section 2.3.3 and in [173], for the query 'Albert Einstein', the relevance of the document with the topic 'Theory of Relativity' is more when it is shown after the topic 'Issac Newton', than when shown before. A particular type of order effect - comparison effect takes place and the relevance of a document is perceived to be higher in comparison when judged after a less relevant document. A similar arrangement can be used in this movie recommendation scenario. So, if we show the user movies in the order, say, M_2, M_4, M_1, and M_3, the movie M_1 might appear more relevant to the user when compared to the lesser relevant movies he or she has browsed before. It remains to be seen whether position bias will have any interaction with such a method. This method can also be used in advertising and e-commerce scenarios. In document ranking, this can be utilised in news search or recommendation engines. This principle has been often used in marketing strategies since many decades and has been recently theorised within behavioural sciences as the 'Nudge Theory'. Within IR and recommender systems, this approach has been applied in food recipe recommendation where the nutrition labels and attractiveness of food items are combined to nudge users to choose low fat recipes that look attractive [62]; and also in nudging users to adopt better privacy practices in search [194]. However, there are ethical concerns of this practice, especially when applied to online news platforms, in light of the current discussion on misinformation and disinformation [141]. Manipulating the user's decisions and opinions by exploiting the order biases goes deeper than fake news. In this case, the facts can be correct but still the users can be manipulated by the way they are presented. However, there is still scope for research in this direction within IR, as IR systems can play a proactive role in preventing such misuse by being better equipped to detect it.

Bibliography

[1] Samson Abramsky. 2013.
Relational Databases and Bell's Theorem. In *In Search of Elegance in the Theory and Practice of Computation - Essays Dedicated to Peter Buneman (Lecture Notes in Computer Science)*, Val Tannen, Limsoon Wong, Leonid Libkin, Wenfei Fan, Wang-Chiew Tan, and Michael P. Fourman (Eds.), Vol. 8000. Springer, 13–35.
`https://doi.org/10.1007/978-3-642-41660-6_2`

[2] Samson Abramsky and Adam Brandenburger. 2011.
The sheaf-theoretic structure of non-locality and contextuality.
New Journal of Physics 13, 11 (nov 2011), 113036.
`https://doi.org/10.1088/1367-2630/13/11/113036`

[3] Diederik Aerts. 2000.
Foundations of Physics 30, 9 (2000), 1387–1414.
`https://doi.org/10.1023/a:1026449716544`

[4] Diederik Aerts. 2009.
Quantum structure in cognition.
Journal of Mathematical Psychology 53, 5 (2009), 314 – 348.
`https://doi.org/10.1016/j.jmp.2009.04.005`
Special Issue: Quantum Cognition.

[5] Diederik Aerts, Jonito Aerts Arguëlles, Lester Beltran, Suzette Geriente, Massimiliano Sassoli de Bianchi, Sandro Sozzo, and Tomas Veloz. 2018.
Spin and Wind Directions I: Identifying Entanglement in Nature and Cognition.
Foundations of Science 23, 2 (01 Jun 2018), 323–335.
`https://doi.org/10.1007/s10699-017-9528-9`

[6] Diederik Aerts and Marek Czachor. 2004.

Quantum aspects of semantic analysis and symbolic artificial intelligence.
Journal of Physics A: Mathematical and General 37, 12 (2004), L123.
http://stacks.iop.org/0305-4470/37/i=12/a=L01

[7] Diederik Aerts and Sandro Sozzo. 2011.
Quantum Structure in Cognition: Why and How Concepts Are Entangled. In *Quantum Interaction*, Dawei Song, Massimo Melucci, Ingo Frommholz, Peng Zhang, Lei Wang, and Sachi Arafat (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 116–127.

[8] Diederik Aerts and Sandro Sozzo. 2013.
Quantum Entanglement in Concept Combinations.
International Journal of Theoretical Physics 53, 10 (dec 2013), 3587–3603.
https://doi.org/10.1007/s10773-013-1946-z

[9] Mojtaba Aliakbarzadeh, Kirsty Kitto, and Peter Bruza. 2019.
Is contextuality about the identity of random variables?
arXiv preprint arXiv:1909.13048 (2019).

[10] Masanari Asano, Takahisa Hashimoto, Andrei Khrennikov, Masanori Ohya, and Yoshiharu Tanaka. 2014.
Violation of contextual generalization of the Leggett–Garg inequality for recognition of ambiguous figures.
Physica Scripta T163 (dec 2014), 014006.
https://doi.org/10.1088/0031-8949/2014/t163/014006

[11] F. Gregory Ashby and Nancy A. Perrin. 1988.
Toward a unified theory of similarity and recognition.
Psychological Review 95, 1 (1988), 124–150.
https://doi.org/10.1037/0033-295x.95.1.124

[12] Harald Atmanspacher and Thomas Filk. 2019.
Contextuality Revisited: Signaling May Differ From Communicating.
Springer International Publishing, Cham, 117–127.
https://doi.org/10.1007/978-3-030-21908-6_10

[13] Harald Atmanspacher and Hartmann Römer. 2012.
Order effects in sequential measurements of non-commuting psychological observables.

Journal of Mathematical Psychology 56, 4 (2012), 274–280.

[14] Carol L Barry. 1994.
User-defined relevance criteria: An exploratory study.
Journal of the American Society for Information Science 45, 3 (1994), 149–159.

[15] Carol L Barry and Linda Schamber. 1998.
Users' criteria for relevance evaluation: a cross-situational comparison.
Information processing & management 34, 2-3 (1998), 219–236.

[16] Irina Basieva, Víctor H. Cervantes, Ehtibar N. Dzhafarov, and Andrei Khrennikov. 2018.
True Contextuality Beats Direct Influences in Human Decision Making.
arXiv:arXiv:1807.05684

[17] J. S. Bell. 1964.
On the Einstein Podolsky Rosen paradox.
Physics Physique Fizika 1 (Nov 1964), 195–200.
Issue 3. `https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195`

[18] Niels Bohr. 1937.
Causality and Complementarity.
Philosophy of Science 4, 3 (1937), 289–298.
`https://doi.org/10.1086/286465` arXiv:https://doi.org/10.1086/286465

[19] Pia Borlund. 2003.
The IIR evaluation model: a framework for evaluation of interactive information retrieval systems.
Inf. Res. 8 (2003).

[20] Max Born. 1926.
Zur Quantenmechanik der Stoßvorgänge.
Zeitschrift für Physik 37, 12 (dec 1926), 863–867.
`https://doi.org/10.1007/bf01397477`

[21] Peter Bruza and Vivien Chang. 2014.
Perceptions of document relevance.
Frontiers in Psychology 5 (2014), 612.
`https://doi.org/10.3389/fpsyg.2014.00612`

[22] Peter Bruza and Vivien Chang. 2014.
Perceptions of document relevance.
Frontiers in Psychology 5 (2014), 612.
`https://doi.org/10.3389/fpsyg.2014.00612`

[23] Peter Bruza and Vivien Chang. 2014.
Perceptions of document relevance.
Frontiers in Psychology 5 (2014), 612.
`https://doi.org/10.3389/fpsyg.2014.00612`

[24] Peter Bruza and Vivien Chang. 2014.
Perceptions of document relevance.
Frontiers in Psychology 5 (2014), 612.
`https://doi.org/10.3389/fpsyg.2014.00612`

[25] Peter Bruza and Richard J. Cole II. 2005.
Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning. In *We Will Show Them! Essays in Honour of Dov Gabbay, Volume One*. 339–362.

[26] Peter Bruza, Kirsty Kitto, Douglas Nelson, and Cathy McEvoy. 2009.
Is there something quantum-like about the human mental lexicon?
Journal of Mathematical Psychology 53, 5 (oct 2009), 362–377.
`https://doi.org/10.1016/j.jmp.2009.04.004`

[27] Peter D. Bruza and Lauren Fell. 2018.
Are decisions of image trustworthiness contextual? A pilot study. (June 2018).
`https://eprints.qut.edu.au/119547/`

[28] P. D. Bruza, K. Kitto, B. Ramm, L. Sitbon, D. Song, and S. Blomberg. 2011.
Quantum-like non-separability of concept combinations, emergent associates and abduction.
Logic Journal of IGPL 20, 2 (jan 2011), 445–457.
`https://doi.org/10.1093/jigpal/jzq049`

[29] P. D. Bruza, K. Kitto, B. Ramm, L. Sitbon, D. Song, and S. Blomberg. 2011.
Quantum-like non-separability of concept combinations, emergent associates and abduction.
Logic Journal of IGPL 20, 2 (jan 2011), 445–457.

https://doi.org/10.1093/jigpal/jzq049

[30] Peter D. Bruza, Kirsty Kitto, Brentyn J. Ramm, and Laurianne Sitbon. 2015.
A probabilistic framework for analysing the compositionality of conceptual combinations.
Journal of Mathematical Psychology 67 (2015), 26 – 38.
https://doi.org/10.1016/j.jmp.2015.06.002

[31] Curt Burgess, Kay Livesay, and Kevin Lund. 1998.
Explorations in context space: Words, sentences, discourse.
Discourse Processes 25, 2-3 (jan 1998), 211–257.
https://doi.org/10.1080/01638539809545027

[32] Jerome R. Busemeyer and Peter D. Bruza. 2012.
Quantum Models of Cognition and Decision (1st ed.).
Cambridge University Press, New York, NY, USA.

[33] Jerome R. Busemeyer, Emmanuel M. Pothos, Riccardo Franco, and Jennifer S. Trueblood. 2011.
A quantum theoretical explanation for probability judgment errors.
Psychological Review 118, 2 (2011), 193–218.
https://doi.org/10.1037/a0022542

[34] Jerome R. Busemeyer, Emmanuel M. Pothos, Riccardo Franco, and Jennifer S. Trueblood. 2011.
A quantum theoretical explanation for probability judgment errors.
Psychological Review 118, 2 (2011), 193–218.
https://doi.org/10.1037/a0022542

[35] Jerome R. Busemeyer, Emmanuel M. Pothos, Riccardo Franco, and Jennifer S. Trueblood. 2011.
A quantum theoretical explanation for probability judgment errors.
Psychological Review 118, 2 (2011), 193–218.
https://doi.org/10.1037/a0022542

[36] Adán Cabello, Simone Severini, and Andreas Winter. 2014.
Graph-Theoretic Approach to Quantum Correlations.
Physical Review Letters 112, 4 (jan 2014).
https://doi.org/10.1103/physrevlett.112.040401

[37] Annalina Caputo, Benjamin Piwowarski, and Mounia Lalmas. 2011.
A Query Algebra for Quantum Information Retrieval. In *Proceedings of the 2nd Italian Information Retrieval (IIR) Workshop, Milan, Italy, January 27-28, 2011*.
http://ceur-ws.org/Vol-704/19.pdf

[38] Jaime Carbonell and Jade Goldstein. 1998.
The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98)*. ACM, New York, NY, USA, 335–336.
https://doi.org/10.1145/290941.291025

[39] Víctor H. Cervantes and Ehtibar N. Dzhafarov. 2018.
Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices.
Decision 5, 3 (jul 2018), 193–204.
https://doi.org/10.1037/dec0000095

[40] Stanley F. Chen and Joshua Goodman. 1999.
An empirical study of smoothing techniques for language modeling.
Computer Speech & Language 13, 4 (oct 1999), 359–393.
https://doi.org/10.1006/csla.1999.0128

[41] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. 1969.
Proposed Experiment to Test Local Hidden-Variable Theories.
Phys. Rev. Lett. 23 (Oct 1969), 880–884.
Issue 15. https://doi.org/10.1103/PhysRevLett.23.880

[42] C.W. CLEVERDON. 1960.
The ASLIB CRANFIELD RESEARCH PROJECT ON The COMPARATIVE EFFICIENCY OF INDEXING SYSTEMS.
Aslib Proceedings 12, 12 (dec 1960), 421–431.
https://doi.org/10.1108/eb049778

[43] W.S. Cooper. 1971.
A definition of relevance for information retrieval.
Information Storage and Retrieval 7, 1 (1971), 19 – 37.

https://doi.org/10.1016/0020-0271(71)90024-6

[44] System Development Corporation. 1967.
Experimental Studies of Relevance Judgments: Final Report.
System Development Corporation.
https://books.google.co.uk/books?id=80GRRwAACAAJ

[45] Erica Cosijn. 2009.
Relevance Judgments and Measurements.
In *Encyclopedia of Library and Information Sciences*. CRC Press, 4512–4519.

[46] Erica Cosijn and Peter Ingwersen. 2000.
Dimensions of relevance.
Information Processing & Management 36, 4 (2000), 533 – 550.
https://doi.org/10.1016/S0306-4573(99)00072-2

[47] Nick Craswell, W. Bruce Croft, Maarten de Rijke, Jiafeng Guo, and Bhaskar Mitra. 2017.
Neural information retrieval: introduction to the special issue.
Information Retrieval Journal 21, 2-3 (nov 2017), 107–110.
https://doi.org/10.1007/s10791-017-9323-9

[48] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008.
An Experimental Comparison of Click Position-Bias Models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM '08)*. Association for Computing Machinery, New York, NY, USA, 87–94.
https://doi.org/10.1145/1341531.1341545

[49] Célia da Costa Pereira, Mauro Dragoni, and Gabriella Pasi. 2009.
Multidimensional relevance: A new aggregation criterion. In *European Conference on Information Retrieval*. Springer, 264–275.

[50] Célia da Costa Pereira, Mauro Dragoni, and Gabriella Pasi. 2012.
Multidimensional Relevance: Prioritized Aggregation in a Personalized Information Retrieval Setting.
Inf. Process. Manage. 48, 2 (March 2012), 340–357.
https://doi.org/10.1016/j.ipm.2011.07.001

[51] P. A. M. Dirac. 1982.

The Principles of Quantum Mechanics (International Series of Monographs on Physics).
Clarendon Press.
https://www.amazon.com/Principles-Quantum-Mechanics-International-Monographs/dp/0198520115?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0198520115

[52] Ehtibar N. Dzhafarov, Víctor H. Cervantes, and Janne V. Kujala. 2017.
Contextuality in canonical systems of random variables.
Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 375, 2106 (oct 2017), 20160389.
https://doi.org/10.1098/rsta.2016.0389

[53] Ehtibar N. Dzhafarov and Janne V. Kujala. 2013.
On Selective Influences, Marginal Selectivity, and Bell/CHSH Inequalities.
Topics in Cognitive Science 6, 1 (nov 2013), 121–128.
https://doi.org/10.1111/tops.12060

[54] Ehtibar N Dzhafarov and Janne V Kujala. 2014.
Contextuality is about identity of random variables.
Physica Scripta T163 (dec 2014), 014009.
https://doi.org/10.1088/0031-8949/2014/t163/014009

[55] Ehtibar N. Dzhafarov and Janne V. Kujala. 2016.
Probabilistic foundations of contextuality.
Fortschritte der Physik 65, 6-8 (sep 2016), 1600040.
https://doi.org/10.1002/prop.201600040

[56] Ehtibar N. Dzhafarov and Janne V. Kujala. 2017.
Contextuality-by-Default 2.0: Systems with Binary Random Variables.
In *Quantum Interaction*. Springer International Publishing, 16–32.
https://doi.org/10.1007/978-3-319-52289-0_2

[57] Ehtibar N. Dzhafarov, Janne V. Kujala, and Victor H. Cervantes. 2016.
Contextuality-by-Default: A Brief Overview of Ideas, Concepts, and Terminology.
In *Quantum Interaction*, Harald Atmanspacher, Thomas Filk, and Emmanuel Pothos (Eds.). Springer International Publishing, Cham, 12–23.

[58] Ehtibar N. Dzhafarov, Janne V. Kujala, Víctor H. Cervantes, Ru Zhang, and Matt Jones. 2016.
On contextuality in behavioural data.
Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 2068 (apr 2016), 20150234.
https://doi.org/10.1098/rsta.2015.0234

[59] E. N. Dzhafarov, Ru Zhang, and Janne Kujala. 2015.
Is there contextuality in behavioural and social systems?
Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 2058 (jan 2015), 20150099.
https://doi.org/10.1098/rsta.2015.0099

[60] A. Einstein, B. Podolsky, and N. Rosen. 1935.
Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?
Phys. Rev. 47 (May 1935), 777–780.
Issue 10. https://doi.org/10.1103/PhysRev.47.777

[61] Michael Eisenberg and Carol Barry. 1988.
Order effects: A study of the possible influence of presentation order on user judgments of document relevance.
Journal of the American Society for Information Science 39, 5 (sep 1988), 293–300.
https://doi.org/10.1002/(sici)1097-4571(198809)39:5<293::aid-asi1>3.0.co;2-i

[62] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017.
Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*. 575–584.

[63] L. Fell, S. Dehdashti, P.D. Bruza, and C. Moreira. 2019.
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'19)*.

[64] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. 2011.
The Feynman Lectures on Physics, Vol. III: The New Millennium Edition: Quantum Mechanics (Feynman Lectures on Physics (Paperback)) (Volume 3).

Basic Books.
https://www.amazon.com/Feynman-Lectures-Physics-Vol-III/dp/0465025013?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0465025013

[65] Ingo Frommholz, Birger Larsen, Benjamin Piwowarski, Mounia Lalmas, Peter Ingwersen, and Keith van Rijsbergen. 2010.
Supporting Polyrepresentation in a Quantum-inspired Geometrical Retrieval Framework. In *Proceedings of the Third Symposium on Information Interaction in Context (IIiX '10)*. ACM, New York, NY, USA, 115–124.
https://doi.org/10.1145/1840784.1840802

[66] Ingo Frommholz, Benjamin Piwowarski, Mounia Lalmas, and Keith van Rijsbergen. 2011.
Processing Queries in Session in a Quantum-Inspired IR Framework.
In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 751–754.
https://doi.org/10.1007/978-3-642-20161-5_86

[67] Norbert Fuhr. 2008.
A Probability Ranking Principle for Interactive Information Retrieval.
Inf. Retr. 11, 3 (June 2008), 251–265.
https://doi.org/10.1007/s10791-008-9045-0

[68] Norbert Fuhr, Anastasia Giachanou, Gregory Grefenstette, Iryna Gurevych, Andreas Hanselowski, Kalervo Jarvelin, Rosie Jones, YiquN Liu, Josiane Mothe, Wolfgang Nejdl, Isabella Peters, and Benno Stein. 2018.
An Information Nutritional Label for Online Documents.
SIGIR Forum 51, 3 (Feb. 2018), 46–66.
https://doi.org/10.1145/3190580.3190588

[69] Liane Gabora and Diederik Aerts. 2002.
Contextualizing concepts using a mathematical generalization of the quantum formalism.
Journal of Experimental & Theoretical Artificial Intelligence 14, 4 (oct 2002), 327–358.
https://doi.org/10.1080/09528130210162253

[70] Ernesto F. Galvão. 2005.

Discrete Wigner functions and quantum computational speedup.
Physical Review A 71, 4 (April 2005).
`https://doi.org/10.1103/physreva.71.042302`

[71] N. Gisin. 1999.
Bell inequality for arbitrary many settings of the analyzers.
Physics Letters A 260, 1-2 (sep 1999), 1–3.
`https://doi.org/10.1016/s0375-9601(99)00428-4`

[72] J. Greene and M. D'Oliveira. 1999.
Learning to Use Statistical Tests in Psychology.
Open University Press.
`https://books.google.co.uk/books?id=wMl-AAAAMAAJ`

[73] Robert B. Griffiths. 2001.
Consistent Quantum Theory.
Cambridge University Press.
`https://doi.org/10.1017/cbo9780511606052`

[74] J.A. Hampton. 1997.
Conceptual combination.
Knowledge, concepts, and categories (1997), 133–159.

[75] Martie G Haselton, Daniel Nettle, and Damian R Murray. 2015.
The evolution of cognitive bias.
The handbook of evolutionary psychology (2015), 1–20.

[76] Robin M Hogarth and Hillel J Einhorn. 1992.
Order effects in belief updating: The belief-adjustment model.
Cognitive Psychology 24, 1 (jan 1992), 1–55.
`https://doi.org/10.1016/0010-0285(92)90002-j`

[77] Yuexian Hou and Dawei Song. 2009.
Characterizing Pure High-Order Entanglements in Lexical Semantic Spaces via Information Geometry. In *QI*.

[78] Yuexian Hou, Xiaozhao Zhao, Dawei Song, and Wenjie Li. 2013.
Mining Pure High-order Word Associations via Information Geometry for Information Retrieval.

ACM Trans. Inf. Syst. 31, 3, Article 12 (Aug. 2013), 32 pages.
https://doi.org/10.1145/2493175.2493177

[79] J. Jiang and J. Allan. 2016.
Adaptive effort for search evaluation metrics.
In ECIR ' 16 (2016), 187–199.

[80] Jiepu Jiang, Daqing He, and James Allan. 2017.
Comparing In Situ and Multidimensional Relevance Judgments. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*. ACM Press.
https://doi.org/10.1145/3077136.3080840

[81] Xiaoran Jin, Marc Sloan, and Jun Wang. 2013.
Interactive Exploratory Search for Multi Page Search Results. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 655–666.
https://doi.org/10.1145/2488388.2488446

[82] KAREN SPARCK JONES. 1972.
A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL.
Journal of Documentation 28, 1 (1972), 11–21.
https://doi.org/10.1108/eb026526 arXiv:https://doi.org/10.1108/eb026526

[83] Daniel Kahneman, Paul Slovic, and Amos Tversky (Eds.). 1982.
Judgment under uncertainty.
Cambridge University Press.
https://doi.org/10.1017/cbo9780511809477

[84] Andrei Khrennikov. 2019.
Basics of Quantum Theory for Quantum-Like Modeling Information Retrieval.
In *Quantum-Like Models for Information Retrieval and Decision-Making*. Springer, 51–82.

[85] Simon Kochen and E. Specker. 1967.
The Problem of Hidden Variables in Quantum Mechanics.
Indiana University Mathematics Journal 17, 1 (1967), 59–87.
https://doi.org/10.1512/iumj.1968.17.17004

[86] A. N. Kolmogorov. 2013.
Foundations of the Theory of Probability.
Martino Fine Books.

[87] A. N. Kolmogorov. 2013.
Foundations of the Theory of Probability.
Martino Fine Books.
https://www.amazon.com/Foundations-Theory-Probability-N-Kolmogorov/dp/1614275149?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1614275149

[88] Carol L. Krumhansl. 1978.
Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density.
Psychological Review 85, 5 (1978), 445–463.
https://doi.org/10.1037/0033-295x.85.5.445

[89] Carol L. Krumhansl. 1988.
Testing the density hypothesis: Comment on Corter.
Journal of Experimental Psychology: General 117, 1 (1988), 101–104.
https://doi.org/10.1037/0096-3445.117.1.101

[90] Janne V. Kujala and Ehtibar N. Dzhafarov. 2016.
Proof of a Conjecture on Contextuality in Cyclic Systems with Binary Variables.
Foundations of Physics 46, 3 (01 Mar 2016), 282–299.
https://doi.org/10.1007/s10701-015-9964-8

[91] Hang Li and Zhengdong Lu. 2016.
Deep Learning for Information Retrieval. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '16)*. ACM, New York, NY, USA, 1203–1206.
https://doi.org/10.1145/2911451.2914800

[92] Jingfei Li, Peng Zhang, Dawei Song, and Yuexian Hou. 2016.
An adaptive contextual quantum language model.
Physica A: Statistical Mechanics and its Applications 456 (aug 2016), 51–67.
https://doi.org/10.1016/j.physa.2016.03.003

[93] Jingfei Li, Peng Zhang, Dawei Song, and Yue Wu. 2017.
Understanding an enriched multidimensional user relevance model by analyzing query logs.
Journal of the Association for Information Science and Technology 68, 12 (2017), 2743–2754.
https://doi.org/10.1002/asi.23868

[94] Qiuchi Li, Sagar Uprety, Benyou Wang, and Dawei Song. 2018.
Quantum-inspired Complex Word Embedding.
arXiv:arXiv:1805.11351

[95] H. P. Luhn. 1957.
A Statistical Approach to Mechanized Encoding and Searching of Literary Information.
IBM J. Res. Dev. 1, 4 (Oct. 1957), 309–317.
https://doi.org/10.1147/rd.14.0309

[96] H. P. Luhn. 1958.
The Automatic Creation of Literature Abstracts.
IBM J. Res. Dev. 2, 2 (April 1958), 159–165.
https://doi.org/10.1147/rd.22.0159

[97] Kevin Lund and Curt Burgess. 1996.
Producing high-dimensional semantic spaces from lexical co-occurrence.
Behavior Research Methods, Instruments, & Computers 28, 2 (jun 1996), 203–208.
https://doi.org/10.3758/bf03204766

[98] Jiyun Luo, Sicong Zhang, Xuchu Dong, and Hui Yang. 2015.
Designing States, Actions, and Rewards for Using POMDP in Session Search.
In *Lecture Notes in Computer Science*. Springer International Publishing, 526–537.
https://doi.org/10.1007/978-3-319-16354-3_58

[99] Yuanhua Lv and ChengXiang Zhai. 2011.
Lower-bounding Term Frequency Normalization. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11)*. ACM, New York, NY, USA, 7–16.
https://doi.org/10.1145/2063576.2063584

[100] A I Lvovsky. 2004.

Iterative maximum-likelihood reconstruction in quantum homodyne tomography.
Journal of Optics B: Quantum and Semiclassical Optics 6, 6 (2004), S556.
http://stacks.iop.org/1464-4266/6/i=6/a=014

[101] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008.
Introduction to Information Retrieval.
Cambridge University Press, New York, NY, USA.

[102] Massimo Melucci. 2005.
Can Vector Space Bases Model Context?

[103] Massimo Melucci. 2005.
Context Modeling and Discovery Using Vector Space Bases. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05)*. ACM, New York, NY, USA, 808–815.
https://doi.org/10.1145/1099554.1099745

[104] Massimo Melucci. 2008.
A basis for information retrieval in context.
ACM Transactions on Information Systems 26, 3 (jun 2008), 1–41.
https://doi.org/10.1145/1361684.1361687

[105] Massimo Melucci. 2010.
An Investigation of Quantum Interference in Information Retrieval.
In *Advances in Multidisciplinary Retrieval*. Springer Berlin Heidelberg, 136–151.
https://doi.org/10.1007/978-3-642-13084-7_11

[106] Massimo Melucci and Ryen W. White. 2007.
Utilizing a geometry of context for enhanced implicit feedback. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07*. ACM Press.
https://doi.org/10.1145/1321440.1321480

[107] David W Moore. 2002.
Measuring new types of question-order effects: Additive and subtractive.
The Public Opinion Quarterly 66, 1 (2002), 80–91.

[108] Michael A. Nielsen and Isaac L. Chuang. 2011.

Quantum Computation and Quantum Information: 10th Anniversary Edition (10th ed.).
Cambridge University Press, New York, NY, USA.

[109] Michael S. Nilan and Patricia T. Fletcher. 1987.
INFORMATION BEHAVIORS IN THE PREPARATION OF RESEARCH PROPOSALS: A USER STUDY.. In *Proceedings of the ASIS Annual Meeting*, Ching chih Chen (Ed.), Vol. 24. Learned Information Inc, 186–192.

[110] M. S. Nilan, R. P. Peek, and H. W. Snyder. 1988.
A Methodology for Tapping User Evaluation Behaviors: An Exploration of Users'Strategy, Source and Information Evaluating.
In *Proceedings of the American Society for Information Science (ASIS) 51st Annual Meeting*, C. L. Borgman and E. Y. H. Pai (Eds.). Learned Information, Medford, NJ, 152–159.

[111] Robert M. Nosofsky. 1984.
Choice, similarity, and the context theory of classification.
Journal of Experimental Psychology: Learning, Memory, and Cognition 10, 1 (1984), 104–114.
`https://doi.org/10.1037/0278-7393.10.1.104`

[112] Robert M Nosofsky. 1991.
Stimulus bias, asymmetric similarity, and classification.
Cognitive Psychology 23, 1 (jan 1991), 94–140.
`https://doi.org/10.1016/0010-0285(91)90004-8`

[113] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer. 2013.
Web credibility: Features exploration and credibility prediction.
In ECIR ' 13 (2013), 557–568.

[114] Kezban Dilek Onal, Ye Zhang, Ismail Sengor Altingovde, Md Mustafizur Rahman, Pinar Karagoz, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, Maarten de Rijke, and Matthew Lease. 2017.
Neural information retrieval: at the end of the early years.
Information Retrieval Journal 21, 2-3 (nov 2017), 111–182.

https://doi.org/10.1007/s10791-017-9321-y

[115] Joao Palotti, Lorraine Goeuriot, Guido Zuccon, and Allan Hanbury. 2016. 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 '16)*. ACM, New York, NY, USA, 965–968. https://doi.org/10.1145/2911451.2914741

[116] Joao Palotti, Guido Zuccon, and Allan Hanbury. 2018. MM: A New Framework for Multidimensional Evaluation of Search Engines. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. ACM, New York, NY, USA, 1699–1702. https://doi.org/10.1145/3269206.3269261

[117] Taemin Kim Park. 1993. The nature of relevance in information retrieval: An empirical study. *The library quarterly* 63, 3 (1993), 318–351.

[118] Francis Jeffry Pelletier. 1994. The Principle of Semantic Compositionality. *Topoi* 13, 1 (mar 1994), 11–24. https://doi.org/10.1007/bf00763644

[119] Benjamin Piwowarski, Ingo Frommholz, Mounia Lalmas, and Keith van Rijsbergen. 2010. What Can Quantum Theory Bring to Information Retrieval. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*. ACM, New York, NY, USA, 59–68. https://doi.org/10.1145/1871437.1871450

[120] Benjamin Piwowarski, Ingo Frommholz, Yashar Moshfeghi, Mounia Lalmas, and Keith van Rijsbergen. 2010. Filtering Documents with Subspaces. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 615–618. https://doi.org/10.1007/978-3-642-12275-0_60

[121] Benjamin Piwowarski and Mounia Lalmas. 2009. A Quantum-Based Model for Interactive Information Retrieval. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 224–231.

https://doi.org/10.1007/978-3-642-04417-5_20

[122] Benjamin Piwowarski and Mounia Lalmas. 2009.
Structured Information Retrieval and Quantum Theory. In *Proceedings of the 3rd International Symposium on Quantum Interaction (QI '09)*. Springer-Verlag, Berlin, Heidelberg, 289–298.
https://doi.org/10.1007/978-3-642-00834-4_24

[123] Benjamin Piwowarski, Mounia Lalmas, Ingo Frommholz, and Keith van Rijsbergen. 2010.
Exploring a Multidimensional Representation of Documents and Queries. In *Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO '10)*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, France, France, 57–60.
http://dl.acm.org/citation.cfm?id=1937055.1937067

[124] Benjamin Piwowarski, Andrew Trotman, and Mounia Lalmas. 2008.
Sound and Complete Relevance Assessment for XML Retrieval.
ACM Trans. Inf. Syst. 27, 1, Article 1 (Dec. 2008), 37 pages.
https://doi.org/10.1145/1416950.1416951

[125] Jay M. Ponte and W. Bruce Croft. 1998.
A language modeling approach to information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '98*. ACM Press.
https://doi.org/10.1145/290941.291008

[126] E. M. Pothos and J. R. Busemeyer. 2009.
A quantum probability explanation for violations of 'rational' decision theory.
Proceedings of the Royal Society B: Biological Sciences 276, 1665 (mar 2009), 2171–2178.
https://doi.org/10.1098/rspb.2009.0121

[127] Emmanuel M. Pothos and Jerome R. Busemeyer. 2011.
A Quantum Probability Explanation for Violations of Symmetry in Similarity Judgments. In *CogSci*.

[128] Emmanuel M. Pothos, Jerome R. Busemeyer, and Jennifer S. Trueblood. 2013.
A quantum geometric model of similarity.

Psychological Review 120, 3 (2013), 679–696.
https://doi.org/10.1037/a0033142

[129] Emmanuel M Pothos, Gavin Perry, Philip J Corr, Mervin R Matthew, and Jerome R Busemeyer. 2011.
Understanding cooperation in the Prisoner's Dilemma game.
Personality and Individual Differences 51, 3 (2011), 210–215.

[130] Yonggang Qiu and Hans-Peter Frei. 1993.
Concept Based Query Expansion. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '93)*. ACM, New York, NY, USA, 160–169.
https://doi.org/10.1145/160688.160713

[131] C. J. van Rijsbergen. 2004.
The Geometry of Information Retrieval.
Cambridge University Press, New York, NY, USA.

[132] C S. Robertson, H. Zaragoza, Stephen Robertson, and Hugo Zaragoza. [n.d.].
The Probabilistic Relevance Framework: BM25 and Beyond.

[133] S.E. ROBERTSON. 1977.
THE PROBABILITY RANKING PRINCIPLE IN IR.
Journal of Documentation 33, 4 (1977), 294–304.
https://doi.org/10.1108/eb026647 arXiv:https://doi.org/10.1108/eb026647

[134] Stephen Robertson and Hugo Zaragoza. 2009.
The Probabilistic Relevance Framework: BM25 and Beyond.
Found. Trends Inf. Retr. 3, 4 (April 2009), 333–389.
https://doi.org/10.1561/1500000019

[135] J. J. Sakurai and Jim Napolitano. 2017.
Modern Quantum Mechanics.
Cambridge University Press.

[136] G. Salton. 1971.
The SMART Retrieval System—Experiments in Automatic Document Processing.
Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[137] Gerard Salton and Chris Buckley. 1988.
Term-Weighting Approaches in Automatic Text Retrieval.
Inf. Process. Manage. 24 (1988), 513–523.

[138] Tefko Saracevic. 1997.
The stratified model of information retrieval interaction: Extension and applications.

[139] Tefko Saracevic. 2007.
Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance.
Journal of the American Society for Information Science and Technology 58, 13 (2007), 1915–1933.
https://doi.org/10.1002/asi.20682

[140] Tefko Saracevic. 2016.
The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really?
Synthesis Lectures on Information Concepts, Retrieval, and Services 8, 3 (Sept. 2016), i–109.
https://doi.org/10.2200/s00723ed1v01y201607icr050

[141] Andreas T Schmidt and Bart Engelen. 2020.
The ethics of nudging: An overview.
Philosophy Compass 15, 4 (2020), e12658.

[142] Hinrich Schütze. 1998.
Automatic Word Sense Discrimination.
Comput. Linguist. 24, 1 (March 1998), 97–123.
http://dl.acm.org/citation.cfm?id=972719.972724

[143] J. Schwarz and M. Morris. 2011.
Augmenting web pages and search results to support credibility assessment.
In CHI ' 11 (2011), 1245–1254.

[144] Ashley Sides, Daniel Osherson, Nicolao Bonini, and Riccardo Viale. 2002.
On the reality of the conjunction fallacy.
Memory & Cognition 30, 2 (mar 2002), 191–198.
https://doi.org/10.3758/bf03195280

[145] Marc Sloan and Jun Wang. 2015.
Dynamic Information Retrieval. In *Proceedings of the 2015 International Conference on Theory of Information Retrieval - ICTIR '15*. ACM Press.
https://doi.org/10.1145/2808194.2809457

[146] Paul Slovic. 1995.
The construction of preference.
American Psychologist 50, 5 (1995), 364–371.
https://doi.org/10.1037/0003-066x.50.5.364

[147] Fei Song and W. Bruce Croft. 1999.
A general language model for information retrieval. In *Proceedings of the eighth international conference on Information and knowledge management - CIKM '99*. ACM Press.
https://doi.org/10.1145/319950.320022

[148] Alessandro Sordoni, Jing He, and Jian-Yun Nie. 2013.
Modeling Latent Topic Interactions Using Quantum Interference for Information Retrieval. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management (CIKM '13)*. ACM, New York, NY, USA, 1197–1200.
https://doi.org/10.1145/2505515.2507854

[149] Alessandro Sordoni, Jian-Yun Nie, and Yoshua Bengio. 2013.
Modeling Term Dependencies with Quantum Language Models for IR. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, New York, NY, USA, 653–662.
https://doi.org/10.1145/2484028.2484098

[150] R. W. Spekkens. 2005.
Contextuality for preparations, transformations, and unsharp measurements.
Physical Review A 71, 5 (may 2005).
https://doi.org/10.1103/physreva.71.052108

[151] James T Townsend, Kam M Silva, Jesse Spencer-Smith, and Michael J Wenger. 2000.
Exploring the relations between categorization and decision making with regard to realistic face stimuli.

Pragmatics & cognition 8, 1 (2000), 83–105.

[152] Jennifer S. Trueblood and Jerome R. Busemeyer. 2011.
A Quantum Probability Account of Order Effects in Inference.
Cognitive Science 35, 8 (sep 2011), 1518–1552.
https://doi.org/10.1111/j.1551-6709.2011.01197.x

[153] Jennifer S. Trueblood and Jerome R. Busemeyer. 2011.
A Quantum Probability Account of Order Effects in Inference.
Cognitive Science 35, 8 (sep 2011), 1518–1552.
https://doi.org/10.1111/j.1551-6709.2011.01197.x

[154] Amos Tversky. 1977.
Features of similarity.
Psychological Review 84, 4 (1977), 327–352.
https://doi.org/10.1037/0033-295x.84.4.327

[155] A. Tversky and D. Kahneman. 1974.
Judgment under Uncertainty: Heuristics and Biases.
Science 185, 4157 (Sept. 1974), 1124–1131.
https://doi.org/10.1126/science.185.4157.1124

[156] Amos Tversky and Daniel Kahneman. 1983.
Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.
Psychological Review 90, 4 (1983), 293–315.
https://doi.org/10.1037/0033-295x.90.4.293

[157] Amos Tversky and Daniel Kahneman. 1983.
Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.
Psychological Review 90, 4 (1983), 293–315.
https://doi.org/10.1037/0033-295x.90.4.293

[158] Amos Tversky and Eldar Shafir. 1992.
The Disjunction Effect in Choice under Uncertainty.
Psychological Science 3, 5 (sep 1992), 305–310.
https://doi.org/10.1111/j.1467-9280.1992.tb00678.x

[159] Amos Tversky and Richard H Thaler. 1990.
Anomalies: preference reversals.
Journal of Economic Perspectives 4, 2 (1990), 201–211.

[160] Hisaharu Umegaki. 1962.
Conditional expectation in an operator algebra. IV. Entropy and information.
Kodai Math. Sem. Rep. 14, 2 (1962), 59–85.
`https://doi.org/10.2996/kmj/1138844604`

[161] Sagar Uprety, Shahram Dehdashti, Lauren Fell, Peter Bruza, and Dawei Song. 2019.
Modelling Dynamic Interactions Between Relevance Dimensions. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval - ICTIR '19*. ACM Press.
`https://doi.org/10.1145/3341981.3344233`

[162] Sagar Uprety, Dimitris Gkoumas, and Dawei Song. 2018.
Investigating Bell Inequalities for Multidimensional Relevance Judgments in Information Retrieval.
arXiv:arXiv:1811.06645

[163] Sagar Uprety, Dimitris Gkoumas, and Dawei Song. 2019.
Investigating Bell Inequalities for Multidimensional Relevance Judgments in Information Retrieval. In *Quantum Interaction*. Springer International Publishing, Cham, 177–188.

[164] Sagar Uprety and Dawei Song. 2018.
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 '18)*. ACM, New York, NY, USA, 191–194.
`https://doi.org/10.1145/3234944.3234972`

[165] Sagar Uprety, Yi Su, Dawei Song, and Jingfei Li. 2018.
Modeling Multidimensional User Relevance in IR Using Vector Spaces. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 993–996.
`https://doi.org/10.1145/3209978.3210130`

[166] Sagar Uprety, Prayag Tiwari, Shahram Dehdashti, Lauren Fell, Dawei Song, Peter Bruza, and Massimo Melucci. 2020.
Quantum-Like Structure in Multidimensional Relevance Judgements. In *Advances in Information Retrieval*. Springer International Publishing, Cham, 728–742.

[167] M. Verma, E. Yilmaz, and N. Craswell. 2016.
On obtaining effort based judgements for information retrieval.
In WSDM ' 16 (2016), 277–286.

[168] John von Neumann. 1955.
Mathematical Foundations of Quantum Mechanics.
Princeton University Press.
https://www.amazon.com/Mathematical-Foundations-Quantum-Mechanics-Neumann/dp/0691028931?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0691028931

[169] Apostol Vourdas. 2019.
Probabilistic inequalities and measurements in bipartite systems.
Journal of Physics A: Mathematical and Theoretical (2019).

[170] Benyou Wang, Qiuchi Li, Massimo Melucci, and Dawei Song. 2019.
Semantic Hilbert space for text representation learning. In *The World Wide Web Conference*. 3293–3299.

[171] Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. 2016.
Exploration of Quantum Interference in Document Relevance Judgement Discrepancy.
Entropy 18, 12 (Apr 2016), 144.
https://doi.org/10.3390/e18040144

[172] Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. 2016.
Exploration of Quantum Interference in Document Relevance Judgement Discrepancy.
Entropy 18, 12 (Apr 2016), 144.
https://doi.org/10.3390/e18040144

[173] Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. 2016.
Exploration of Quantum Interference in Document Relevance Judgement Discrepancy.
Entropy 18, 12 (Apr 2016), 144.
https://doi.org/10.3390/e18040144

[174] Kuansan Wang, Nikolas Gloy, and Xiaolong Li. 2010.
Inferring Search Behaviors Using Partially Observable Markov (POM) Model. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)*. ACM, New York, NY, USA, 211–220.
https://doi.org/10.1145/1718487.1718514

[175] Zheng Wang and Jerome R. Busemeyer. 2013.
A Quantum Question Order Model Supported by Empirical Tests of anA Prioriand Precise Prediction.
Topics in Cognitive Science (sep 2013), n/a–n/a.
https://doi.org/10.1111/tops.12040

[176] Zheng Wang and Jerome R. Busemeyer. 2016.
Interference effects of categorization on decision making.
Cognition 150 (2016), 133 – 149.
https://doi.org/10.1016/j.cognition.2016.01.019

[177] A. Wawer, R. Nielek, and A. Wierzbicki. 2014.
Predicting webpage credibility using linguistic features.
In WWW ' 14 (2014), 1135–1140.

[178] Dominic Widdows. 2003.
Orthogonal Negation in Vector Spaces for Modelling Word-Meanings and Document Retrieval. In *ACL*.

[179] Dominic Widdows. 2004.
Geometry and Meaning. In *CSLI lecture notes series*.

[180] Peter Wittek, Bevan Koopman, Guido Zuccon, and Sándor Darányi. 2013.
Combining word semantics within complex Hilbert space for information retrieval. In *International Symposium on Quantum Interaction*. Springer, 160–171.

[181] William K Wootters. 1987.
A Wigner-function formulation of finite-state quantum mechanics.
Annals of Physics 176, 1 (May 1987), 1–21.
https://doi.org/10.1016/0003-4916(87)90176-x

[182] Yunjie (Calvin) Xu and Zhiwei Chen. 2006.
Relevance judgment: What do information users consider beyond topicality?
Journal of the American Society for Information Science and Technology 57, 7 (2006), 961–973.
https://doi.org/10.1002/asi.20361

[183] Y. Yamamoto and K. Tanaka. 2011.
Enhancing credibility judgment of web search results.
In CHI ' 11 (2011), 1235–1244.

[184] James M Yearsley and Jonathan J Halliwell. 2019.
Contextuality in Human Decision Making in the Presence of Direct Influences: A Comment on Basieva et al. (2019).
arXiv:q-bio.NC/1905.12570

[185] E. Yilmaz, M. Verma, N. Craswell, F. Radlinski, and P. Bailey. 2014.
Relevance and effort: An analysis of document utility.
In CIKM ' 14 (2014), 91–100.

[186] Cheng Xiang Zhai, William W Cohen, and John Lafferty. 2003.
Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 10–17.

[187] Peng Zhang, Jingfei Li, Benyou Wang, Xiaozhao Zhao, Dawei Song, Yuexian Hou, and Massimo Melucci. 2016.
A Quantum Query Expansion Approach for Session Search.
Entropy 18 (2016), 146.

[188] Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. 2018.
End-to-End Quantum-like Language Models with Application to Question Answering. In *AAAI*.

[189] Peng Zhang, Dawei Song, Yuexian Hou, Jun Wang, and Peter Bruza. 2010.

Explanation of relevance judgement discrepancy with quantum interference. In *AAAI-Fall 2010 Symposium on Quantum Informatics for Cognitive, Social, and Semantic Processes (QI2010)*.
http://oro.open.ac.uk/35126/

[190] Peng Zhang, Dawei Song, Xiaozhao Zhao, and Yuexian Hou. 2011.
Investigating Query-Drift Problem from a Novel Perspective of Photon Polarization. In *ICTIR*.

[191] Yinglong Zhang, Jin Zhang, Matthew Lease, and Jacek Gwizdka. 2014.
Multidimensional Relevance Modeling via Psychometrics and Crowdsourcing. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14)*. ACM, New York, NY, USA, 435–444.
https://doi.org/10.1145/2600428.2609577

[192] Xiaozhao Zhao, Peng Zhang, Dawei Song, and Yuexian Hou. 2011.
A Novel Re-ranking Approach Inspired by Quantum Measurement.
In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 721–724.
https://doi.org/10.1007/978-3-642-20161-5_79

[193] Liron Zighelnic and Oren Kurland. 2008.
Query-drift Prevention for Robust Query Expansion. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08)*. ACM, New York, NY, USA, 825–826.
https://doi.org/10.1145/1390334.1390524

[194] Steven Zimmerman, Alistair Thorpe, Chris Fox, and Udo Kruschwitz. 2019.
Privacy Nudging in Search: Investigating Potential Impacts. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 283–287.

[195] Guido Zuccon. 2016.
Understandability Biased Evaluation for Information Retrieval. In *Advances in Information Retrieval*, Nicola Ferro, Fabio Crestani, Marie-Francine Moens, Josiane Mothe, Fabrizio Silvestri, Giorgio Maria Di Nunzio, Claudia Hauff, and Gianmaria Silvello (Eds.). Springer International Publishing, Cham, 280–292.

[196] Guido Zuccon. 2016.

Understandability biased evaluation for information retrieval. In *European Conference on Information Retrieval*. Springer, 280–292.

[197] Guido Zuccon and Leif Azzopardi. 2010.
Using the Quantum Probability Ranking Principle to Rank Interdependent Documents. In *Proceedings of the 32Nd European Conference on Advances in Information Retrieval (ECIR'2010)*. Springer-Verlag, Berlin, Heidelberg, 357–369.
https://doi.org/10.1007/978-3-642-12275-0_32

[198] Guido Zuccon, Leif A. Azzopardi, and Keith van Rijsbergen. 2009.
The Quantum Probability Ranking Principle for Information Retrieval.
In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 232–240.
https://doi.org/10.1007/978-3-642-04417-5_21

[199] Guido Zuccon, Benjamin Piwowarski, and Leif Azzopardi. 2011.
On the use of complex numbers in quantum models for information retrieval. In *Conference on the Theory of Information Retrieval*. Springer, 346–350.

Appendix A

Consider a state vector in two different basis of a two dimensional Hilbert space,

$$\begin{aligned} |\psi\rangle &= a\,|A\rangle + b\,|B\rangle \\ &= c\,|C\rangle + d\,|D\rangle \end{aligned} \tag{8.1}$$

We want to represent the vectors of one basis in terms of the other. To do that, consider the vector orthogonal to $|\psi\rangle$, which is

$$\begin{aligned} |\tilde{\psi}\rangle &= b\,|A\rangle - a\,|B\rangle \\ &= d\,|C\rangle - c\,|D\rangle \end{aligned} \tag{8.2}$$

Using the above representations, we get

$$\begin{aligned} |C\rangle &= c\,|\psi\rangle + d\,|\tilde{\psi}\rangle \\ |D\rangle &= d\,|\psi\rangle - c\,|\tilde{\psi}\rangle \end{aligned} \tag{8.3}$$

Substituting $|\psi\rangle = a\,|A\rangle + b\,|B\rangle$ and $|\tilde{\psi}\rangle = b\,|A\rangle - a\,|B\rangle$ in 8.3, we get:

$$\begin{aligned} |C\rangle &= (ac+bd)\,|A\rangle + (bc-ad)\,|B\rangle \\ |D\rangle &= (ad-bc)\,|A\rangle + (ac+bd)\,|B\rangle \end{aligned} \tag{8.4}$$

We can thus represent one basis of a Hilbert Space in terms of another basis.

Appendix B

This appendix contains details of the crowd sourced user studies reported in this thesis, including the ethical approval process - applications to the Open University Human Research Ethics Committee, consent forms and information sheet, questionnaire design and materials.

Ethical Approval Process

The general process for ethical approval in the Open University requires one to submit an initial application to the Human Research Ethics Committee (HREC) with brief details about the study. This is used by HREC to determine whether the study needs an approval or not. Following this, a detailed application is submitted. The detailed application involves the following main documents to be submitted:

- **Application Form:** This form introduces the experiment and brief theory behind it with an abstract and literature review. Besides, one is required to furnish details about the experiment methodology, participant descriptions, recruitment procedures, participant compensation, ethical issues, and data protection and management.

- **Participant Information Sheet:** This document contains information about contact details of the investigators, description of the study, procedure to take part or withdraw and ethical and data security concerns.

- **Consent Form:** This document is a consent form as shown to participants at the beginning of the study and contains the terms of the study. Should the participants disagree with any of the terms, they are informed of their right to withdraw from the study. Note that the participants always have the right to withdraw from the study even after the study has been completed by them.

- **Study/Survey/Questionnaire Design** This document gives a rough overview of the actual survey which will be shown to participants including the visuals of the interface.

Survey Design and Crowdsourcing

Here, I discuss the tools used for conducting my crowdsourced user studies. The main tasks were survey design and then distributing it to the participants.

Survey Design

I used the Qualtrics platform licensed by the Open University (`https://openss.eu.qualtrics.com/`) for designing my surveys. Qualtrics provides an interface for building complex surveys and questionnaire along with javascript customisation. The surveys can be distributed using a url. Once participants click on the url and start the survey, the data is logged in the Qualtrics platform and can be downloaded as a CSV file at the end of the survey.

Crowdsourcing Platform

I used the crowdsourcing platform Prolific (`www.prolific.co`). Prolific is generally used by researchers in Psychological sciences and thus it seemed to be well-suited to IR user behaviour studies. I also considered other well-known platforms like Amazon Mechanical Turk (`https://www.mturk.com/`) and Appen (`https://appen.com/` - formerly known as Figure-Eight and Crowdflower). As of 3rd August, 2020, Prolific has more than 115,000 participants, majority from The United States and United Kingdom. It provides various types of filters for pre-screening participants, e.g. based on demographics, personal interests, education and employment, etc. In all my studies I have used the filter for approval rate, which is the fraction of studies approved of a participant. Participants in prolific are compensated for all approved studies with a minimum compensation of £5.00 per hour. As an investigator, one starts a new study by giving brief description of the study, selecting the expected completion time, number of participants and hourly compensation rate. Then a url for the study is required, which for me is the Qualtrics url. One cannot begin a study unless the total compensation to be paid to all the participants is paid to Prolific via online transfer.

Each participant has a unique prolific id which is captured at the beginning of the study. Prolific also records other participant variables such as their location, demographics, start and end times, device type, etc. and all of this data is sent to the server hosting the survey (here, Qualtrics). When the required number of participants have finished the survey, the investigator can check their submissions and approve or disapprove them. Participants can contact the investigators via their email or a in-built messaging service in case they experienced some issues during the survey, or want to contest the rejection of their submission, or want to withdraw their submission later on.

Key Ethical Concerns

Ethical concerns for my studies were primarily of three types:

- **Participant Information Sheet:** Since all my user studies were conducted on the crowdsourcing platform, an important thing to consider was how to distribute the participant information sheets. In a lab based study with participants physically present, investigators often ask the participants to keep the information sheet with themselves. This sheet contains the contact information of the investigators and and other important details which the participants might need in the future to either withdraw from the study or know the results of the study. For my crowdsourced studies, since the participants undertook the study in the computers or mobile phones, the participant information sheet was in the form of a webpage and it was the first page which the participants saw when they started the study. I asked them to download and save the webpage or take screenshot of the screen, in case they are unable to download the webpage.

- **Consent Form:** It is also important to record the informed consent of the participants. The consent form for my study was designed as a webpage which was shown immediately after the Participant Information Sheet webpage. The consent form webpage displayed the terms and conditions of the study and the participants were required to click 'Yes' or 'No' buttons accordingly whether they agree or disagree with each statement. For later studies I modified this webpage so that the participants first read all the terms and conditions and then selected whether they agreed with all of them and wanted to continue to the study or they did not agree with some of them and wanted to withdraw.

- **Data privacy:** This is a major ethical concern in experiments on human subjects. For my crowdsourced studies using Prolific, the primary identifier of each participant was the Prolific Id which was an alphanumeric string of 24 characters. Prolific also captured other information about the participants such as their age, gender, nationality, country of birth, employment status, native language, location coordinates, type of device used. However, I disregard all variables other than the Prolific Id as they are not useful for my experiments. Thus, the only information stored about a participant is their prolific id and their responses to questions which are generally 'Yes' or 'No' type options. Thus, the participants can never be identified. All data collected by prolific is automatically sent to the Qualtrics platform where I use the account based on the Open University License. From there, it is downloaded to OneDrive cloud platform licensed by the Open University.

Attachment of Various Forms Used in the Ethical Approval Process

The consent form was same for all the user studies and the participant information sheets for each study differed in only a few things related to the study - number of questions, compensation rate, etc. My user studies can be divided into two categories - (1) **contextuality**, where I investigated contextuality via Contextuality-by-Default inequality and (2) **multidimensional relevance** studies which were focused on multidimensional relevance. The various studies within the second category had a very similar design and differed only in the type or combinations of questions asked. Hence the application form for ethical approval of second category of studies are similar as they were mostly extensions of the first study in that category. Therefore, I attach two application forms, one for the **contextuality** study and othr for **multidimensional relevance** category of user studies.

Participant Information Sheet

Information about the Study

Study title: Investigating contextuality in relevance judgments of documents
Principle Investigator: Sagar Uprety, The Open University, UK
Contact: sagar.uprety@open.ac.uk
Alternate Contact: Professor Dawei Song, The Open University, UK
Contact: dawei.song@open.ac.uk

As a user of the crowdsourcing platform Prolific, you are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully

General information about the research study and collected research data

- This study is about investigating contextuality in user behaviour in judgment of documents of a search engine. It is run by researchers from The Open University (OU), United Kingdom. It is funded by the European Commission.
- This study has been approved by the OU Human Research Ethics Committee with reference number HREC/xxxx/Uprety.
- The study will involve answering simple questions like whether a document is relevant to a query.
- There will be **4 questions** and each question will take about 3 minutes to answer. (total around 5 minutes for the study, including time for reading this information and giving consent)
- You will be paid according to the rate of £7.08/hr (£0.82 for the study)

What will I be asked to do if I agree to take part?

- In each question you will be shown a query and one document snippet. Following which there will be 3 questions asked pertaining to relevance of the document to the query where you will be asked to give your answer corresponding to a 4 point scale (choose one of - definitely yes, probably yes, probably no, definitely no)

- It is up to you to decide whether or not to take part. If you do decide to take part you can download this information sheet (Right click and save on desktop/go to browser settings and save on mobile) asked to consent to some questions on the next page. Even If you decide to take part you are still free to withdraw at any time during completing the study without giving a reason by clicking the '**Withdraw**' button on the screen.

- In case you are unable to download this page, kindly take a screenshot.

- This study would help in furthering the understanding of the researchers as to how users of search engines make judgments of document relevance while considering difference contexts.

Participant Information Sheet

How will the data I provide be used?

- The data will be stored on the servers of Qualtrics UK and on the servers of The Open University, UK located in the campus at Walton Hall, Milton Keynes.

- The information obtained from the participants will be anonymous, as there are no means to ascertain the identity of an individual based on the documents judged as relevant.

- No personal information of the participants would be collected.

- The anonymous answers obtained from the participants will be used for analysis and mathematical modelling and will be published in peer-reviewed scientific conferences and journals.

- The anonymous relevance judgments data will be archived for use in future analysis tasks.

The data will be collected on the Qualtrics survey platform which complies with the UK/EU data protection regulations. In case the data is shared with any collaborators, it will be conditioned on them demonstrating their compliance with the EU data protection laws.

Your right to withdraw from the study

- You have the right to withdraw from the study at any time during your answering of questions by choosing the '**Withdraw**' button **at the end of each question.**
- You have the right to ask for your data to be removed after your participation in the study by emailing at **sagar.uprety@open.ac.uk** or at **dawei.song@open.ac.uk** up until 10 days from the date you completed the study. Kindly make a note of these email ids if you are unable to download this information sheet.

Queries regarding the study
Should you have any questions before beginning the study, please direct them to the above email addresses (along with your prolific id) and you will get a response as soon as possible.

How do I agree to take part?
Consent to the seven questions asked in the next pages. By clicking on the '**I agree to proceed with the study**' button at the end of these questions, you will be taken to the questionnaire.

Thank you very much for taking time to read this information sheet.

Consent Form

The Open University

Consent Form

Informed Consent for Investigating contextuality in relevance judgments

Name, position, and department/faculty of researcher
Sagar Uprety, School of Computing and Communications, Faculty of STEM, The Open University, UK

Please tick the appropriate boxes	Yes	No
1. Taking part in the study		
I have read and understood the study information dated 14th February 2019 or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time up until publication of the results in peer-reviewed conferences or journals, without having to give a reason.	☐	☐
I understand that taking part in the study involves answering a questionnaire completed via web browser	☐	☐
2. Use of the information in the study		
I understand that information I provide will be used for publications in peer reviewed scientific conferences. and journals	☐	☐

Consent Form

I understand that there will be no personal information collected about me that can identify me, such as my name or where I live. ☐ ☐

I understand that my data will be stored in Open University campus servers for 10 years after collection. ☐ ☐

3. Future use and reuse of the information by others

I give permission for the relevance judgment answers that I provide to be deposited in a specialist data centre after it has been anonymised, so it can be used for future research and learning. ☐ ☐

NOTE:

- The data will be deposited in the form of a survey database and will be anonymised by default (see information sheet).
- It will not be for any commercial use.

I agree to proceed with the study | Cancel participation

This research project has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number: XXXX.

http://www.open.ac.uk/research/ethics/

Application Form for Contextuality

HUMAN RESEARCH ETHICS COMMITTEE (HREC)
APPLICATION FORM

All OU research involving human participants or materials requires assessment by the HREC.

Where you have determined your research requires a full review or you have completed the HREC Project Registration and Risk Checklist and been advised that will need to complete this form as part of the full review process, please complete and email this form to Research-REC-Review@open.ac.uk. Attach any related documents, for example: a consent form, information sheet, questionnaire, or publicity leaflet to ensure that the HREC Review Panel has everything they need to carry out a full review. If there are more than one group of participants, relevant documents for each research group need to be included.

If you have any queries about completing the proforma please check the Research Ethics website, in particular the FAQs - http://www.open.ac.uk/research/ethics/human-research/faqs which include sample documents and templates, or email Research-REC-review@open.ac.uk.

The deadline for applications is **every Thursday** by **5.30pm**. Applications are then sent to the HREC Review Panel with a minimum response time of 21 working days. However, the process can take a month or longer, so when planning your research and ethics application, you need to build in sufficient time for the HREC review to avoid any delays to your research. Particularly, when you are planning overseas travel or interviews with participants as it is essential that no potential participants are approached until your research has been fully assessed by the HREC.

Please complete all the sections below – deleting the instructions in italics.

Project identification and rationale

1. Title of project

Modelling User's Cognitive Dynamics in Information Access and Retrieval using Quantum Probability

---

2. Abstract

Contextuality is a phenomenon encountered in Quantum Theory wherein a measurement done under different contexts gives different results. Recently in the field of Quantum Cognition, there is growing speculation and some empirical evidence that this phenomenon is also present in human decision making.

We seek to further investigate contextuality in human information processing, particularly in decisions or judgments of relevance of documents to queries. Users will be asked to read a query and a related document/documents and will be asked a set of questions. The different set of questions and/or documents are analogous to different measurement contexts and we want to know if the answer to a question (relevance of a document) is influenced by what set of questions (documents) are asked (presented) alongside. The answers will be typically yes/no type (relevant/non-relevant) and a probability measure can be attributed to each answer by calculating the frequency of yes and no answers to a question over all the users. Then we plug in these probabilities into an inequality from Quantum Theory - the violation (non-violation) of which shows the presence of contextuality (non-contextuality) in the decision-making process.

Application Form for Contextuality

STRICTLY CONFIDENTIAL

Project personnel and collaborators

Application Form for Contextuality

STRICTLY CONFIDENTIAL

STRICTLY CONFIDENTIAL

3. Investigators

Give names and institutional attachments of all persons involved in the collection and handling of individual data and name one person as Principal Investigator (PI). Research students should name themselves as PI and include a supervisor's electronic signature and/or comments (below) as evidence of supervisor support. Without this the application cannot be processed.

Principal Investigator/ (or Research Student): Sagar Uprety, The Open University, UK

Other researcher(s): Dawei Song, The Open University, UK

For students only:
Please note that this application cannot be processed without your supervisor's signature and/or supporting comments which should be provided below (Research ethics review: guidance for research students and supervisors):

Postgraduate research degree: PhD

Personal identifier: G2581947

Supervisor (preferably primary): Dawei Song

Email: dawei.song@open.ac.uk

Supervisor's electronic signature:

Supervisor's supporting comments:

Approved

Research protocol

4. Schedule

Time frame for the research and its data collection phase(s):

From: 3rd December 2018 To: 7th December 2018

Earliest date participants will be contacted: 3rd December 2018

5. Methodology

5. Methodology

Outline the method(s) that will be employed to collect and analyse data. Any relevant

Application Form for Contextuality

documents, such as interview or survey questions or a participant information sheet, should be sent with the completed proforma. Where there are more than one group of participants, please provide separate consent forms and participant information sheets. If, for any reason, any of this is not possible please explain why.

Survey will be designed with relevant questions using the Qualtrics (https://www.qualtrics.com/uk/) survey management tool. This would include participant information and consent form as the first two pages of the survey. If the user does not consent to the study, he or she will be re-directed to exit the survey using skip logic built within Qualtrics.

An anonymous survey link generated will be distributed to users of an online crowdsourcing platform called Prolific (We have been recommended to use this platform by researchers working in the same area as it leads to better quality of data collection for decision-making related studies - https://prolific.ac) . Qualtrics will also be used for storing the responses collected. Also, the prolific id of each participant would be requested in the beginning of the survey, so that we can uniquely identify them if they want to pull out of the study later or want the responses given by them to be deleted.

Participants can ask for withdrawing from the study and deletion of their data within 10 days of their participating in the study. After 10 days, the data might have been used for doing numerical analysis and reported in a paper submitted to a conference.

6. Participants

Give details of the population targeted or from which you will be sampling and how this sampling will be done. Give information on the diversity of the sample.

The study will be published on the online crowdsourcing platform Prolific which has more than 25000 people registered who participate remotely in different user studies and surveys. They are from many different countries in the world, although more than 70% are from UK and US. They are predominantly between the ages of 20 and 40, and more than half of them are female. Participants who find the task of this user study easy and interesting to accomplish and who are satisfied by the compensation offered will enrol to participate in the study.

The participants are completely anonymous and there is no way that we can know of their identity. As such, I do not foresee any ethical issues arising in sampling of participants.

The available participants will be pre-screened by a particular criterion by the Prolific platform, which is given below in the Recruitment Procedures section.

7. Recruitment procedures

Give details of how potential participants will be identified and approached. Also any possibility of coercion or conflict of interest and how this will be addressed. For example, where the participants are known to the researcher either personally or professionally.

The study will be published on the online crowdsourcing platform Prolific. Participants willing to participate in the study will be internally filtered in the crowdsourcing platform based on their approval rating (a metric demonstrating effectiveness in participating in surveys). We will pre-screen participants who have at least an approval rating of 96% (i.e. 96% of their previous submissions have been correct and accepted). This is to ensure that we have the best performing participants. They are anonymous participants, identified only by a user id of the crowdsourcing platform Prolific. Their identity is not revealed, hence there will not be any conflict of interest.

Application Form for Contextuality

STRICTLY CONFIDENTIAL

The first step is to record the user id of each participant. Thereafter, the participants will be shown an information sheet and will be asked for their consent. If they do not consent, they will automatically exit from the study.

We plan to have 240 participants after the pre-screening stage.

8. Recompense to participants

Give details of any recompense which will be offered to research participants or volunteers, e.g. a small payment or gift voucher. Participants should not be disadvantaged so it is usual to compensate them for their time, although it should not be considered a benefit or inducement. More guidance is available in FAQ 12.

The participants will be compensated in cash for their participation in the study at the rate of £7.03/hr. Estimating that the completion time for the whole study is 7 minutes, the participants will be paid £0.82 for the completing the study. This amount multiplied by the number of participants (240) will be pre-paid to the Prolific platform (without which we cannot start the study). The platform in turn will distribute the compensation to the participants who complete the study.

9. Consent

*Provide information on how valid consent will be sought from participants and attach copies of information sheet(s) and consent form(s). See FAQ 13 and FAQ 14 for guidance and templates. Consent forms and/or information sheets **have** to include the following or a rationale as to why not:*

- *Contacts; the PI and an alternative contact, e.g. Head of Department or supervisor, with respective OU email addresses.*
- *Clear information on how and when a participant may withdraw from the research. This should include a date or timeframe so it is clear that after the data gathering phase, when data may have been anonymised, it may not be possible to withdraw.*
- *Separate forms for each participant group - where applicable*
- *Information on how research data will be stored and disseminated/published and destroyed or retained (also see the OU data retention policy – internal link).*

The first page of the study would give all required information to the participant, as displayed in the attached Information Sheet. Upon reading it, the participant will be directed to a consent page where the questions given in the consent form (attached) will be asked. If the participant answers negatively to any of the question, they will automatically exit from the study.

10. Location(s) of data collection

Give details of where and when data will be collected, with an explanation of why the research needs to be conducted in the chosen setting or location. If it will take place on private, corporate or institutional premises, indicate what approvals are gained/required.

Data will be collected on the online crowdsourcing platform Prolific in January 2019 following OU Human Research Ethics approval. The need for crowdsourced data collection is because of the need of large number of participants (240) and less time for lab-based user study.

Application Form for Contextuality

STRICTLY CONFIDENTIAL

11. Literature review

Provide a brief review of the existing literature or previous research. Clarify whether the proposed study replicates prior work and/or duplicates work done elsewhere and/or has an element of originality (maximum 200 words).

The field of Quantum Cognition [1] investigates the presence of Quantum-like statistics in human decision making. Particularly the aspect of contextuality which has so far been observed only in case of Quantum systems. In contextual systems, the value of a variable is not pre-defined and depends upon the context of measurement of the variable. Recent works [1,2] have found the presence of contextuality in decision making. Drawing from the theory and some techniques of these studies, I plan to investigate contextuality in user's decision of relevance of documents. This will be the first attempt to investigate contextuality in the area of Information Retrieval.

[1] Jerome R. Busemeyer and Peter D. Bruza. 2012.Quantum Models of Cognition and Decision (1st ed.). Cambridge University Press, New York, NY, USA.

[2] Víctor H. Cervantes and Ehtibar N. Dzhafarov. 2018. Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. Decision5, 3 (July 2018), 193–204

[3] Irina Basieva, Víctor H. Cervantes, Ehtibar N. Dzhafarov, and Andrei Khrennikov. 2018. True Contextuality Beats Direct Influences in Human Decision Making. arXiv:arXiv:1807.05684

Key Ethics considerations

12. Published ethics and legal guidelines to be followed

Detail which guidelines will be followed by the researchers.
For example: BERA, BPS, BSA, SRA, MRS, SPA, UK Evaluation Society (see FAQ 5 on the Research Ethics website for more information).

British Psychological Society (BPS) guidelines

13. Data protection and information security

If your research involves the collection of information about individuals, you need to be aware of and follow the Data Protection Registration process - please confirm that this has been done (see FAQ 7). Also, re: storage and disposal of data, you need to detail below the procedures and schedule (including dates) you will be following. Indicate the earliest and latest date for the destruction of original data, where it is required, or any archiving arrangements that have been agreed/permitted, and ensure this is included in the project schedule. You should also be aware of OU information security policy and guidance (see FAQ 8).

No information about any individual is collected.

14. Research data management, disseminating and publishing research outcomes

Application Form for Contextuality

STRICTLY CONFIDENTIAL

If not covered elsewhere in your application, please give details of how your research data will be managed including publishing and data retention. It is recommended that all researchers applying to HREC write a Data Management Plan (DMP), and guidance and templates for writing a DMP are available on the Library Research Support website, with links to OU Open Access and ORDO (Open Research Data Online). If you need further help contact the Library Research Support team and FAQ 16 for links and guidance. Any funding body requirements should also be provided, e.g. the Economic Social Research Council (ESRC) requests data is deposited in a repository.

15. Deception

Give details of the withholding of any information from participants, or misrepresentation or other deception that is an integral part of the research. Any such deception should be fully justified.

No deception

16. Risk of harm

Detail any foreseen risks to participants or researchers, e.g. home visits, and based on a risk assessment, the steps that will be taken to minimise or counter these. Consider the Lone working guidance (FAQ 18) and project risk assessment matrix (FAQ 14). If the proposed study involves contact with children or other vulnerable groups, you should comply with the OU Safeguarding policy and procedures FAQ 10. Also, confirm that the requirements of the Disclosure and Barring Service have been met by providing the relevant reference number and period covered - for each person involved in the research.

17. Debriefing

Give details of how information will be given to participants after data collection to inform them of the outcomes of their participation and the research more broadly.

The participants will know the name and affiliations of the investigators and can check out their publications after 6 months of data collection to know about the results of their study.

Project Management

18. Research organisation and funding

Please provide details of the principal funding body (internal or external). If your project is part of a current or successful externally funded bid, enter your Award Management System (AMS) reference number below. For further guidance contact your Faculty Research Administrator (FRA) or refer to the Research and Enterprise website (internal site).

Funding body: European Union's Marie Curie Horizon 2020 Framework

Application Form for Contextuality

STRICTLY CONFIDENTIAL

AMS reference number:
721321

19. Other project-related risks

Indicate how research risks will be limited by detailing anticipated or potential problems. If you are carrying out fieldwork in the UK or overseas, you should be aware of the OU Fieldwork (FLD) travel advice and International Travel Risk webpages FAQ 18 (internal link).

No such risks

20. Benefits and knowledge transfer

State how the research may be of general benefit to participants and society in general (100 words maximum).

Presence of contextuality in user decisions in search can radically change our understanding of the nature of human judgments. We would have to redefine our models all of which assume that there is no contextuality in the data.

21. Supporting documents

Include as attachments or appendices, any documents related to your research proposal. Add the HREC reference number to each (if already known), and list below, for example:

Consent form and Participant information sheet – for each participant group	☒
Questionnaire	☒
Email or letter from the organisation agreeing that the research can take place	☐
Draft bid or project outline	☐
Publicity leaflet	☐

Application Form for Contextuality

STRICTLY CONFIDENTIAL

STRICTLY CONFIDENTIAL

21. Declaration

I declare that:

- The research will conform to the above protocol and that I will inform the HREC of any significant changes or new ethics issues and have these agreed before they are implemented.
- I have read and will adhere to the following OU policy:
    - OU Code of Practice for Research - http://www.open.ac.uk/research/ethics
    - OU Ethics Principles for Research involving Human Participants - http://www.open.ac.uk/research/ethics

Principal Investigator(s)	SAGAR UPRETY
School/Unit/Faculty	COMPUTING AND COMMUNICATIONS, FACULTY OF STEM
Telephone	+447424781376
E-mail	sagar.uprety@open.ac.uk
Signature(s) (scanned or electronic)	Sagar Uprety
Date	27/11/2018

HREC Final report

At the end of a HREC reviewed research project, Principal Investigators are required to assess their research for any ethics-related issues and/or major changes. Where these have occurred, the PI should return a completed copy of the HREC final report form to Research-REC-Review@open.ac.uk. Final reports are confidential and only made available to the HREC Chair and Committee members, and are requested to inform the HREC process, to assess how any ethics-related issues and major changes have been dealt with and to ensure OU research has been carried out as agreed. If you could add the date when your research is due to finish below, you will be sent a reminder.

Proposed date for final report: 30th August 2020

NB. Research students should enter their end of research project date

Research ethics applications - collection and use of data

Application Form for Contextuality

STRICTLY CONFIDENTIAL

titles of all projects considered by the HREC (either by HREC checklist or proforma) with HREC reference number, Faculty/dept. and HREC decision date, will be added to the Research Ethics website - http://www.open.ac.uk/research/ethics/human-research.

Information provided as part of a research ethics application, e.g. from research students or staff, is stored so the HREC has an accurate record. All data is managed and held securely by the Research Ethics Administrative Team and only shared with HREC members as part of the research ethics review process. Occasionally, and only where relevant, applications are discussed with like OU research review panels, e.g. the Staff Survey Project Panel (SRPP) and Staff Survey Project Panel (SSPP), predominately to avoid delays where applications are being made in tandem.

If, as part of a research ethics application sensitive personal data is disclosed, it will be stored securely and only shared as above. If such data is volunteered but then needs to be withdrawn, the researcher should contact Research-REC-Review@open.ac.uk. More information is available in the OU Student privacy notice and Staff, workers and applicants privacy notice.

Application Form for Multidimensional Relevance

HUMAN RESEARCH ETHICS COMMITTEE (HREC)
APPLICATION FORM

The Open University

All OU research involving human participants or materials requires assessment by the HREC.

Where you have determined your research requires a full review or you have completed the HREC Project Registration and Risk Checklist and been advised that will need to complete this form as part of the full review process, please complete and email this form to Research-REC-Review@open.ac.uk. Attach any related documents, for example: a consent form, information sheet, questionnaire, or publicity leaflet to ensure that the HREC Review Panel has everything they need to carry out a full review. If there are more than one group of participants, relevant documents for each research group need to be included.

If you have any queries about completing the proforma please check the Research Ethics website, in particular the FAQs - http://www.open.ac.uk/research/ethics/human-research/faqs which include sample documents and templates, or email Research-REC-review@open.ac.uk.

The deadline for applications is **every Thursday** by **5.30pm**. Applications are then sent to the HREC Review Panel with a minimum response time of 21 working days. However, the process can take a month or longer, so when planning your research and ethics application, you need to build in sufficient time for the HREC review to avoid any delays to your research. Particularly, when you are planning overseas travel or interviews with participants as it is essential that no potential participants are approached until your research has been fully assessed by the HREC.

Please complete all the sections below – deleting the instructions in italics.

Project identification and rationale

1. Title of project

Modelling User's Cognitive Dynamics in Information Access and Retrieval using Quantum Probability

2. Abstract

Contextuality is a phenomenon encountered in Quantum Theory wherein a measurement done under different contexts gives different results. Recently in the field of Quantum Cognition, there is growing speculation and some empirical evidence that this phenomenon is also present in human decision making. Explicit contextuality or context effects manifest as Order effects in decisions.

In this study we investigate context effects in questions pertaining to relevance of a document for a query. 3 questions regarding different perspectives of judging a document are asked to participants one group and the same 3 question with a different order are asked to another group. They report their answers on a 4-point likert scale so that uncertainty behind the judgements is captured. A Quantum model (a complex valued vector space) is assumed underlying such decisions and non-zero value of the complex number from the data will validate the Quantum model behind the context effects (order effects).

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

Project personnel and collaborators

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

STRICTLY CONFIDENTIAL

3. Investigators

Give names and institutional attachments of all persons involved in the collection and handling of individual data and name one person as Principal Investigator (PI). Research students should name themselves as PI and include a supervisor's electronic signature and/or comments (below) as evidence of supervisor support. Without this the application cannot be processed.

Principal Investigator/ (or Research Student): Sagar Uprety, The Open University, UK

Other researcher(s): Dawei Song, The Open University, UK

For students only:
Please note that this application cannot be processed without your supervisor's signature and/or supporting comments which should be provided below (Research ethics review: guidance for research students and supervisors):

Postgraduate research degree: PhD

Personal identifier: G2581947

Supervisor (preferably primary): Dawei Song

Email: dawei.song@open.ac.uk

Supervisor's electronic signature:

Supervisor's supporting comments:

Approved

Research protocol

4. Schedule

Time frame for the research and its data collection phase(s):

From: 3rd December 2018 To: 7th December 2018

Earliest date participants will be contacted: 3rd December 2018

5. Methodology

5. Methodology

Outline the method(s) that will be employed to collect and analyse data. Any relevant

Application Form for Multidimensional Relevance

documents, such as interview or survey questions or a participant information sheet, should be sent with the completed proforma. Where there are more than one group of participants, please provide separate consent forms and participant information sheets. If, for any reason, any of this is not possible please explain why.

Survey will be designed with relevant questions using the Qualtrics (https://www.qualtrics.com/uk/) survey management tool. This would include participant information and consent form as the first two pages of the survey. If the user does not consent to the study, he or she will be re-directed to exit the survey using skip logic built within Qualtrics.

An anonymous survey link generated will be distributed to users of an online crowdsourcing platform called Prolific (We have been recommended to use this platform by researchers working in the same area as it leads to better quality of data collection for decision-making related studies - https://prolific.ac) . Qualtrics will also be used for storing the responses collected. Also, the prolific id of each participant would be requested in the beginning of the survey, so that we can uniquely identify them if they want to pull out of the study later or want the responses given by them to be deleted.

Participants can ask for withdrawing from the study and deletion of their data within 10 days of their participating in the study. After 10 days, the data might have been used for doing numerical analysis and reported in a paper submitted to a conference.

6. Participants

Give details of the population targeted or from which you will be sampling and how this sampling will be done. Give information on the diversity of the sample.

The study will be published on the online crowdsourcing platform Prolific which has more than 60000 people registered who participate remotely in different user studies and surveys. They are from many different countries in the world, although more than 70% are from UK and US. They are predominantly between the ages of 20 and 40, and more than half of them are female. Participants who find the task of this user study easy and interesting to accomplish and who are satisfied by the compensation offered will enrol to participate in the study.

The participants are completely anonymous and there is no way that we can know of their identity. As such, I do not foresee any ethical issues arising in sampling of participants.

The available participants will be pre-screened by a particular criterion by the Prolific platform, which is given below in the Recruitment Procedures section.

7. Recruitment procedures

Give details of how potential participants will be identified and approached. Also any possibility of coercion or conflict of interest and how this will be addressed. For example, where the participants are known to the researcher either personally or professionally.

The study will be published on the online crowdsourcing platform Prolific. Participants willing to participate in the study will be internally filtered in the crowdsourcing platform based on their approval rating (a metric demonstrating effectiveness in participating in surveys). We will pre-screen participants who have at least an approval rating of 96% (i.e. 96% of their previous submissions have been correct and accepted). This is to ensure that we have the best performing participants. They are anonymous participants, identified only by a user id of the crowdsourcing platform Prolific. Their identity is not revealed, hence there will not be any conflict of interest.

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

The first step is to record the user id of each participant. Thereafter, the participants will be shown an information sheet and will be asked for their consent. If they do not consent, they will automatically exit from the study.

We plan to have 80 participants after the pre-screening stage for a pilot study and if successful, enrol further 220 participants, totalling 300.

8. Recompense to participants

Give details of any recompense which will be offered to research participants or volunteers, e.g. a small payment or gift voucher. Participants should not be disadvantaged so it is usual to compensate them for their time, although it should not be considered a benefit or inducement. More guidance is available in FAQ 12.

The participants will be compensated in cash for their participation in the study at the rate of £7.08/hr. Estimating that the completion time for the whole study is 5 minutes, the participants will be paid £0.59 for the completing the study. This amount multiplied by the number of participants (300) will be pre-paid to the Prolific platform (without which we cannot start the study). The platform in turn will distribute the compensation to the participants who complete the study.

9. Consent

*Provide information on how valid consent will be sought from participants and attach copies of information sheet(s) and consent form(s). See FAQ 13 and FAQ 14 for guidance and templates. Consent forms and/or information sheets **have** to include the following or a rationale as to why not:*

- *Contacts; the PI and an alternative contact, e.g. Head of Department or supervisor, with respective OU email addresses.*
- *Clear information on how and when a participant may withdraw from the research. This should include a date or timeframe so it is clear that after the data gathering phase, when data may have been anonymised, it may not be possible to withdraw.*
- *Separate forms for each participant group - where applicable*
- *Information on how research data will be stored and disseminated/published and destroyed or retained (also see the OU data retention policy – internal link).*

The first page of the study would give all required information to the participant, as displayed in the attached Information Sheet. Upon reading it, the participant will be directed to a consent page where the questions given in the consent form (attached) will be asked. If the participant answers negatively to any of the question, they will automatically exit from the study.

10. Location(s) of data collection

Give details of where and when data will be collected, with an explanation of why the research needs to be conducted in the chosen setting or location. If it will take place on private, corporate or institutional premises, indicate what approvals are gained/required.

Data will be collected on the online crowdsourcing platform Prolific in February 2019 following OU Human Research Ethics approval. The need for crowdsourced data collection is because of the need of large number of participants (300) and less time for lab-based user study.

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

11. Literature review

Provide a brief review of the existing literature or previous research. Clarify whether the proposed study replicates prior work and/or duplicates work done elsewhere and/or has an element of originality (maximum 200 words).

The field of Quantum Cognition [1] investigates the presence of Quantum-like statistics in human decision making. An interesting property of Quantum systems is incompatibility, or inability to measure two properties of a system at the same time (popularly known in terms of the Uncertainty Principle). This incompatibility is also present in human judgements when it is not possible to make a decision based on two different perspectives at the same time. The more a person is certain about one perspective, the less certain he or she is about the other. Thinking from one perspective influences the judgement about the other. This uncertainty manifests in forms of Order Effects in Psychology [2], where changing order of evidence presented effects the final judgement. Order effects are a form of context effect where asking a question (or considering a perspective) changes or creates context for the second question. Recent research has been successful in modelling, explaining and predicting order effects using the mathematical tools of Quantum Theory [3]. Recently, order effects have been investigating in relevance judgements of documents [4], where multiple perspectives of judging a document (e.g. whether judging relevance with respect to topicality or credibility of the information) seem to be incompatible and thus interfere with each other in the decision making. We propose to investigate incompatibility in three perspectives as opposed to two in previous literature, which enables us to represent the interference between the perspectives using a complex number or a phase angle, thus establishing a strong Quantum nature of judgement in case a non-zero phase angle is calculated from the data.

[1] Jerome R. Busemeyer and Peter D. Bruza. 2012.Quantum Models of Cognition and Decision (1st ed.). Cambridge University Press, New York, NY, USA.

[2] Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. Cognitive Psychology, 24, 1–55

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

[4] Bruza P and Chang V (2014) Perceptions of document relevance. *Front. Psychol.* **5**:612. doi: 10.3389/fpsyg.2014.00612

Key Ethics considerations

12. Published ethics and legal guidelines to be followed

Detail which guidelines will be followed by the researchers.
For example: BERA, BPS, BSA, SRA, MRS, SPA, UK Evaluation Society (see FAQ 5 on the Research Ethics website for more information).

British Psychological Society (BPS) guidelines

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

13. Data protection and information security

If your research involves the collection of information about individuals, you need to be aware of and follow the Data Protection Registration process - please confirm that this has been done (see FAQ 7). Also, re: storage and disposal of data, you need to detail below the procedures and schedule (including dates) you will be following. Indicate the earliest and latest date for the destruction of original data, where it is required, or any archiving arrangements that have been agreed/permitted, and ensure this is included in the project schedule. You should also be aware of OU information security policy and guidance (see FAQ 8).

No information about any individual is collected.

14. Research data management, disseminating and publishing research outcomes

If not covered elsewhere in your application, please give details of how your research data will be managed including publishing and data retention. It is recommended that all researchers applying to HREC write a Data Management Plan (DMP), and guidance and templates for writing a DMP are available on the Library Research Support website, with links to OU Open Access and ORDO (Open Research Data Online). If you need further help contact the Library Research Support team and FAQ 16 for links and guidance. Any funding body requirements should also be provided, e.g. the Economic Social Research Council (ESRC) requests data is deposited in a repository.

15. Deception

Give details of the withholding of any information from participants, or misrepresentation or other deception that is an integral part of the research. Any such deception should be fully justified.

No deception

16. Risk of harm

Detail any foreseen risks to participants or researchers, e.g. home visits, and based on a risk assessment, the steps that will be taken to minimise or counter these. Consider the Lone working guidance (FAQ 18) and project risk assessment matrix (FAQ 14). If the proposed study involves contact with children or other vulnerable groups, you should comply with the OU Safeguarding policy and procedures FAQ 10. Also, confirm that the requirements of the Disclosure and Barring Service have been met by providing the relevant reference number and period covered - for each person involved in the research.

17. Debriefing

Give details of how information will be given to participants after data collection to inform them of the outcomes of their participation and the research more broadly.

The participants will know the name and affiliations of the investigators and can check out their publications after 6 months of data collection to know about the results of their study.

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

Project Management

18. Research organisation and funding

Please provide details of the principal funding body (internal or external). If your project is part of a current or successful externally funded bid, enter your Award Management System (AMS) reference number below. For further guidance contact your Faculty Research Administrator (FRA) or refer to the Research and Enterprise website (internal site).

Funding body: European Union's Marie Curie Horizon 2020 Framework

AMS reference number:
721321

19. Other project-related risks

Indicate how research risks will be limited by detailing anticipated or potential problems. If you are carrying out fieldwork in the UK or overseas, you should be aware of the OU Fieldwork (FLD) travel advice and International Travel Risk webpages FAQ 18 (internal link).

No such risks

20. Benefits and knowledge transfer

State how the research may be of general benefit to participants and society in general (100 words maximum).

Establishing quantum nature of human judgements, particularly the context effects due to incompatibility in decisions can help us understand complex user behaviour better and thus build better models to predict human decisions under uncertainty.

21. Supporting documents

Include as attachments or appendices, any documents related to your research proposal. Add the HREC reference number to each (if already known), and list below, for example:

Consent form and Participant information sheet – for each participant group	☒
Questionnaire	☒
Email or letter from the organisation agreeing that the research can take place	☐
Draft bid or project outline	☐

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

Publicity leaflet	☐

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

STRICTLY CONFIDENTIAL

21. Declaration

I declare that:

- The research will conform to the above protocol and that I will inform the HREC of any significant changes or new ethics issues and have these agreed before they are implemented.
- I have read and will adhere to the following OU policy:
  - OU Code of Practice for Research - http://www.open.ac.uk/research/ethics
  - OU Ethics Principles for Research involving Human Participants - http://www.open.ac.uk/research/ethics

Principal Investigator(s)	SAGAR UPRETY
School/Unit/Faculty	COMPUTING AND COMMUNICATIONS, FACULTY OF STEM
Telephone	+447424781376
E-mail	sagar.uprety@open.ac.uk
Signature(s) (scanned or electronic)	Sagar Uprety
Date	27/11/2018

HREC Final report

At the end of a HREC reviewed research project, Principal Investigators are required to assess their research for any ethics-related issues and/or major changes. Where these have occurred, the PI should return a completed copy of the HREC final report form to Research-REC-Review@open.ac.uk. Final reports are confidential and only made available to the HREC Chair and Committee members, and are requested to inform the HREC process, to assess how any ethics-related issues and major changes have been dealt with and to ensure OU research has been carried out as agreed. If you could add the date when your research is due to finish below, you will be sent a reminder.

Proposed date for final report: 30th August 2020

NB. Research students should enter their end of research project date

Research ethics applications - collection and use of data

Application Form for Multidimensional Relevance

STRICTLY CONFIDENTIAL

titles of all projects considered by the HREC (either by HREC checklist or proforma) with HREC reference number, Faculty/dept. and HREC decision date, will be added to the Research Ethics website - http://www.open.ac.uk/research/ethics/human-research.

Information provided as part of a research ethics application, e.g. from research students or staff, is stored so the HREC has an accurate record. All data is managed and held securely by the Research Ethics Administrative Team and only shared with HREC members as part of the research ethics review process. Occasionally, and only where relevant, applications are discussed with like OU research review panels, e.g. the Staff Survey Project Panel (SRPP) and Staff Survey Project Panel (SSPP), predominately to avoid delays where applications are being made in tandem.

If, as part of a research ethics application sensitive personal data is disclosed, it will be stored securely and only shared as above. If such data is volunteered but then needs to be withdrawn, the researcher should contact Research-REC-Review@open.ac.uk. More information is available in the OU Student privacy notice and Staff, workers and applicants privacy notice.

Appendix C

This appendix shows all the document snippets used in my various user study based experiments. The queries are shown in table 1. Corresponding to each query the document snippets are shown in the following pages.

Comparison of the validity of two image ...

www.gov.uk › government › publications › breast-scre...

breast tomosynthesis, is a quite new breast **imaging** .... This prospective study comprised 1,000 consecutive bilateral screening **mammograms** (2,000 CC and 2,000 MLO images) ... Inclusion of the pectoral muscle on the CC was stated at 32% (inter-radiographer variability 22% - 60%) ... most challenging tasks in **mammography** identified by radiographers ... Focus on communication skills is pivotal for the patient's experience.

Figure 1: Document Snippet for query Q_1

Inbreeding benefits in wild populations? - Our Research

research.freelancers.org › users › rfr › inbreeding

5 Mar 2020 - .... reduced frequency of heterozygotes will reduce opportunities ... Such **inbreeding** results in the reduction in fitness that we term **inbreeding** depression.......most **inbreeding** depression observed in Drosophila and higher plants .... This could eventually lead to a 'mutational meltdown' for populations with an effective size (Ne) of <100 ... and some snails, **inbreeding** depression is estimated by comparing the fitness of self- and cross-fertilized progeny ... Many **animal** populations experience gene flow at high enough rates to reintroduce genetic load quickly via immigrants, violating the fourth condition ...

Figure 2: Document Snippet for query Q_2

16 cognitive biases that kill your decision making...

https://middleriddle.com › cracking-the-data-science-interview › decision-trees-...

Sep 8, 2018 - When you remove sub-nodes of a **decision** node, this process is called pruning .... Introduction to **Statistical** Learning by Gareth James ... One of the paradoxes of life is that **our** big decisions are often less calculated than .... choices; **facts** go undiscovered, ignored, or misunderstood; **decision-makers** are .... Eventually, an **aide** helped **my** wife into a wheelchair.

Figure 3: Document Snippet for query Q_3

Lead interactions- Global Health

www.wholeadthechemical.in › introduction › Fact sheets › Detail

23 Aug 2019 - Chronic **lead** nephropathy occurred due to years of **lead** exposure manifested in kidney biopsy by moderate focal atrophy, loss of proximal tubules and interstitial fibrosis ... **Lead** directly targets testicular spermatogenesis in the epididymis inducing reproductive toxicity ... prepubertal animals .... **lead** interactions with **human** chorionic gonadotropin (HCG) .... by UV-vis absorption spectroscopy, circular dichroism spectroscopy ... deterioration of chronic renal insufficiency ...

Figure 4: Document Snippet for query Q_4

Anti-smoking 'wonder pill' launched | Daily Mail Online

www.dailymail.co.uk › news › article-420473 › Anti-s...

4 Dec 2006 - ... average insured payment for health care of $1,145, while non-**smoking** employees average $762...'wonder pill' that claims to be ... way on the pleasure centre of the brain to cut the satisfaction ... The **cost** of the drug at around £2 a day could result in ... U.S. Environmental Protection Agency (EPA) estimates smoke-free restaurants save about $190 per 1,000 square feet in lower maintenance costs ... **Smokers**, on average, miss 6.16 days of work per year due to sickness ...

Figure 5: Document Snippet for query Q_5

Here are some possible radiation dangers in your environment....

https://hustlebustlenews.com/here-are-some-radiation..

May 5, 2014 - A study examined the role of occupational RF/ MW-EMF exposure in the risk of meningioma.....the International Commission on Non-Ionizing Radiation Protection. Several conditional logistic regressions performed for glioma and meningioma. No significant association.....However, the slight increase in risk...merits further research....

Figure 6: Document Snippet for query Q_6

What Chronic Wasting Disease and Mad Cow Disease can Teach Us ...

https://rightasrain.uwmedicine.org/.../what-chronic-wasting-disease.../

Apr 16, 2018 - The most well-known prion **disease in humans** is variant Creutzfeldt-Jakob disease. ... bovine spongiform encephalopathy, a.k.a. **mad cow disease**, in the ... If there is lead or mercury in the environment, pets often show **signs** ...

Figure 7: Document Snippet for query Q_7

7 Ways that Social media is affecting us positively - Currati

https://curatti.com/social-media-positive-effects/

Feb 14, 2018 - Recently, we published the article "Is **Social Media** Really an Existential ... approach candidates through **social networking sites** like LinkedIn. ... **Social Media** Has a Lot of **Benefits** for Students and Teachers ... Did you know- 59% of schools say their students use....

Figure 8: Document Snippet for query Q_8

Neural Networks – Nature

www.nature.com › journals › neural-networks

Neural Networks provides a forum for developing an international community of scholars interested in all aspects of neural networks. Fields of specialization: Nonlinear dynamics, theoretical neural networks ... neural networks and fuzzy methods, synchronization of oscillators and chaotic systems ... Halliburton Energy Services Inc. owns several patents directed to drilling improvements that rely on neural network technology. Sports and fitness companies are also innovating with neural networks, with companies like Adidas AG and Fitbit Inc. receiving multiple patents in this area..

Figure 9: Document Snippet for query Q_9

Future of European democracy | Social & technological change

https://nextbillionnaire.com/categories/technology/

30th January 2018 - Once upon a time, many thought the internet would spawn a digital democratic utopia .... potentially harmful effects of facial-recognition software .... Its greatest contribution is also its greatest threat. The greatest thing the internet has done is to allow for the elevation of the individual voice .... Yet because everything is now on the same playing field, hate speech is on the same level in your daily life.

Figure 10: Document Snippet for query Q_{10}

Q No.	Query	Information Need
Q1	Education programs for mammographic image quality assurance	Find out about education programs available to provide training to improve mammographic image quality
Q2	swamp dwelling animals which could face genetic extinction due to interbreeding	To find information about swamp dwelling animals which could face genetic extinction due to interbreeding
Q3	statistics to aid our decision making	To find out how use of statistics can enhance our decision making process
Q4	areas of the human body which are the main targets for lead	Find which parts of the body are affected by the chemical lead
Q5	Companies policy against smoking lower maintenance costs	What evidence is there that companies which adopt a policy against smoking can lower their maintenance costs?
Q6	Radio Waves and Brain Cancer	Look for evidence that radio waves from radio towers or mobile phones affect brain cancer occurrence
Q7	symptoms of mad cow disease in humans	Find information about mad cow disease symptoms in humans
Q8	educational advantages of social networking sites	What are the educational benefits of social networking sites?
Q9	Patent applications for nonlinear neural network oscillators	Find out if there are patent applications filed in the area of nonlinear neural network oscillators.
Q10	impact of technology on democracy	Find about how the increasing spread of technology in our lives impacting democracy and the democratic institutions.
Q11	Positive effects of forest fires	Forest fires are increasing becoming bigger and more frequent in recent times. Are there any positive effects of forest fires?

Table 1: All queries and information needs

Should we let raging wildfires burn?

https://mrstree-services.org/blog/should-we-let-raging-wild-fires-burn

Large increases in the extent of deciduous forests as a result of increasing fire activity .... Conifer trees need fire .... moderate fire activity, trees like black spruce often regenerate immediately. But when northern forests burn too severely, deciduous trees like aspen and birch can outcompete conifers during post-fire succession .... renewal of soil chemistry .... replenishment of streamside vegetation, recycling of nutrients, and fire-adapted plants being dispersed.

Figure 11: Document Snippet for query Q_{11}