



Towards Automated Reasoning for Drug Discovery and Pharmaceutical Business Intelligence

Barry Robson*

Abstract

Many workers involved in drug discovery will have some early familiarity with the principles of quantum mechanics as applied in chemistry. Certainly those involved in computational chemistry, and particularly molecular modeling, will do so. This familiarity will however be in regard to the algebraic systems based on the imaginary number i that is required for wave mechanics and hence study of molecular properties. This does not exhaust the scope of quantum mechanics, and following an argument by Dirac, the larger picture including more flavors of imaginary number should be applicable to all aspects of human thought where numbers are involved. This includes probabilistic semantics and the construction of probabilistic networks that not only capture knowledge, but perform inference of a more general value to the pharmaceutical industry.

Background

Escalation of Pharmaceutical and Biomedical Data:

Large reservoirs of chemical or biomedical data, such as all US patents now accessible on the Internet, contain a huge amount of information for chemical business intelligence and drug discovery [1]. Large archives of structured biomedical data, such as patient records, contain a great deal of information about interrelations between clinical factors such as drug prescriptions and outcomes [2]. Collections of protein and nucleic acid sequences and structures have long been available and continue to grow for exploitation by the biotechnology and pharmaceutical industries [3]. Their impact on healthcare has been recently reviewed in historical and futurological context [4]. The ability to data mine such very large sources provides us with more reliable and objective probabilities associated with *rules* or *statements* about things in the world and the relationships between them. (e.g. Refs. [5,6,7]).

Purpose of this Report and Status of the Project:

The true value of the above resides not simply in sifting hundreds and thousands of rules arbitrarily in order to find interesting new ones, because the true importance of a single isolated rule lies in the context of other related rules, collectively providing a more complete weight of evidence. Rather, one wishes to use many rules collectively for automated inference and decision support. The present report describes the theoretical and algorithmic basis for such an inference and knowledge discovery approach, the detailed results of which we hope to report in further papers. The basic idea has already been applied in prototype applications to inference networks comprising probabilistic rules from large collections of medical data, including drug outcomes [2]. To be of value to drug discovery, this is now being extended by the generation of appropriate rules from 6.7 million proto-rules that link molecular formulae to patent text, derived

by automatic reading of all US patents [1]. The size and complexity of these data sets, and of many of the rules derived from them, motivates us to on-going theoretical and algorithmic development, and so the ultimate value is yet to be demonstrated. Nonetheless, we have already shown [1] that we can evolve rules about what appear to be novel drug candidates by combining the 6.7 million proto-rules with screening *in silico*, meaning here the simulation of candidate drug binding to appropriate protein targets. What we also want our methodology to include are *probabilistically qualified statements* about the particular relationships of types of molecule to various protein targets, the relevant disease states, possible drug side effects, synthetic method, shelf-life stability, and so forth. So far, pharmaceutical companies have for the most part used e.g. English to express such knowledge and, acting as “rules” exchanged between humans, these have served fairly well. However, as rules for use in automated reasoning they are at best qualitative, and not in an efficient canonical form. The basis of a *probabilistic semantics* is required that can capture rules from both data and human experts.

Lack of Universal Best Practice in Inference:

Unlike the simplicity of examples conveyed in most standard statistics textbooks, many rules from many data sources can interact in a complicated way as a probabilistic-logical network to determine the final prediction to support decisions. Largely that is because classical text approaches focus on hypothesis tests to establish what is in effect a *single* rule such as “A associates with B”, “The value of A correlates negatively with the value of B”, or “The value of A cannot be exceeding x by chance” or simply test that potential rules are implied somewhere in the data (e.g. χ^2 test). Data mining requires no prior hypotheses or hunches, and seeks to find all interesting rules subject to the computing power available, so raising the issue of how to combine them. Apparently tidy textbook treatments of probability and decision theory have fallen short of what is needed in practice, and in consequence the early innovative MYCIN Expert System approach [8] set the stage for a subsequent great diversity of approaches. This diversity begs the question of *best practice* [9].

*correspondence: barry.robson@quantalsemantics.com

Quantal Semantics Inc, North Carolina, US; St. Matthew's University School of Medicine, Grand Cayman; Department of Mathematics and Computer Science, University of Wisconsin-Stout, US; The Dirac Foundation, UK.

A familiar approach that is theoretically well founded on the classical view is the Bayes Net [10], but this is traditionally confined to networks that are acyclic directed graphs, and involving only AND logic as implied by the multiplication of conditional probabilities of general form $P(A | B \& C \& \dots) = P(A \& B \& C \& \dots) / P(B \& C \& \dots)$ as the rules. It is equivalent to a fully connected graph in which rules that would result in cyclic paths through the network are assigned probability 1, and hence need not be expressly included. This is consistent with information theory ($I = -\log P$, $0 = -\log 1$) and the thesis of Popper [11], but it is a bad assumption made in many instances even given the available raw data that could correct it. In consequence, non-traditional Bayes Nets that can have cyclic paths have been developed, but are seen as requiring iteration [12]. For a large network expressing knowledge, this is time consuming.

The Semantic Web:

All this uncertainty about best approach now becomes pressing with the emerging worldwide semantic web (SW), where universal best practice for harnessing *probabilistic* statements in semantic format is recognized as desirable but still problematic [13]. A truly probabilistic SW, i.e. one that employs probabilistic semantics throughout with certainty simply as a limiting case of probability 1, would imply an enormous reservoir of rules or statements obtained by data mining or expert opinion that can be used in methods of inference. These rules either recognize uncertainty about a statement, or the fact that they are only observed as true in a fraction of observed cases; either way, they need to be associated with a probability. Importantly, compared with Bayes Nets as our reference case, they introduce the further feature of not simply a conditional probability relating, in the simplest instance, two states events observations or measurements A and B, but also a relationship description or *relator* with the linguistic force of a verb or preposition (or verb or prepositional phrases) relating them. Comprising three things, nouns or noun phrases A and B, and the relator, they are said to constitute *triples*. This is still a simplification compared with the more complex human sentence, but the approach below does provide a basis for a richer treatment to be described elsewhere.

Limitations of Conditional Probabilities:

Only in simple categorical cases that are typically matters of documented history such as $P(\text{"carbenoxolone is a synthetic derivative of glycyrrhetic acid"})$, or of definition, can we interpret them as conditional probabilities with probability one. If it were less certain, and without loss of generality, it can be expressed in conditional probability form $P(\text{"a synthetic derivative of glycyrrhetic acid"} | \text{carbenoxolone})$. Here the very specific nature of the first argument A in $P(A|B)$ alerts us that we would need a vast number of arguments A, B, C,... to represent knowledge that way. It also begs the question of how more profitably to write, evaluate, and use probabilistic triples for relators such as verbs of action that are *not* purely categorical, such as $P(\text{carbenoxolone} | \text{inhibits} | \text{11Beta-hydroxysteroid-dehydrogenase})$. Unfortunately, the format $P(A| \text{relator} | B)$ has no classical probability counterpart, except that we might break it down into its three components where P(A) is a row vector, the relator is a matrix, and P(B) is a column vector. This idea has a strong *relation* to the mathematical system and notation

developed by Dirac [14] for quantum mechanics (QM), which is also a probabilistic inference system. In that notation one writes $\langle A |$ relator $| B \rangle$. It conveniently looks like an extension of XML to handle semantic relationships, but in fact, as discussed below, it is what Dirac's $\langle A |$ relator $| B \rangle$ *more precisely* implies algebraically that is interesting, and it shows that it is not simply $P(A| \text{relator} | B)$ that is wanted.

Appeal to Quantum Mechanics:

The QM approach is compelling for several reasons. One is that Dirac's approach in particular provides many tools and ideas to draw from. It is widely used even in traditional (pre-Dirac) quantum mechanics which is based on complex algebra using the imaginary number i (the square root of minus one). Importantly for present purposes, however, a further aspect of Dirac's mathematics often called the *Clifford-Dirac calculus* is not confined to i -complex algebra. This extended algebra forms the basis of modern particle physics, both the so-called *standard model*, and the spinor, twistor, and string theories developed from it [15]. A more compelling strategic reason for using QM in general, however, is that it is widely held by physicists to be the required *universal best practice* for representing knowledge and inference from it on all scales from the subatomic to the cosmological [15]. The Feynman path integral approach [15] may be in particular be seen as performing inference on the particle physics analogue of a knowledge network. But no less compelling is the remarkable relation to semantics. More specifically, Dirac saw his treatment of QM as part of general language that would certainly encompass probabilistic semantics: "*The methods of theoretical physics should be applicable to all those branches of thought in which the essential features are expressible with numbers*" (Dirac's Nobel Prize Banquet speech, 1933 - that he included human thought *per se* and its communication is clear because, rightly or wrongly, he excluded those with emotive content, poetry and economics). He did not however leave clear instruction on how this application was to be done, as if it could more-or-less be applied as-is.

Semantic Significance of the Adjoint Operation:

At first, features of QM that will be familiar to many theoretical and computational chemists do look promising. Many of QM's tools relate remarkably well to transformations and symmetries implied in linguistics, at least symbolically, because $\langle A |$ relator $| B \rangle$ encodes rather more grammar information than even $P(A| \text{relator} | B)$ might suggest. One way of looking at this is through the fundamental QM algebraic operation known as taking the *adjoint* † which relates to potential reversal in time and causality, or more generally conditionality [15]. Potentially all algebraic symbols s in an expression e are subject to e^\dagger as the action of the adjoint on the expression, i.e. $s^\dagger \neq s$. Even bracket duals like (and) but which are not reflectively symmetrical, notably | and >, can change. However whether and how any s is changed by it depends on its susceptibility to two algebraic operations known as taking the *complex conjugate* $*$ (changing the sign of the imaginary part, if any), and taking the *transpose* T (interchanging rows and columns, if any) of the algebraic entity that the symbol implies. This is because $s^\dagger = s^{*T} = s^{T*}$. The syntax of most, and certainly Indo-European, languages is well constructed to reflect a subset of choices of symmetries related to these, through word order and active-passive tenses of verbs, and with

highly inflected languages like Latin even more so, but the underlying meanings are general to semantics because the knowledge in a network is a *directed* graph. In the semantic approach based on triples, one may imagine the relator as a label associated with the directed arcs (edges) of the graph envisaged as an arrow \rightarrow between two nouns or noun phrases as labels of the nodes (vertices), e.g. A and B. Here $A \rightarrow B$ equivalent to $B \leftarrow A$, with \leftarrow linguistically seen as the active-passive inversion of \rightarrow as a verb, e.g. between *chase* and *chased by*, and $A \leftarrow B$ is equivalent to $B \rightarrow A$, again an active-passive inversion, but both are distinct in meaning from $A \rightarrow B$. The fact that $\langle A |$ relator $| B \rangle$ is *some kind of scalar complex value* (on occasion it can be purely real) provides one sufficient criterion for making it susceptible to the action of taking the adjoint, and this is what allows the value of $\langle A |$ relator $| B \rangle$ to encode $P(A | \text{relator } P|B)$ and $P(B | \text{relator } P|A)$. In contrast, classical probability P of any kind is always a *scalar real value*, and so not susceptible to taking the adjoint. We can attempt to talk about the adjoint transformation of a classical conditional probability $P(A|B)$ to $P(B|A)$ as a symbolic manipulation, but quantitatively $P(A|B)^{\dagger} = P(A|B)$ so there is no way to calculate the value of one from the other alone. $P(A | \text{relator } P|B)$ does much better but it is only a half way house to encoding the above-mentioned four relationships involving A, B, \rightarrow , and \leftarrow , and getting them to behave in the required way. As a vector-matrix approach, $P(A | \text{relator } P|B)$ is susceptible to the action of † , but not $*$. $\langle A |$ relator $| B \rangle$ covers both, and in this sense, probabilistic semantics is much more related to QM than it is to classical probability theory.

Difficulties of Pre-Dirac Quantum Mechanics for Semantics:

In the above, we are in effect asking to replace A, B, C, etc. as normally applied to fundamental particles and their properties, and molecules and their properties, by macroscopic everyday objects or properties of them. It is well known that Schrödinger noted that nothing in the algebra of QM appears to prohibit this, and it led to his famous thought experiment that a cat can be alive and dead at the same time, in superposition of states. Notoriously, QM gives bizarre predictions such as superposition of states, and non-locality, on the range of scale that is everyday human experience. We here examine how this difficulty is overcome within a QM formalism, and how it may be utilized for probabilistic semantics to describe the everyday world in a quantitative way. It is arguable as to whether this *is* QM or simply some math borrowed from it. The former may be argued, because standard basic QM calculations can also be performed with the same prototype software.

Methods

General Description:

There is a blurring between the theory and methods of this approach, because to a significant extent theory is represented by what the user writes as a program to be compiled. The prototype system constructed is essentially a compiler and executor for QM expressions on an input file. A great deal of what the executing program finally does is defined by input, where many other approaches to inference might “hard code” the actions. The fixed and brief content of the compiler is compensated by the effort represented in input but this confers great flexibility of use, and those parts of input which are very general in nature can be retained from project to project. The term “user” is

often employed below for the person who programs the input in this way, though the program’s actual end user will not usually be the programmer, because programming requires some expertise. Notably, the focus is on complex valued vectors and matrices in Dirac bra-ket notation, along with the operators that act on them. An example bra is $\langle 1 | \text{Beta-hydroxysteroid-dehydrogenase} |$ and an example bra-relator-ket is $\langle 1 | \text{Beta-hydroxysteroid-dehydrogenase} | \text{binds} | \text{carbenoxolone} \rangle$. Again, here *binds* is the *operator* in QM terms, or relator (predication) in semantic theory terms. Such entities are used to define a network for inference purposes, called the *Dirac Net*. As discussed below, and as for any algebra, such entities can contain variables such as $\$A$, $\$B$,... However, their role is different, such that, for introductory purposes, we can think of the role of the input as assigning values to single bra-relator-kets $\langle A |$ relator $| B \rangle$ that describe the probabilistic relations between nodes A and B. This is analogous to assembling a set of conditional probabilities $P(A | B, C, \dots)$ to define a Bayes Net. Nonetheless, initial emphasis will be placed on definitions using expressions with variables like $\$A$, $\$B$, etc, because that is a considerable differentiator, and the relatively fixed part, whereas expressions lacking variables are essentially data that may frequently change.

Program Flow:

In the present prototype, there is no program flow control (‘go to’, loops etc.) in the QM language; what comes earlier is regarded as potentially *definitional* of what comes later. Interacting entities must be defined so there is no use of algebraic () brackets, and $a = b(c+d)$ with the parenthetic expression $(c+d)$ would be rendered as $e = c + d$ defining e , followed by be . That said, an operation can always be represented as a sequence of operators, $a b c d e$, and bras and kets can contain brackets and bra-relator-kets with ‘|’ and ‘>’ which do have similar effect as parenthetic expressions in ‘(’ and ‘)’. These are beyond present scope and will be discussed elsewhere. Unlike normal programming languages but like expressions in classical logic, a form such $\langle A | B \rangle$ or $\langle C | D \rangle$ can be to the left of an assignment. Such are stored as a unit representing a part of the “giant expression” that represents the Dirac Net, and in this case represents an OR “gate” whereas otherwise AND is implied.

Association Variables:

bras, kets, brackets, bra-relator kets (and ketbras of form $| A \rangle \langle B |$ which are matrices) are called *association variables* as they are really stored by the compiler as association (or hash) arrays, the keys of which are string constants such as *ethanol* which variables match (see discussion later below).

Defining Basic Format:

For semantic application, there is no constraint on input to mean that input and hence output is, in effect, English, nor any human language except for convenience and standardization. To allow this, a few more-or-less fixed format defining forms should appear early in input, albeit that the current implementation as following is somewhat a matter of taste, and is not fundamental to the approach save to illustrate flexibility.

$\langle \$A | \$A \rangle = \langle \$A | \$A \rangle$ #Define the symbol, here ‘=’, for semantic equivalence.

$\langle \$A | \$A \rangle \text{not} \langle \$B | \$B \rangle$ #Define the symbol, here ‘not’, for non-equivalence.

Note that, as throughout, there is one expression per line, almost always an assignment or symbolic representation of a class of assignments, and

the use of # to indicate comment. The '=' so defined is also subsequently an assignment of the value of the expression on its right to that on the left, which is usually a bra-relator-ket. To highlight that, one could choose ':=' rather than '=', for more general reasons called the *metadata operator*. As an example of power in the hands of the programmer, for better or worse, subsequent basic format definitions such as $\langle \$A := \$B \mid \$C \rangle = \langle \$B \mid \$A \ \$C \rangle$ could then define the algebra that the operator implies. Such approaches that are theoretically controversial, or are arbitrary although self-consistent, are left to the user as programmer. Basic format definitions and the kind of expressions that are now to be discussed are in many respects a research bench for the inference system developer.

Binding Variables:

In practice the above kind of statements are not further used once the job of defining basic symbols is done, but others that define format, and a large number that do not, can have a role that persists, as follows.

$\langle \$A \mid \$B \rangle = \langle \$A \mid \$C \rangle \text{and} \langle \$C \mid \$B \rangle$ #Define logical AND, and a basic type of syllogism.

$\$A$ etc are as before obviously general symbolic variables. The difference from the basic format definers is that the variables are distributed in a way that will provide meaningful match in expressions. They are here called *binding variables*, and the whole expression above is a *template*. Note that two binding variables of differing name such as $\$A$ and $\$B$ cannot normally stand for the same thing such as the word *ethanol*. $\langle \$A \mid \text{bind} \mid \$B \rangle$ will match $\langle \text{macromolecules} \mid \text{bind} \mid \text{ligands} \rangle$ but not $\langle \text{macromolecules} \mid \text{bind} \mid \text{macromolecules} \rangle$, which requires a separate $\langle \$A \mid \text{bind} \mid \$A \rangle$ or $\langle \$B \mid \text{bind} \mid \$B \rangle$. It is axiomatic that different universal variables $\$A$, $\$B$, etc. stand for different things, while any one e.g. $\$A$ always stands for the same thing. Otherwise we can use any variable names we want, as their scope does not extend beyond the expression. Templates do not normally have assignment of constant values as input, i.e. writing numbers to the right of '='. It is in fact possible to do so, but the values then naturally relate to degrees of confidence in the resulting probabilistic rule or statement, and we will here assume for simplicity that confidence is the default of probability one, as in all cases when numerical assignment is absent. Formally, assignment expressions per line that are templates represent *program* and those remaining (because they do not have any variables of this kind) represent *data*. Recall, however, that the whole set of data even by itself does imply an expression representing a static network that can be evaluated. "Program" and "data" lines may normally be mixed in order for human readability. The above template example is, incidentally, an example of what is fixed inside the compiler. Once logical *and* is so defined as a "symbol", whatever it may be (say French "et") it can be omitted by default, and it will imply multiplication of the two brackets.

Defining Relators as Algorithms:

In this report, operators as relators such as verbs and verb phrases (or prepositional and other relationships) that cannot be defined in any basic logical way from what is defined so far, can be defined directly either as algorithms or as matrices. In practice, both algorithms and

matrices are defined within subroutines which are currently written in Perl 5, and which the user includes in the input file, though in emerging versions, constant values of matrices can be applied to variables by an assignment statement. If a new operator is encountered which is not defined by the thread of previous definitions, it searches all the following lines for the subroutine of same name to define it. The following illustrates how an explicit expression with a new operator can be followed by its general definitional subroutine, here for brevity showing the start of that only.

```
<6|more than|3>
sub more_than #This is an example of a basic action defined by a
subroutine.
{
```

This defines the evaluation $\langle 6 \mid \text{more than} \mid 3 \rangle$ which is in this case the scalar value 1 if true, and 0 if false. Functions such as *log* are similarly seen as operators, and are definable by the user in input. Expressions with operators that are not defined in this way may still have meaning by preceding definitions, and if not can still be manipulated by grammars represented in templates, and if not even that, they can still be used as entities having a value like a conditional probability in a Bayes Net. The following is the most common alternative to defining operators as algorithms.

Defining Relators as Descendants of Defined Relators:

Extensions of definitions that are possible in terms of the ideas of adjoint, transpose, and complex conjugate need not necessarily be defined as algorithms and fall under the scope of the input:

$\langle \$A \mid \text{equal to or less than} \mid \$B \rangle = \langle \$B \mid \text{more than} \mid \$A \rangle$

This defines the converse such as active-passive inverse, and we may note the following examples. Again, defining relators first time by a subroutine is not essential for all purposes.

$\langle \$A \mid \text{includes} \mid \$B \rangle = \langle \$A \mid \$B \rangle$ #conversion of brackets to categorical bra-relator-kets, an important "seed step"

$\langle \$A \mid \text{include} \mid \$B \rangle = \langle \$A \mid \text{includes} \mid \$B \rangle$ #reduction to canonical form

$\langle \$A \mid \text{be} \mid \$B \rangle = \langle \$B \mid \text{include} \mid \$A \rangle$ #definition of active-passive inversion, and choosing use of 'be' as canonical

$\langle \$A \mid \text{be} \mid \$B \rangle = \langle \$A \mid \text{is} \mid \$B \rangle$ #reduction to canonical form

$\langle \$A \mid \text{be} \mid \$B \rangle = \langle \$A \mid \text{are} \mid \$B \rangle$ #reduction to canonical form

$\langle \$A \mid \$R \mid \$B \rangle = \langle \$A \mid \text{be} \mid \$B - \$R \text{ers} \rangle$ #An example generation of a non-categorical form, e.g. of # $\langle \text{cats} \mid \text{be} \mid \text{fish-eaters} \rangle$ to $\langle \text{cats} \mid \text{eat} \mid \text{fish} \rangle$

$\langle \$A - \$R \text{ers} \mid \$B \rangle = \langle \$A \mid \$R \mid \$B \rangle$ #Conversely, an example of conversion to bracket canonical form

$\langle \$A \mid \text{pays} \mid \$B \rangle = \langle \$A \mid \text{gives} \mid \text{money} \rangle \langle \text{money} \mid \mid \$B \rangle$ #Definition of 'pays' from 'gives' and money

Relators defined as above are stored with their definitions in a memory space called the *thesaurus*, for inspection of correct threading of multiple definitions dependent on each other. Those that have no such origins are described as *root*, which may mean that a subroutine of that name was used for the definition, or one was not found. A root relator is still capable of active-passive inversion, negation etc, and association variables such as a bra-relator-ket containing such can still be assigned values. In the above we were not concerned with plurality of nouns and the corresponding verb forms, and so reduce to a categorical form

based on the infinitive, but note for example

$\langle \$A | \text{are } \$B \rangle = \langle \$A \text{ | are } \$B \rangle$ #example of treatment of plural to canonical form

$\langle \$A | \text{is } \$B \rangle = \langle \$A \text{ | are } \$B \rangle$ #example of treatment of plurality conversion canonicalization

$\langle \text{a sheep} | \text{is } \$B \rangle = \langle \text{sheep} | \text{are } \$B \rangle$ #specific example of irregular noun

$\langle \$A | \text{some are } \$B \rangle = \langle \$A \text{ | are } \$B \rangle \langle \$A \text{ | are } \$B \rangle^*$ #specification of one kind of evaluation of the existential case

Treatment of these cases is elaborate and depend on the extent to which the user wishes to go in treating semantics as linguistics, say as good English, and not least a tool for exploring the underlying generative grammar for correct forms. It depends on whether we want to render all statements into standard form using, say, a canonical form of Ogden Basic English [16], or start from the basic forms and define more complex verbs, as in e.g. “pays” ← “gives money to”, or constantly explore both. One purpose here is to reconcile apparently different forms that are really semantically equivalent and hence mutually redundant. Another is to deduce one rule from two or more, or conversely decompose it into simpler rules.

Default Definitions and Manipulations:

In contrast to the above, there is a higher order of more fundamental symbolic representation (e.g. which applies to any relator as $\$R$, without specifying it). The following examples may be noted.

$\langle \$A | = (\$B, \$C)$

$\langle \$A | = | \$A \rangle^*$

$| \$A \rangle = \langle \$A |^*$

$| \$A \rangle = \$A | \rangle$

$\langle \$A | = \langle | \A^*

$| \$A \$B \rangle = \$A | \$B \rangle$

$\langle \$A \$B | = \$A^* \langle \$B |$

$\langle \$A | \$B \rangle^* = \langle \$B | \$A \rangle$

$\langle \$A | \$B \rangle = \langle \$B | \$A \rangle^*$

$\langle \$A | \$R | \$B \rangle = \langle \$B | \$R | \$A \rangle^*$ #Relators R are Hermitian

$\langle \$A | \$R | \$B \rangle = \langle \$B | \$R^* | \$A \rangle$ #Relators R are not necessarily trivially Hermitian

$\langle \$A | \$R^* | \$B \rangle = \langle \$B | \$R^* | \$A \rangle^*$ #Relators R are not necessarily trivially Hermitian

The difference is that covering all cases of interest for these symmetries would be extensive and computationally inefficient, so the essential features for semantic computation are default in the program, which is indeed concerned with matters of adjoint, complex conjugate, and transpose. The action of the above would be merely to redefine the notation as that the user wishes to use. But also, while relators are usually *non-trivially Hermitian* as defined by the above examples, which is the default relator = relator[†] but relator ≠ relator*, we can have special cases or exceptions to specify.

$\langle \$A | \text{is } | \text{green} \rangle = \langle \text{green} | \text{is a quality of } | \$A \rangle$

$\langle \$A | \text{marries } | \$B \rangle = \langle \$B | \text{marries } | \$A \rangle$

Active Variables:

Expressions such as $\$molecule = chloramphenicol$ are not templates, and the variable, distinguished by starting in lower case, is not a binding variable. Rather, once the variable is defined, they are active during the reading and interpretation of input, line by line. They can be seen as Perl variables. When { *executable Perl* } is encountered in an expression or stand alone on a line, it may return a value and may return a string that substitutes for the string ‘ { *executable Perl* }’. The returned value may be an empty string, in which case the executable Perl may do other things, such as re-compute what is stored in variable $\$molecule$. Note that $\$molecule = \{ \text{executable Perl} \}$ is permissible. The idea is evidently extensible to other programming languages than Perl.

Dirac Net Definitional Phase:

All input expressions are “definitions”; the word “definitional” here refers those expressions that define it with constant values, and so do not contain $\$A, \$B, \$R$ etc. As the first step, the compiler builds the Dirac Net from these; it represents a large expression in the Dirac algebra. It is important to understand that binding variables and their templates do not play a role in this first pass, and need not be present in input. Then there is only one pass and the NET it is said to be *static*, like a Bayes Net. Conversely, there “must” be at least one expression in input without binding variables, meaning that if there is not, the null net will return in the subsequent evaluation phase the scalar probability 1. When the expression to the left of the value assignment is a bracket of form $\langle A | B \rangle$ or bra-relator-ket of form $\langle A | \text{relator } | B \rangle$, or an expression that implies such, it is stored in the *network memory*; those to the right are not stored. Contrast this with the template which is the whole binding expression to be stored in *template memory*. The job of the expression to the right of ‘=’ is to assign probability values to those stored in network memory, directly as constant values or indirectly through expressions with active variables. These values are stored alongside the entities in network memory, actually as the values of bra-relator-kets etc. as associative variables, thus forming the analogue of a Bayes Net. They are generally but not always algebraic-complex quantities (with real and imaginary parts) as discussed soon below.

Dirac Net Evaluation Phase:

In the evaluation phase, the collective degree of truth of the Dirac Net as a knowledge network is evaluated as a complex number encoding two probabilities, called forward probability P_{fwd} and backward probability P_{bwd} , as discussed later below. It is sufficient for the moment that the following can be done, but for a basic understanding note that (a) given a “network” of only $\langle A | B \rangle$, $P_{\text{fwd}} = P(A|B)$ and $P_{\text{bwd}} = P(B|A)$, and given two or more such the resulting P_{fwd} and P_{bwd} is the product of all P_{fwd} and all P_{bwd} . More correctly, this is as long as logical *and* is applied throughout, i.e. multiplication is applied between brackets and bra-relator kets as for conditional probabilities in a Bayes Net. Similarly, for the AND-only case, *order is immaterial*. They comprise a set. Statements about the world that are not included in that set have the same effect as if they were included with probability one, and note that including many irrelevant statements with lower values that one can only lower the overall probabilities. Implicit semantic triple forms, $\langle A | B \text{ and } C \text{ and } D \rangle$, or explicit triple form with categorical relators

<B and C and D | are |A>, are still said to be triples despite joint multiple arguments B and C and D. They provide the counterpart of $P(A | B \text{ and } C \text{ and } D)$ in a Bayes Net. In many cases, the implication of non-categorical relators may not have particular insightful consequences. We cannot compute from multiplying <dogs| chase |cats> and <cats| chase | mice> the probability that dogs chase mice, only the probability that dogs chase mice-chasers. However, a causal system is more impactful in interpretation of final values. <locals| eat |fish><fish | swim in| lake><lake| is | contaminated> gives the probability P_{fwd} that locals are contaminated, and the probability P_{bwd} that the etiology of that contamination is the lake (see discussion below on bidirectionality). <flu> <exposed| to |flu> <infected | if |exposed> <symptomatic| if |exposed> <seriously ill| if |symptomatic> <die | if |seriously ill> is similarly a typical epidemiological calculation. Input carefully chosen so that the rules are indeed relevant and impactful is called a *relevancy set*. The interface allows one to write, save, edit, open and run multiple relevancy sets as text files. They can also be joined into one relevancy set. The P_{fwd} and P_{bwd} result from that will be the product of P_{fwd} for the two sets and P_{bwd} for the two sets, though if further steps using binding are implied, that is not generally true, as follows.

Dirac Net Evolution Phase with Binding Variables:

A more advanced aspect of flow is that the network is *dynamic*. In this phase the binding variables come into play, if present in input. The dynamic aspect arises in that templates may also be seen as *editing instructions* that can convert one or more data bra-relator kets to one or more other data bra-relator kets. A template <\$A | are not | \$B> = <non \$B | are |\$A> uses the logical law of the contrapositive to express the desire of the programmer to convert all forms such as <birds| are |non mammals> to a canonical form <mammals| are not | birds>. Another example template that illustrates the fuller nature of the process is one that defines the valid forms of syllogisms. An example of a syllogism given above was <\$A| are |\$C> = <\$A| are |\$B><\$B| are |\$C>, which is the whole template. The *match part* comprises the bra-relator-kets <\$A| are |\$B> and <\$B| are |\$C> in the expression on right hand side of the assignment. These component bra-relator-kets, as features of “program”, hunt out those as bra-relator-kets as “data” in the network that match to them. They insert them into a copy of the expression on the right side of the template, and force evaluation of that expression, in this case simply a product of two bra-relator-kets. Note that there is only a match if different binding variables match different constant parts in the “data” bra-relator-ket, and if the same binding variables match the same constant parts in the “data” bra-relator-ket, and if all remaining constant parts of the bra-relator-ket match. The *edit part* of the template is the bra-relator-ket <\$A| are |\$C> on the left side of the assignment. From the relationship between the right hand side of the template and the bra-relator-kets matched, and the bra-relator-ket on the left side, it deduces the specific form of <\$A| are |\$C> on the left of the template, i.e. with the binding variables replaced by constants. At present, only one bra-relator ket can replace one or more. This shrinks the network. However, a mode may be applied that reverses the editing process and expands the network. The order in which templates are applied to edit is arbitrary, and the resulting network and its value often order independent, but not generally so, hence the following.

Dirac Net Optimization Phase:

The order of template application can be randomized and for each random choice the process of net evolution and net evaluation is applied in a step called *local optimization*. This process repeated many times in the hunt to achieve what is optimal attempts *global optimization*, and may consist of various heuristic algorithms to direct the search in addition to randomizing the order in which templates are applied. For example, recall that confidence in the resulting probabilistic rule or statement can be applied to a template; templates with more confidence can be applied earlier. These issues otherwise lie beyond present scope and are under ongoing development, and will be described elsewhere, but some interesting general or typical findings are discussed in Results. The overall process is halted when no better optimum is found after a specified number of iterations, which has the appearance of convergence of the evaluation of the network when plotted.

Reconciliation:

The above has omitted an important issue, except by brief reference to statements which look different but are semantically equivalent. Network evolution as editing of the network can generate rules for the network that can be detected as having similar semantic content to rules already present. This is facilitated by reducing all statements to a canonical form, say with the verb “to be” and negatives reflected in negation of the verb, but that is a matter of the threading of definition presented in input. Reconciliation of detected similar forms is by

$$\langle \text{canonical form} \rangle = \langle \text{canonical form} \rangle + \langle \text{new rule} \rangle - \langle \text{canonical form} \rangle \langle \text{new rule} \rangle$$

It can be shown to be order independent and not to artificially increase information content of the system. Actually, even in a static net, this is applied, and is part of every evaluation step. The reason is that different experts could enter the same rule twice or more with the same or different probabilities. We almost always run the Net with an evolution phase as local optimization, because recursive use of suitable templates can then detect that the apparently distinct rules are really semantically equivalent when the relationship is not obvious. This generates the canonical forms, defined by the templates in input. The above reconciliation algorithm is not arbitrary and has a deeper significance. It is a kind of hard wired template rule for stating the probabilities associated with two rules reconciled as one are computed as *randomly associated OR*, that is, the rules are independent, but can be distinguished as statements about the world that can recur, such that they are countable. The reconciliation mechanism applied repeatedly to remove the same or semantically equivalent rule is actually a counting process, and repeated application in any order implies the binomial theorem and binomial expansion. In that sense, data mining can be done, and is not distinct from the inference process. If we really wish to count in the Bernoulli sampling sense, it is recognized that seeing one such specific relation out of an as yet unknown and potentially large number implies small probabilities associated with duplicate rules. A small arbitrary and constant probability value is assigned and later normalized, as will be discussed elsewhere. However, assignment of probabilities to rules by a preceding separate step of data mining, or by human experts, or both, are the norm and have meaning, as follows.

Empirical Assignment and Interpretation of Probabilities:

The probability assignment statement for the bra-relator-ket is the most important and is also the algebraic way of expressing what the content of any $\langle A |$ relator $| B \rangle$ is, in probability terms.

$$\langle A | \text{relator} | B \rangle = (P_{fwd}, P_{bwd}) \quad (1)$$

For example,

$$\langle \text{overeating} | \text{causes} | \text{obesity} \rangle = (0.9, 0.7)$$

The brackets may be omitted for such a scalar quantity, and we may also use 90%, 70%. In this case it implies simply

$$\langle \text{overeating} | \text{causes} | \text{obesity} \rangle = (P(\text{"overeating causes obesity"}), P(\text{"obesity causes overeating"}))$$

We can do much with a system based on this idea alone, but alone it says nothing about many useful things. These include how we involve $P(\text{overeating})$, $P(\text{obesity})$, other operators than AND, the role of the relator as a matrix and the consequence of relators acting on relators, the role and significance of mutual information, and the emergent properties of networks. We certainly could not show consistency with QM by doing *typical* QM calculations with this idea. To expand on and exploit the relationship to QM, we need to see how QM relates. QM usually calculates probabilities *ab initio* from the physics of the system of interest, while we want to derive bras and kets like those above with empirical probabilities data mined from the everyday world of human experience, or from human expertise. So, along with one other important modification relating to interpretation of the complex number, these probabilities somehow replace the normalized statistical weights $ke^{i\theta}$ of pre-Dirac QM. The part $e^{i\theta}$ certainly has probability-like qualities when $e^{\theta} \leq 1$, but will be seen as relating to *association constant* $K(A; B) = P(A, B) / P(A)P(B)$ such that $P(A)K(A; B) = P(A|B)$, $P(B)K(A; B) = P(B|A)$. $P(A)$ or $P(B)$ relates to k . In QM it is set by the nature and scale of the system under consideration such as $P(x)$ of position x of a particle on a circular orbit of length L , whence $k = L^{-1/2}$. The probability of any one *exact real* value with indefinite precision is vanishingly small, and QM overcomes this with the idea of a *Dirac delta function* to express that [14,15], but we are either interested in *nominal categorical data* such as *Lipitor*, or *range data* such as *Low_Density_Lipoproteins(mg/dL):='<130'*, being states with a non-vanishing probability, more akin to spin states fixed at one of two values. Note the appearance again of the metadata operator ':=', which is used to express QM's *projection* of values by the action of the relator. We now need to see (P_{fwd}, P_{bwd}) as a complex number. It would be possible to write the complex scalar value to the right of the assignment equality as (real part, imaginary part), as is customary in the i -complex case, but the following general approach taken allows the above particularly convenient method, and applies to any complex number j (see later below). The scalar complex value assigned is simply $(P_{fwd}, P_{bwd}) = \frac{1}{2} [P_{fwd} + P_{bwd} + j (P_{bwd} - P_{fwd})] = \frac{1}{2}(1-j)P_{fwd} + \frac{1}{2}(1+j)P_{bwd}$ (2)

Note that $(P_{fwd}, P_{bwd}) = (P_{bwd}, P_{fwd})^* = (P_{bwd}, P_{fwd})^*$. This equation is the usual form used in QM based on the *commutator* to obtain the required symmetry properties such as $\langle A | B \rangle = \langle B | A \rangle^*$ [15]. Here for the moment j is *some kind of* imaginary number with adjoint $j^* = j^*$. Above and in the following account, any (P_{fwd}, P_{bwd}) can

be replaced by a scalar real value. Inspection of Eqn. 2 shows this to be mathematically correct: if $P_{fwd} = P_{bwd}$, as in (0.6, 0.6), then the result of Eqn. 2 is a scalar real value, here 0.6.

Semantic Interpretation of Real and Imaginary Parts:

There is a semantic significance to this by the categorical interpretation of conditional probabilities, extended to bra-relator-kets later below. The real part $\frac{1}{2} (P_{fwd} + P_{bwd})$ of Eqn. 2 is the *degree of existential qualification*, the extent to which we can interpret $\langle A |$ relator $| B \rangle = \langle B | A \rangle$ as "some A are B" \equiv "some B are A". The imaginary part and commutator $\frac{1}{2}(P_{bwd} - P_{fwd})$ is the *degree of universal qualification*, being -1 for the strongest case of "All A are B" and $+1$ for the strongest case of "all B are A". These strongest values are not always achieved on a numerical interpretation of the categorical case. Whilst by definition $P(\text{"enzymes are catalysts"}) = 1$, $P(\text{"catalysts are enzymes"})$ can be considered as the fraction of individual cases observed as catalysts that are more precisely enzymes. Such a fraction is, nonetheless, typically a very small value, even smaller in $P(\text{"vertebrates are cats"})$ than in $P(\text{"mammals are cats"})$, and in cases like $P(\text{"inhibitors inhibit enzymes"})$ and $P(\text{"enzymes inhibit inhibitors"})$ the latter can be considered zero by definition. These considerations are universal and in QM or any other system have some kind of counterparts for any definition of j that allows an adjoint, as long as we are looking at one rule. However, when we bring rules together in different circumstances such as when, for example, we form syllogisms, it requires a deeper understanding of j .

The Hyperbolic Imaginary Number h:

Theoretical discussion is now required in reference to the nature imaginary number j . Dirac applied a *relativistic correction* to QM [14,15] by the transformation $i \rightarrow h$, often called a *Lorentz rotation*. Here h is the hyperbolic number such that $hh = +1$. The everyday world of human experience is a subset of the relativistic one at lower velocities and masses, at least for the things (other than such as light and the gravity of the Earth) which we experience directly. The Lorentz rotation implies a generalization of the Wick rotation (time $t \rightarrow it$). Feynman simplified the path integral by first rendering the system classical in a comparable way [15]. h was first defined algebraically much earlier, by Cockle [17], and also goes by his name, or as the Lorentz number, or as the *split complex number* amongst many others. It becomes manifest in one flavor in the relativistic correction when solving the square root of the wave equation as γ_{time} , and in equations comparable with those of particle physics, identifies with the physicist's hyperbolic imaginary number γ_5 . It is to be understood below that $\langle A |$, $| B \rangle$, $\langle A | B \rangle$, relator, $\langle A |$ relator $| B \rangle$ are in this report are h -complex unless stated otherwise. The important point is that the transformation $e^{i\theta} \rightarrow e^{h\theta}$ is not to another wave function, but a hyperbolic function.

$$e^{h\theta} = \iota^* e^{-\theta} + \iota e^{+\theta} \quad (3)$$

This follows the "iota notation" of Ref. [9]. By analogy with operators in quantum field theory [15], $\iota = \frac{1}{2}(1+h)$ and its complex conjugate $\iota^* = \frac{1}{2}(1-h)$ [9], and Eqn. 3 arises as $e^{h\theta} = \cosh(\theta) + h \sinh(\theta) = \iota^* e^{-\theta} + \iota e^{+\theta}$. This "iota notation" is simple to use algebraically, as supposed to constantly addressing h . It is readily shown to have the idempotent property $\iota = \iota$ and $\iota^* \iota^* = \iota^*$, the annihilation property $\iota \iota^* = \iota^* \iota = 0$,

and the normalization property $\iota + \iota^* = 1$. From these alone, example consequences are that $\iota^*e^{h\theta} = \iota^*e^{-\theta}$, $\iota e^{h\theta} = \iota e^{+\theta}$, $e^{h\theta} = \iota^*e^{h\theta} + \iota e^{h\theta}$, $\iota^*x + \iota x = x$, $(\iota^*a + \iota^*b)(\iota^*c + \iota^*d) = (\iota^*ac + \iota^*bd)$, $\log_c \iota^*x = \iota^* \log_c x$, $e^{(\iota^*x)} = \iota^*e^x$. As a practice, and illustration of relative ease of use as well as of its broader significance in underlying adjoint symmetries, one may deduce the following for the Riemann zeta function used in one approach to data mining [1,2,5,6,7]. It will also be needed later.

$$\begin{aligned} \zeta(s = \sigma + ht, n) &= \sum_{x=1 \dots n} x^{-(\sigma+ht)} = \sum_{x=1 \dots n} x^{-\sigma} e^{+ht \log_e x} \\ &= \sum_{x=1 \dots n} x^{-\sigma} \iota^* e^{-t \log_e x} + \iota e^{+t \log_e x} = \iota^* \sum_{x=1 \dots n} x^{-\sigma-t} + \iota \sum_{x=1 \dots n} x^{-\sigma+t} \\ &= \iota^* \zeta(s = \sigma+t, n) + \iota \zeta(s = \sigma-t, n) \end{aligned} \quad (4)$$

Eigenvalues of h :

With practice ι algebra is simple because ι has eigenvalues 0 and +1 when ι^* has eigenvalues +1 and 0, and *vice versa*. We can always see a meaning by substituting these eigenvalues. In the above example, $\zeta(s = \sigma + t, n)$ and $\zeta(s = \sigma - t, n)$ are the two real valued solutions, related by $\iota^* \zeta(s = \sigma + t, n) = (\iota \zeta(s = \sigma - t, n))^\dagger$. That follows from h having real eigenvalues -1 and +1, which is already simpler than i algebra because i and $-i$ are analogous eigenvalues of i , i.e. still complex. It has fundamental physical significance. Dirac introduces h very early in his book [14] as σ (not the above σ !) a linear operator such that $\sigma^2 = 1$ (his Eqn. 23), and $|P\rangle = \frac{1}{2}(1+\sigma)|P\rangle + \frac{1}{2}(1-\sigma)|P\rangle$. “It is easily verified that the two terms on the right here are eigenkets of σ , belonging to the eigenvalues 1 and -1 respectively, when they do not vanish”. They are always behind the scenes in QM and “do not disappear” algebraically. Multiplying by a bra and in present notation, this is

$$\langle A|B\rangle = \iota^* \langle A|B\rangle + \iota \langle A|B\rangle \quad (5)$$

It later appears as a decomposition called the *Dirac field* into left and right handed projections of the wave function which we relate to our P fwd and P bwd, and with each term seen as a spinor it is a dual spinor called the *Dirac spinor* [15], and idea that will become important below. The roles of ι^* and ι or their counterparts nonetheless are otherwise rather sparser in traditional i -complex QM equations than the above would suggest, precisely because what is usually written corresponds to partial or final solutions *after* substituting eigenvalues +1 and -1.

The Generalization as j :

The above being so, QM really rests on the *larger system* of the Dirac-Clifford calculus which is partly captured by $e^{h\theta}$. We will let $j = hi$ with Dirac’s eigenvalues $+i$ for $h = +1$ and $-i$ for $h = -1$.

$$\begin{aligned} e^{j\theta} &= e^{hi\theta} = \iota^* e^{-\theta} + \iota e^{+\theta} = \iota^* [\cos(-\theta) + i \sin(-\theta)] + \iota [\cos(\theta) + i \sin(\theta)] \\ &= \iota^* [\cos(\theta) - i \sin(\theta)] + \iota [\cos(\theta) + i \sin(\theta)] \\ &= \cos(\theta) + hi \sin(\theta) \end{aligned} \quad (6)$$

This indicates $(e^{h\theta})^* = e^{-h\theta}$ so that $j^* = -j$. It makes no comment on commutation properties of h and i , and so it does not automatically follow that $hi = -ih$, a typical anti-commutative feature in manipulation of quaternions and Clifford and Dirac calculus. With h and i seen as appropriate operators, we can have $(hi)^\dagger = (-i)(-h) = ih$. In any event, in practice, we simply set $j = i, j = -i, j = h, j = -h$, as these relate to general (not necessarily real) eigenvalues of j . We now wish to consider how to express wave systems and classical systems in terms

of empirical probabilities, though in semantics the former is only to demonstrate consistency with QM.

Conjugate Symmetry:

The remaining hurdle is that $e^{j\theta}$ is fundamental but has a very restrictive symmetry that may be called *conjugate symmetry* $e^{-j\theta} (e^{-j\theta})^* = 1$. The value of one conjugate variable determines the value of the other, so as noted by Chester [18] for $e^{i\theta} = e^{ix/p/h}$ it follows that, suitably normalized, then $P(x|p) = P(p|x)$ (his Eqn. 2.18): the *event reversal theorem*. It is very much what is *not* wanted for something like $P(A|B)$ if we want a distinct and useful adjoint $P(B|A)$. In physics, asymmetric examples are interaction with the Higgs particle and other external fields in Quantum Field Theory [15]. The role of the observer and experiment implied in Dirac’s ket normalization is an example of breaking that symmetry, and relevant here. Dirac’s ket normalization is part of his Recipe for obtaining observable probabilities [14] involving (1) normalization with the respect to the ket $\langle A|B\rangle$ such that in $\langle A|B\rangle$ the implied probability $P(B|A) = 1$ and (2) taking the product $P(A|B) = \langle A|B\rangle^\dagger (\langle A|B\rangle)^\dagger = \langle A|B\rangle^\dagger (\langle A|B\rangle)^*$. To obtain $P(B|A)$ one can first form $\langle A|B\rangle^\dagger$ and then proceed as above, or equivalently replace ket by bra normalization $\langle A|B\rangle$, though this replacement is unphysical for conjugate variables A and B, like p and x in QM, which is why ket normalization is the more general recipe. We need to break conjugate symmetry in a more general way. Perhaps the most general statement follows from Eqn. 5. If we have two algebraic expressions $\iota^* \langle A|B\rangle_1$ and $\iota \langle A|B\rangle_2$ and we wish them to be the parts $\iota^* \langle A|B\rangle$ and $\iota \langle A|B\rangle$ such that $\langle A|B\rangle = \langle B|A\rangle^*$, then we can form the following linear combination.

$$\langle A|B\rangle = \iota^* \langle A|B\rangle_1 + \iota \langle B|A\rangle_2 = \iota^* \langle A|B\rangle_1 + \iota \langle A|B\rangle_2^* \quad (7)$$

Non-Conjugate Asymmetry and Simple Empirical Assignments:

Eqn. 7 is of immediate practical importance because forms with relators R certainly do not in general have conjugate symmetry. We can write by definition

$$\langle A|R|B\rangle = \iota^* \langle A:=r|B:=r^*\rangle + \iota \langle B:=r|A:=r^*\rangle \quad (8)$$

In general operators of interest here are Hermitian and cast orthodata values to metadata A and B. Here ‘:=’ is again the metadata operator. In QM what we really mean by x and p is e.g. position(nm):=6.4 and momentum(Kg nm/sec):= 2.3, and in medicine we have e.g. Systolic_BP(mmHg):=140. More analogous to spin states of QM are the active and passive forms of R cast as the active form r and the passive form r^* (relative to the active-passive tense of the relator). A mundane example is

$$\langle \text{dogs} | \text{chase} | \text{cats} \rangle = \iota^* \langle \text{dogs} := \text{chase} | \text{cats} := \text{chased} \rangle + \iota \langle \text{cats} := \text{chase} | \text{dogs} := \text{chased} \rangle$$

Using these in a network, however, would assume the presence of $\langle B:=r^* | A:=r^* \rangle$. Even allowing an assumption of independence of $B:=r^*$ and $A:=r^*$ such that $\langle B:=r^* | A:=r^* \rangle = \iota^* P(A:=r^*) + \iota P(B:=r^*)$ this requires having a plethora of nodes for nouns associated with relators. Therefore however probabilities are assigned, it is useful to store these forms as follows.

$$\langle A|R|B\rangle = \iota^* P(A:=r, B:=r^*) + \iota (B:=r, A:=r^*) \quad (9)$$

It is relatively easy for a human expert to assign the condition probabilities as $\langle \text{dogs} | \text{chase} | \text{cats} \rangle = \iota * P(\text{"dogs chase cats"}) + \iota P(\text{"cats chase dogs"})$. This is still bi-directional in conditionality. By Eqn. 9, \mathbf{R} in $\langle A | \mathbf{R} | B \rangle$ is non-trivially Hermitian in that

$$\begin{aligned} \langle A | \mathbf{R} | B \rangle &= \langle B | \mathbf{R} | A \rangle^* \\ \langle A | \mathbf{R} | B \rangle &= \langle B | \mathbf{R}^* | A \rangle \end{aligned} \quad (10)$$

There are certainly trivially Hermitian relators, though, e.g. $\langle \text{Jack} | \text{marries}^* | \text{Jill} \rangle = \langle \text{Jack} | \text{marries} | \text{Jill} \rangle$. All such are real valued, because $\langle A | \mathbf{R} | B \rangle = \langle B | \mathbf{R} | A \rangle^*$.

The Importance of Mutual Information:

This topic will be needed for further development, but it also relevant to the above empirical treatment.

$$I(A; B) = I(A) + I(B) - I(A, B) = \log_e P(A, B) / P(A) P(B) = \log_e K(A; B) = \zeta(s=1, o[A, B]) - \zeta(s=1, e[A, B]), \mathcal{L}_e o[A] \rightarrow \infty, \mathcal{L}_e o[B] \rightarrow \infty \quad (11)$$

is the *Fano mutual information* [19] between A and B, with Robson’s treatment [5,6,7] for finite data, with observed and expected frequencies $o[\]$ and $e[\]$. The practical importance is that we get $I(A; B)$ by data mining. The theoretical importance is that the limit expressed for indefinitely large data means that we can always write $I(A; B)$ in terms of zeta functions ζ . One practical importance of $K(A; B)$ is that P fwd and P bwd do not carry enough knowledge to calculate $P(A)$, $P(B)$ etc with nodes A, B, and use these as prior probabilities if required. Nor can we evaluate probabilities as complementary or negative states such as $P(\sim A)$ and $P(\sim B)$, $P(\sim A, B)$ and so on. If we could, then a large variety of scientific measures such as predictive odds, likelihood ratios, odds ratios, number needed to treat can be determined [20]. All this is possible if we provide the association constant *assoc*, $K(A; B) = e^{I(A; B)}$, as well as P fwd and P bwd. For a simple bracket, $\langle A | B \rangle$ provides P fwd and P bwd as $P(A|B)$ and $P(B|A)$, and $P(A) = P(A|B)/K(A; B)$, and $P(B) = P(B|A)/K(A; B)$, from which the above may be calculated. So in provision of these kinds of probabilistic rules and statements over the internet, something like the following would be appropriate [20].

$\langle \text{overeating} | \text{P fwd} = 0.9 | \text{causes} | \text{assoc} = 6.8 | \text{obesity} | \text{P bwd} = 0.7 \rangle$
 $K(A; B)$ is conveniently written with the relator in such cases because its exponent may be seen as the eigenvalue of a linear Hermitian operator such that

$$\langle A | \mathbf{R} | B \rangle = e^{I(A; B)} \langle A | B \rangle \quad (12)$$

It is possible to show that the required form for a categorical relator like *include* or *if* (see further formal discussion below) is

$$\begin{aligned} \langle A | \mathbf{R} | B \rangle &= [\iota * P(A) + \iota P(B)] e^{I(A; B)} \\ &= \iota * P(A|B) + \iota P(B|A) \\ &= \frac{1}{2} [P(A|B) + P(B|A)] + \mathbf{h} \frac{1}{2} [P(B|A) - P(A|B)] \\ &= (\text{P fwd}, \text{P bwd}) \end{aligned} \quad (13)$$

If we consider a default bracket as $\langle A | B \rangle = \iota * P(A) + \iota P(B)$ which would be the case when mutual information is not available such that A and B are held to be randomly associated, then for a linear Hermitian operator \mathbf{R} , association constant $K(A; B) = e^{I(A; B)}$ is an eigenvalue. That is, it represents the one or more solutions to $\langle A | \mathbf{R} | B \rangle = K(A; B) \langle A | B \rangle$. At first consideration, Eqn. 13 needs no further insight for a *categorical*

relator \mathbf{R} such that $\langle A | \mathbf{R} | B \rangle = \langle A | B \rangle$ or $\langle A | \mathbf{R} | B \rangle = \langle B | A \rangle$. But also an expert given P fwd and P bwd for $\langle A | B \rangle$ could assign a reasonable $K(A; B)$ for the non-categorical case that represents the multiplicative *modification* that $\langle A | B \rangle$ needs to be multiplied by to give any $\langle A | \mathbf{R} | B \rangle$. One might consider this essentially Bayesian, meaning manipulation of subjective but common sense notions about probability values. To assign objective probabilities by data mining, however, requires specific equations.

h-Complex Mutual Information:

When we consider $e^{hI(A; B)}$ or $e^{iI(A; B)}$ one could speak of $hI(A; B)$ and $iI(A; B)$ as being complex-valued mutual information. That seems inappropriate when the $I(A; B)$ is considered real. However, there is not necessarily such clear distinction, *a priori*. When we write $e^{hI(A; B)}$, it can be expressed as a real value multiple of $e^{\iota I(A; B)}$. Information $I_{1+hI}(A; B)$ has the following meaning. It has a complex “base” in the zeta function parameter $s = 1 + hI$ such that

$$I_{\sigma+hI}(A; B) = \zeta(s = \sigma+hI, o[A, B]) - \zeta(s = \sigma+hI, e[A, B]) = \iota * \zeta(s = \sigma+t, o[A, B]) + \iota \zeta(s = \sigma-t, o[A, B]) - \iota * \zeta(s = 1+t, e[A, B]) - \iota \zeta(s = \sigma-t, e[A, B]), \text{ so}$$

$$I_{\sigma+hI}(A; B) = \iota * I_{\sigma+t}(A; B) + \iota I_{\sigma-t}(A; B) \quad (14)$$

For $I_h(A; B) = \iota * I_1(A; B) + \iota I_{-1}(A; B)$. For increasing data, meaning increasing n in the Riemann zeta function, $\zeta(s = -1, n) \rightarrow -1/12$, so $I_{-1}(A; B) \rightarrow 0$. Then $I_h(A; B) = \iota * I(A; B)$. But by *iota algebra*, $I(A; B) = \iota * I_h(A; B)$, $\iota * I(A; B) = I_h(A; B)$, $\iota * I(A; B) = \iota * I_h(A; B)$, so $I_h(A; B) = I(A; B)$. All together, we can summarize the equalities, starting with the above just shown.

$$\begin{aligned} I_h(A; B) &= I(A; B) \\ I_{\iota}(A; B) &= \iota I(A; B) \\ I_{\iota}(A; B) &= \iota * I(A; B) \\ hI(A; B) &= \iota I(A; B) = -\iota * I(A; B) = -(\iota I(A; B))^* \end{aligned} \quad (15)$$

The second and third of these are easily shown by substituting for s ; the last begs explanation that follows from that. With $s = \iota *$ and $s = \iota$, we have $e^{\frac{1}{2}I(A; B)} e^{-\frac{1}{2}hI(A; B)} = e^{\frac{1}{2}I(A; B)}$ and $e^{\frac{1}{2}I(A; B)} e^{+\frac{1}{2}hI(A; B)} = e^{\iota * I(A; B)}$. It follows that we now have the mapping between $e^{I(A; B)}$ and $e^{+hI(A; B)}$. We multiply the former by $e^{-2I(A; B)}$. But this simply means multiplying by ι . We have the equalities $e^{+hI(A; B)} = e^{-I(A; B)} e^{2I(A; B)} = e^{-I(A; B)} \iota e^{2I(A; B)} = \iota e^{I(A; B)} = e^{+I(A; B)} e^{-2I(A; B)} = \iota * e^{-2I(A; B)} = \iota * e^{-I(A; B)}$. But we know from Eqn. 3 that $e^{hI(A; B)} = \iota * e^{-I(A; B)} + \iota e^{+I(A; B)}$. What has been picked up here is the partial coexisting eigensolutions to $e^{hI(A; B)}$. The above implies

$$\begin{aligned} e^{hI(A; B)} &= \iota * e^{-I(A; B)} + \iota e^{+I(A; B)} = \iota e^{+I(A; B)} + \iota e^{+I(A; B)} = e^{I(A; B)} + h e^{I(A; B)} \\ (e^{hI(A; B)})^* &= \iota * e^{+I(A; B)} + \iota e^{-I(A; B)} = \iota * e^{+I(A; B)} + \iota * e^{+I(A; B)} = e^{I(A; B)} - h e^{I(A; B)} \end{aligned} \quad (16)$$

It is the situation for conjugate variables, say $\langle A | B \rangle = P(A) e^{hI(A; B)} = P(B) e^{hI(A; B)}$.

Some issues of mixed i and h complex systems:

This is a theoretical aside as a discussion to show issues of consistency with QM, but it will become important as part of the method should it be shown that the *i*-complex algebra is also relevant to inference, as some workers have suggested. One should really think in terms of

the “mother system” as comprising complex terms involving hi (and actually think of i in three flavors if it were appropriate to follow Dirac’s fuller treatment). This raises some difficulties, but also some insights. First note that the Pure Phase θ is Negative Mutual Information, i.e.

$$\theta = 2\pi\mathcal{A}(A; B)/h = -I(A; B) \quad (17)$$

where $\mathcal{A}(A; B)$ is the physical action expressed in terms of A and B such as p and x or equivalent energy E and time t , h is Planck’s constant, and $I(A; B)$ implies mutual information given the uncertainty in measurement according to the uncertainty principle. The choice of $-I(A; B)$ rather than $+I(A; B)$ may appear to be for certain consistencies with text-book QM but depends on “directional” frame of conditionality reference, and on $hi = -ih$. In fact, the reason for Dirac’s recipe for observable probabilities involving ket normalization was that in the case of quantum mechanical *conjugate variables* we cannot in general normalize with respect to the bra without risking physical realism. $\theta = 2\pi\mathcal{A}(A; B)/h$ cannot be such that $P(B|A) = P(B) e^{2\pi\mathcal{A}(A; B)/h} > P(B)$, and certainly not for a general value of $P(B)$ such that $P(B) e^{2\pi\mathcal{A}(A; B)/h} > 1$, and we cannot breach the uncertainty principle by having precision such that $2\pi\mathcal{A}(A; B) \ll h$, nor have the spread of the wave function decreasing entropy with time. This is discussed further in Conclusions. Appropriate normalizations are required to resolve this. First note the tendency to relate weights in QM to the square roots of probabilities, but recall that here we are deal with bra-ket products not bras or kets alone. $c^*c\langle A|A\rangle = \langle A|c^* c|A\rangle$. Traditional QM scaling normalization factors for kets are a scalar real divisor $\sqrt{\langle A|A\rangle}$ called the norm. The norm for the ket $|B\rangle$ is $\sqrt{\langle B|B\rangle}$. It is easy to show that the norm is scalar real, and precisely $\sqrt{\langle B|B\rangle} = \sqrt{P(B)}$ if events B in $\langle B|B\rangle$ are indistinguishable but not by recurrence, and recurrent B ’s are independent and countable. Similarly $\sqrt{\langle A|A\rangle} = \sqrt{P(A)}$. The form for conjugate variables A and B

$$\langle A|B\rangle = \iota^* \sqrt{P(A)} e^{+iI(A; B)} + \iota \sqrt{P(B)} e^{-iI(A; B)} = \iota^* \sqrt{P(A)} e^{-i\theta} + \iota \sqrt{P(B)} e^{+i\theta} \quad (18)$$

must be correct according to traditional considerations of normalization because with eigenvalues $h = -1$ and $h = +1$ as it brings us to the same status as is usual in wave mechanics for *conjugate variables*, recalling that $\theta = -I(A; B)$ and is, for example, $+2\pi px/h$. $\sqrt{P(A)} e^{+iI(A; B)}$ gives us two problems to solve $\langle A|B\rangle = \sqrt{P(A)} e^{+i\theta}$ and its complex conjugate $\langle B|A\rangle = \sqrt{P(A)} e^{-i\theta}$ [18]. However it is the final observable probabilities that matter, and the above view is questionable in the larger h -complex picture even though both give the right answers. In preparing B to measure A conditional upon that, we set $P(B|A) = P(B) = 1$, and in preparing A to measure B conditional upon that, we set $P(A|B) = P(A) = 1$. The ket normalized form must thus be $\langle A|B\rangle' = \iota^* P(A|B) + \iota$ and the observable probability $P(A|B) = [\iota^* P(A) e^{iI(A; B)} + \iota][\iota^* P(A) e^{iI(A; B)} + \iota]^* = [\iota^* P(A) e^{iI(A; B)} + \iota][\iota^* + \iota P(A) e^{iI(A; B)}]$. By the h -complex algebra, this cannot be distinguished in final outcome from several other starting options for $\langle A|B\rangle'$, including the form $\sqrt{P(A)}[\iota^* e^{iI(A; B)} + \iota]$ for which multiplication by its own complex conjugate gives again $P(A|B)$. Analogous arguments apply to the bra normalization, including starting first from $(\langle A|B\rangle')^*$ followed by ket normalization if we stick to the Dirac recipe for the general, not just conjugate variable, case.

In any event, $e^{+iI(A; B)} = e^{-i(I(A; B) + n2\pi)} = e^{-i(I(A; B))} e^{in2\pi}$ for any negative, zero, or positive integers as n . Requiring $P(B) e^{-iI(A; B)} = P(B) e^{+iI(A; B)} = P(A|B)$ is correctly normalized under periodic boundary conditions of length $L = x$, we can increase or decrease $I(A; B)$ only in packets worth 2π . For conjugate variables p and x , $2\pi n = 2\pi px/h$, so $p = nh/L$, which is correct [18]. Hence we know that

$$\begin{aligned} e^{+iI(A; B)} &= [\cos(I(A; B)) + i \sin(I(A; B))] \\ e^{-iI(A; B)} &= [\cos(-I(A; B)) + i \sin(-I(A; B))] = [\cos(I(A; B)) - i \sin(I(A; B))] \end{aligned} \quad (19)$$

Electing to use $P(A)$ and $P(B)$ rather than their square roots, we obtain

$$\begin{aligned} \langle A|B\rangle &= \iota^* P(A) [\cos(I(A; B)) - i \sin(I(A; B))] + \iota P(B) [\cos(I(A; B)) + i \sin(I(A; B))] \\ &= \frac{1}{2} [P(A) + P(B)] [\cos(I(A; B)) - i \sin(I(A; B))] \\ &\quad + h \frac{1}{2} [P(B) - P(A)] [\cos(I(A; B)) + i \sin(I(A; B))] \end{aligned} \quad (20)$$

The simplest approach is to consider partial solutions with eigenvalues of $h = -1$ and $h = +1$ giving $P(A) [\cos(I(A; B)) - i \sin(I(A; B))]$ and $P(B) [\cos(I(A; B)) + i \sin(I(A; B))]$ respectively. With the square root format preferred, we require division by the norms $\sqrt{\langle A|A\rangle} = \sqrt{P(A)}$, and $\sqrt{\langle B|B\rangle} = \sqrt{P(B)}$.

$$\begin{aligned} \langle A|B\rangle &= \sqrt{P(A)} [\cos(I(A; B)) - i \sin(I(A; B))] \\ \langle B|A\rangle &= \sqrt{P(B)} [\cos(I(A; B)) + i \sin(I(A; B))] \end{aligned} \quad (21)$$

We are now required to extend beyond the specific case of conjugate variables, and some brief observations should be made which really relate to what complex conjugation means when at least two kinds of complex numbers are present. $\langle A|B\rangle = \iota^* \langle A|B\rangle_1 + \iota \langle A|B\rangle_2^*$ was an arbitrary form, but we can build from the brackets for conjugate variables $\langle A|B\rangle$ as above, now highlighted in bold to show that meaning, using $\langle A|B\rangle = \iota^* \langle A|B\rangle + \iota \langle A|B\rangle^*$. The order of j -complex conjugation operations on i and h is important. Whereas we think of the Lorentz transformation as $i \rightarrow h$, from the perspective of j this would yield transformation $j = hi \rightarrow hh = 1$, and so scalar real would result. The correct procedure given $\langle A|B\rangle = \iota^* \langle A|B\rangle + \iota \langle A|B\rangle^*$ appears to be to apply complex conjugation to i first which is really the transformation $hi \rightarrow -ih$, i.e. $j \rightarrow j^*$, then apply the transformation $ih \rightarrow -h = ji$. Not to labor these purely theoretical points, it appears that we can apply reasonable complex conjugation operations followed by applying multiplication of appropriately formed brackets by ι^* and ι , combined with bra and ket normalization. Be that as it may, it remains simpler and sufficient for present purposes to think of wave mechanics as a purely i -complex algebraic system, and apply the Lorentz transformation as $i \rightarrow h$ to it. However, it remains insightful to consider the impact of mixed complex systems when one comes to consider the relationship between mutual information and quantization. By extension of thought regarding the eigenket-eigenvalue solution $\langle A| \mathbf{R} |B\rangle = K(A; B) \langle A|B\rangle$, we can conceive of information as an operator \mathbf{I} (distinguish from \mathbf{I} the identify operator), and now consider i and h as vectors in their complex planes, i.e. the classical Argand interpretation. The fundamental difference in considering $\langle A|B\rangle =$

$[t^*P(A) + tP(B)] e^{i(A;B)}$ and comparing the wave equivalent is that $e^{i(A;B)} \rightarrow \cos(I(A;B)) + i \sin(I(A;B))$. The major distinction is that in the second and wave case we can advance or decrease $I(A; B)$ by 2π and it returns the same value for the probabilities calculated, and so $I(A; B)$ can only convey information about that the system has progressed a fraction through one wave cycle with I as an operator.

$$Ij = 2\pi n j \text{ x fraction}$$

$$nj = nhi = -nih = \text{integer } x j \text{ (quantized)} \quad (22)$$

We can think of the transformations

$$j \rightarrow i: ni = \text{integer } x i \text{ (quantized)}$$

$$j \rightarrow h: nh = \text{real } x h \text{ (unquantized)} \quad (23)$$

We write these in place of the pure phase θ , and we do not need a deeper understanding of the structure of I to use it. However there is evidence that real solutions exist that shortcut the Dirac recipe for observable probabilities like $P(x|y)$. For the analogue of the traditional treatment of the quantized case $j \rightarrow i: ni = \text{integer } x i$ we note that $i^{-i} = e^{e^{i/2}}$, so $Ii = 2\pi ni = 4n \log_e i^{-i} = -4n i \log_e i$. Here $\log_e i$ is multivalued, $\frac{1}{2}\pi i + 2\pi mi$ for integer i , but its principle value is $\frac{1}{2}\pi i$, which turns $-4n i \log_e i$ to $2\pi n$. The general form $Ii = -4n i (\frac{1}{2}\pi i + 2\pi mi) = 2\pi n(1 + 4m)$ is intriguing but basically still a statement that we can only jump the information in packets of 2π or 2π x fraction

h implies Classical Probabilistic Behavior:

$\langle A|B \rangle$ for $j \rightarrow h$ obeys classical probability laws of addition but also of multiplication. For example the chain rule

$$[t^*P(A|B) + tP(B|A)] x [t^*P(B|C) + tP(C|B)] = [t^*P(A|B)P(B|C) + tP(C|B)P(B|A)] \quad (24)$$

and the exact calculation by the classical law of composition of probabilities representing the sum over all possible arguments X.

$$\langle A|B \rangle = \sum_X [t^*P(A|X) + tP(B|X)] [t^*P(X|B) + tP(X|C)] = t^*P(A|B) + tP(B|A) \quad (25)$$

It now also holds as the law of composition of probability amplitudes as in wave mechanics

$$\langle A|B \rangle = \sum_X \langle A|X \rangle \langle X|B \rangle \quad (26)$$

Orthogonal Vectors:

QM makes great use of these as a simple and tractable case, and it may serve to illustrate how in input the user can program the required functionalities. The difference here is that for classical and semantic purposes we define them as **h**-complex.

$$|B \rangle = [t^*, t P(B)]^T$$

$$\langle A | = |A \rangle^* = [t^*, t P(A)]^* = [t, t^* P(A)] \quad (27)$$

The importance of these is that we can think of all nodes in the network with self probabilities $P(A)$, $P(B)$ etc. as prior probabilities expressible as these vectors, and associate the relators with mutual information from data mining. However it is often useful to think of some or all as relatively fixed and change the mutual information instead. Typically, leaf nodes in the network are considered priors, though with

a bidirectional network the distinction is not so meaningful. We may change any self probabilities. Usually we envisage input and output as special brackets of observation that relate to the vectors, as e.g. $\langle ?|A \rangle = [t^* + t P(A)]$, where $?$ indicates the fact of an observation as occurring with probability one. The simple example $\langle ?|A \rangle \langle A|B \rangle \langle B|C \rangle \langle C|? \rangle$ shows that there is no sense in which the preparer of input and the measurer of results can be distinguished. Note that the above illustrates a cyclic path (see Conclusions). To define non-orthogonal vectors, it suffices in input to have e.g.

$$|B \rangle = (1,0), (0, 0.7)$$

$$\langle A | = (1,0), (0, 0.6)$$

Here (1,0) is P fwd and P wd setting the value that implies t , and (0, 0.7) similarly setting the value that define $t^*P(B)$, and (0,0.6) similarly for $t^*P(A)$. We do this for every node A, B, C, in the network. But in Eq. 27 every node is then *orthogonal* to every other node: $\langle A|B \rangle = 0$. Whilst all those brackets and bra-relator-kets that we have not specified imply that they are there with probability one, all those that are expressly defined make, in effect, the assumption that all nodes are *mutually exclusive*, until corrected otherwise by the addition of a relator. In many cases that will stand as correct. $\langle \text{dogs} | \text{cats} \rangle$ interpreted categorically still means the same thing with same probabilities if $\langle \text{cats} | \text{be} | \text{dogs} \rangle$, which evidently should have a zero value for both P fwd and P bwd. We could indicate so specifically by $\langle \text{cats} | \text{be} | \text{dogs} \rangle = (0,0)$, but having the template $\langle \$A | \text{be} | \$B \rangle = \langle \$B | \$A \rangle$ present takes care of that. By introducing $\langle \text{dogs} | \text{chase} | \text{cats} \rangle$ in input, the value that implied by its absence, $\langle \text{dogs} | \text{chase} | \text{cats} \rangle = 1$ now jumps to a particular **h**-complex value. Non-categorical relators in general have the effect of removing the categorical meaning, and specifying some other probabilistic relationship. We should more specifically, at least in algebra if not input, write $R(A, B)$ for a relator to indicate that it is specifically a matrix relating to A and B which does that, and indeed for categorical relators too.

Operators acting on Orthogonal Vectors:

Recall that $[a, b; c, d] [p, q]^T = [ap+bq, cp+dq]^T$ is the product of a matrix with a column vector, and $[p, q] [a, b; c, d] = [pa+qc, pb+qd]$ is the product of a row vector with a matrix. They are defined in the following non-trivially Hermitian form, and the actions on bra and ket are respectively as follows.

$$R(A,B) |B \rangle = [0, t^* + t e^{i(A;B)}; t^* e^{i(A;B)+t}, 0] [t^*, t P(B)]^T = [t P(B) e^{i(A;B)}, t^* e^{i(A;B)}]^T \quad (28)$$

$$\langle A | R(A,B) |B \rangle = [t, t^* P(A)] [t P(B) e^{i(A;B)}, t^* e^{i(A;B)}]^T = t P(B) e^{i(A;B)} + t^* P(A) e^{i(A;B)}$$

$$= t^* P(A|B) + t P(B|A) \quad (29)$$

We may now consider the notion of ‘‘casting a value’’ into the bra and ket. In the above sense, $R(A, B)$ projects into $|B \rangle$ gives $|RB \rangle = [[t P(B) e^{i(A;B)}, t^* e^{i(A;B)}]^T]$ and $R(A, B)$ projects into $\langle A |$ gives $\langle AR | = [\langle t^* P(A) e^{i(A;B)}, t e^{i(A;B)} \rangle^T]$. This concept may be tested. Recall that the $[a, b; c, d][e, f; g, h] = [ae+bg, af+bh; ce+dg, cf+dh]$ is the product of a matrix with a matrix. Here I is an identity matrix.

$$R(A, B)^2 = R(A, B) R(A, B) = [1, 0; 0, 1] e^{i(A;B)} = I e^{i(A;B)} \quad (30)$$

Here $R(A, B) \neq R(A, B)^2$, so evidently $R(A, B)$ is not a *projection matrix*, given the standard definition that a projection matrix is one which is equal to its square. And whilst forms $[0, \alpha; 0, \alpha]$ are promising as example non-Hermitian projection operators, it is readily shown that $[0, \iota^* + \iota e^{i(A;B)}; 0, \iota^* e^{i(A;B)} + \iota]^2 = [0, 1; 0, 1] e^{i(A;B)}$, so if $R(A, B)$ were defined that way, again $R(A, B) \neq R(A, B)^2$ and hence $R(A, B)$ so redefined is also not a projection operator. Nor is the result $[0, 1; 0, 1] e^{i(A;B)}$, although $[0, 1; 0, 1]$ is. However, returning to the original definition, we have so far achieved for these orthogonal vectors

$$\begin{aligned} \langle A | R(A, B) | B \rangle &= \iota^* P(A|B) + \iota P(B|A) \\ \langle A | R(A, B)^2 | B \rangle &= \langle A | I e^{i(A;B)} | B \rangle = \iota^* P(A|B) + \iota P(B|A) \quad (31) \end{aligned}$$

Hence $R(A, B) = [0, \iota^* + \iota e^{i(A;B)}; \iota^* e^{i(A;B)} + \iota, 0]$ is *effectively* a projection matrix for the form $\langle A | B \rangle \rightarrow \langle A | R(A, B) | B \rangle = \iota^* P(A|B) + \iota P(B|A)$ that really interests us. We just have to always include a $\langle A | R | B \rangle$ to stand for $\langle A | B \rangle$. To understand the process of antisymmetric projection in the absence of bra or ket normalization, we can factorize the actions of $R(A, B)$ into $R(A, B)^{1/2}$ acting on the bra and $R(A, B)^{1/2}$ acting on the ket.

$$\langle A | R(A, B)^{1/2} (R(A, B)^{1/2} | B \rangle) = \langle A | R(A, B)^{1/2} | B \rangle = \iota^* P(A|B) + \iota P(B|A) \quad (32)$$

Results

The system developed is rich in capabilities and for brevity attention will be paid here in Results and Discussion to those summary findings and insight of general importance.

To appreciate the significance of following comments, recall that for a net which is not static, the process of local optimization repeated many times in the hunt to achieve *global optimization*. The usual idea of optimal in this case relates to the notion of knowledge is what results when given data is processed such that maximal knowledge is carried in the least number of bits of information. In practice, it is the *information density* as the average information content of the rules, the total information in the net divided by the number of rules N . Like the network overall, each rule is associated with P_{fwd} and P_{bwd}, which relate to the semantic statement and its adjoint. It turns out that an equivalent semantic statement can always be written by use of negation or some qualification that replaces a probability P by $1 - P$, at least in principle. This is done if any $P < 0.5$, so information $= -\log_2 P$ lies between 0 and 1. The maximum information that a network of N rules can have is N bits in each direction (in the sense of overall P_{fwd} and overall P_{bwd}) or $2N$ considering both directions. The theoretical achievable upper limit of information density is thus 1 bit in each direction of conditionality. Recall that by information theory and Popper's argument [11] it is rules of 100% probability in a direction of conditionality that are not interesting, since the same effect would be obtained by not having them there at all. Popper's position is that asserted statements mean little unless refuted. If all probabilities are 100% in a given direction of conditionality, that information density by the above definition is 0%. Note also that the closer the two directional probabilities are to being equal, the less directionality matters, and the rules could on average be expressed

as the symmetric existential or "some" case.

Many computations to date have been concerned with the ordering of diagnostics and selection of therapy, but well illustrate the above principles. In one study regarding predictions of tuberculosis in the newborn based on suspected exposure to tuberculosis by the mother 46 rules implying multiple cyclic paths were reduced to the 23 in 30 local optimizations and no improvement was obtained in up to 2000 local optimizations. The overall forward % probability of the Net was 0.323% and the overall reverse probability was 57.684%, with real and imaginary components of 29.0 and 28.7 on a "percentage basis" (0.290 and 0.286 in actuality). The 0.323% largely, but not solely, reflects the low probability that tuberculosis is transmitted from mother to the baby in the womb (compared with say, HIV, which has a high probability). The 57.684% of are the collective causes of etiologies which *could have* caused tuberculosis in the newborn. The forward information density of 0.55 bits is considerably less than the theoretical upper limit of 1.00 nats, but the backward information density 0.053 reflects much less evidence from refutation in the Popper sense [18]. Whilst comparable studies involving pharmaceutical chemistry are at an early stage, there is a tantalizing much larger number of available rules from the reading of all US patents [1]. The 6.7 million protocols were originally of form $\langle \text{formula} | \text{assignee and patent number} \rangle$, where assignee is most often the company such as AstraZeneca, and the formula is a compound described in SMILES code [1]. The formula basically reflects the nomenclature of names advised by the International Union of Pure and Applied chemistry. The system has to recognize that the same formulae can be written in different ways, which is particularly problematic when relating parts of compounds. This latter is probabilistic in the sense of a degree of similarity between molecules with similar parts [1]. Initially we simply assigned the above formula-assignee rule as 100% true in both directions of conditionality. While containing a huge amount of information, this does not add much to *probabilistic* inference, and the following will also serve as illustrating use of P_{fwd} and P_{bwd}. The probabilistic interpretation is up the user but, as an example, we recently we employed the rule $\langle \text{formula} | \text{if} | \text{patent number} \rangle$, with a distinct P_{fwd} and P_{bwd}. Given a patent, several compounds may appear in it, and the same compounds can appear in different patents. P_{fwd} = $P(\text{formula} | \text{patent number}) = n(\text{formula}, \text{patent number}) / n(\text{patent number})$ is often less than 100% because the chance of picking one precise compound from several on a patent at random is less than 100%. Compare $P(\text{males} | \text{New Yorkers}) = n(\text{males}, \text{New Yorkers}) / n(\text{New Yorkers}) \approx 50\%$ to see the idea of this. More interesting from an intellectual property perspective is that the same molecule may appear on different patents, so that P_{bwd} = $P(\text{patent number} | \text{formula}) = n(\text{formula}, \text{patent number}) / n(\text{formula})$, which is certainly less than 100%, and more interesting still that P_{bwd} = $P(\text{assignee} | \text{formula})$ which is less than 100%. The interpretation could be that given a molecule, assignees do not really have 100% clear ownership. However, most often it is likely that prior art may have been quoted, or a known compound used in a synthetic process, or new use is being patented for a compound which is not a novel composition of matter.

This area is nonetheless controversial because it is a good way of detecting controversial issues. It remains, however, a matter of data mining until several rules are combined to make a network to be used in inference. At present the networks for chemical compounds are very small, and therefore more in the nature of queries.

Discussion

General theoretical findings:

A referee raised the issue of how the indeterminate aspects of QM theory could be applied in this area, noting how this aspect is being extensively utilized in quantum computing, and wondering how this powerful aspect of QM theory might be applied to probabilistic semantics. Quantum indeterminacy is the seemingly necessary incompleteness in the QM description of a physical system. It is also true that, along with the related idea of fundamental uncertainty, this feature of QM has attracted attention in regard to uncertainty, fuzziness, and sometimes unexplained leaps of insight in human language and thought, albeit it usually approached rather symbolically rather than through use of a system of complex algebra. In large part the issue of indeterminacy relates to what can be characterized by a *probability distribution* on the set of measurement outcomes of an observable. Probability distributions are non-discrete vectors and can be represented finitely by vectors, while matrices describe the dependencies between them. We have built in vectors and matrices from the outset precisely to accommodate probability distributions in future, and not least because they are of course no less important in classical data analytics. We could for example use a density function to express Bayesian degrees of belief for different values of an observed quantity, or indeed of an information value or classical probability, given say a binomial distribution as a likelihood. It can of course imply a scalar value as an expected or average value derived from distributions: expected information in terms of zeta functions rose formally from this idea, but preserving the original distributions allows averaging of other (perhaps as yet unforeseen) measures, and permits other statistical summaries such as maximum likelihood (as opposed to expectation). It seems intuitively obvious that we could introduce appropriate indeterminacy in this way.

Highly relevant to the above is that the $i \rightarrow h$ transformation does not get rid of distributions implied in QM nor does it get rid of Planck's constant and the uncertainty principle, and yet the interpretation becomes classical. The distribution however changes, and in a very useful way. Consider the particle on a circular orbit of length L . Proceeding as described in Methods, $\langle x|p \rangle = P(A)\exp(I(A;B)) = L^{-1/2}\exp(2\pi n h p x/h) = \iota^* L^{-1/2}\exp(-2\pi n i p x/h) + \iota L^{-1/2}\exp(2\pi n i x/L)$. Consequently

$$P(x|p) = \langle x|p \rangle' (\langle x|p \rangle')^* = L^{-1}\exp(2\pi n p x/h) = L^{-1}\exp(-2\pi n m x^2/ht) \quad (33)$$

It is a *particle function* as a *Gaussian function* of x spreading with time, although that will only be apparent for everyday objects of large mass m over cosmological time t because of the small value of h . It will be classically interpretable as increasing error with which we can interpret the position of the particle, and increasing entropy along with that increasing lack of knowledge. Planck's constant merely sets the least possible error on our measurements, which for any

realistic apparatus will be much larger. "Measurements" really include perturbing interactions with other objects and fields that do not involve human observers but in some way mimic the interaction implied in true observation. Whatever the meaning of " n " now, say such that the mass of the object $M = nm$, we should presumably stick with that value based on initial state, assuming observations or other interactions do not modify it. *Between* observations on an entity as an object rather than a wave, QM teaches that $\langle x|p \rangle$ switches to the wave function, but this is in practice indistinguishable from the uncertainty arising from our observations and other interactions. These aspects have been discussed elsewhere (e.g. Ref. [21]), but there are slight differences and there is no inclusion of n in those references. The appearance of the Gaussian function is a blessing for everyday practical use because we can apply transformations to express uncertainties in data and distributions in populations that follow the ubiquitous normal distribution.

This raises the question of what a distribution can mean for a state A when it is categorical data: say the observation that a patient is male. We can certainly think of an error in making that assessment of being male, but there is a more fundamental analogy with QM. The point is to see it in conditional probability terms. When we write $\langle A|B \rangle$ to embody $P(A|B)$ and $P(B|A)$, it resembles the QM particle seen as being in *oscillation* between states A and B , at any moment of time being in these with a certain weight, $P(A)$ and $P(B)$. An obvious choice for equations like Eqn. 33 is to consider the ground state. This has the interesting analogies with setting up the i -complex wave equations for a harmonic oscillator [18], in which the particle is seen as oscillating around its mean position. The ground state is then a *Gaussian function*, i.e. a result similar to that obtained using h . It appears plausible that applying this "oscillation interpretation" analogy more generally, including bias to generate skewed Gaussian functions, no transformation from i -complex wave mechanics may be required, but rather classical inference becomes wave mechanics in its ground state.

Cyclic Paths:

As noted above, being algebraic-complex, $\langle A|B \rangle \langle B|C \rangle \langle C|D \rangle$ encodes both directions of conditionality and the Dirac Net is thus *bidirectional*. A consequence of bi-directionality is that the network can be a general graph allowing cyclic paths. The emergent property of a cyclic path such as $\langle A|B \rangle \langle B|C \rangle \langle C|A \rangle$ is that its value is scalar real, as can be shown algebraically. This would not of course be seen if two separate Bayes Nets were used to encode each direction of conditionality. In consequence of the real value, the notion of events ultimately affecting their own cause that led to the restrictions of a Bayes Net traditionally defined as an acyclic directed graph [10] do not apply. A Dirac Net does not require iteration for solution.

Work in progress:

An assembly of rules linking common drug names to formulae of compounds is in hand. Also by reading patents, a growing number links the compounds to protein targets, disease targets, species (typically but not always humans) in which results are obtained, and relevant biochemical and methodological details. Of considerable interest is the deduction of related formulae that may have intended or other biological actions, which until experimentally qualified, are inherently probabilistic. Novel formulae can be automatically generated by evolving under the

“natural selection” of not being in the scope of those already in patents, but binding appropriately in simulation to the protein targets. In addition, they may satisfy other requirements such as stability, relative ease of synthesis, and implications of minor toxicity effects by common use of related chemistries [1]. Probabilistic inference from these kind of rules will be described elsewhere (See also Refs. [20,21]).

Conclusions

Networks based on quantitative QM concepts, and on Dirac algebra and notation in particular, appear to be very promising as a basis for probabilistic semantics. The issue of what is Best Practice in the specific implementation here is inevitably still open to argument on a detailed basis, as it is an early work in progress. Nonetheless it is felt that the overall strategic position that a correct rendering of QM is Best Practice is persuasive. We have not here discussed in depth the important topic of negation and the significance of definite and indefinite article, and more complex statements with more than one relator. These and other linguistic aspects, as well as Artificial Intelligence issues relating mapping of language to conceptual spaces of thought, will be discussed elsewhere.

Competing interests

The author has no competing interest.

Article history

Received: 29-Jan-2012 Accepted: 22-Mar-2012

Published: 27-Mar-2012

References

1. Robson B, Dettinger R, Peters A, Boyer SKP: **Drug discovery using very large numbers of patents: general strategy with extensive use of match and edit operations.** *J. Computer Aided Molecular Design* 2011, **25(5)**:427-41
2. Mullins IM, Siadaty MS, Lyman J, Scully K, Garrett CT, Miller WG, Robson B, Apte C, Weiss, S., Rigoutsos I, Platt D, Cohen S, Knaus WA: **Data mining and clinical data repositories: Insights from a 667,000 patient data set.** *Computers in Biology and Medicine* 2006, **36(12)**:1351-77
3. Robson B, Garnier J: **Introduction to Protein and Protein Engineering.** 1984, 1988, Elsevier Press.
4. Robson B, Baek OK: **The Engines of Hippocrates. From the Dawn of Medicine to Medical and Pharmaceutical informatics.** 2009, Wiley.
5. Robson B: **Clinical and pharmacogenomic data mining: 1. Generalized theory of expected information and application to the development of tools.** *J. Proteome Res* 2003; **2**;(3.);283-302
6. Robson B: **Clinical and Pharmacogenomic Data Mining: 3. Zeta Theory As a General Tactic for Clinical Bioinformatics.** *J. Proteome Res. (Am. Chem. Soc.)* 2005, **4(2)**:445-455
7. Robson B: **Clinical and Pharmacogenomic Data Mining: 4. The FANO Program and Command Set as an Example of Tools for Biomedical Discovery and Evidence Based Medicine.** *J. Proteome Res. (Am. Chem. Soc.),* **7(9)**:3922–3947
8. Buchanan BG, Shortliffe EH: **Rule Based Expert Systems. The Mycin Experiments of the Stanford Heuristic Programming Project** 1982, Addison-Wesley: Reading, Massachusetts.
9. Robson B: **The New Physician as Unwitting Quantum Mechanic: Is Adapting Dirac's Inference System Best Practice for Personalized Medicine, Genomics and Proteomics?** *J. Proteome Res. (Am. Chem. Soc.)* 2007, **6(8)**: 3114 – 3126
10. Pearl J, Russell S: **Bayesian Networks. Handbook of Brain Theory and Neural Networks;** M. A. Arbib, Ed. 2003, MIT Pres, Cambridge: 157-160
11. Popper K: **The Logic of Scientific Discovery,** (as *Logik der Forschung*, 1934), English translation (1959), Routledge).
12. Klopotek MA: **Cyclic Bayesian Network – Markov Process Approach,** *Studia Informatica* 2006, **1/2(7)** Systemy i Technologie Informacyjne http://www.studiainformatica.ii.ap.siedlce.pl/download/51.pdf
13. Prediou L, Tuckenschmidt H: **Probabilistic Models for the SW – A Survey.** http://ki.informatik.uni-mannheim.de/fileadmin/publication/Prediou08Survey.pdf 2009, (last accessed 4/29/2010).
14. Dirac PAM: **The Principles of Quantum Mechanics.** 1930, Oxford University Press, Oxford.
15. Penrose R: **The Road to Reality. A Complete Guide to the Laws of the Universe.** 2004, Jonathan Cape, Random House, London .
16. Ogden CK, Basic English: A General Introduction with Rules and Grammar. 1930,1940, Paul Treber & Co.,Ltd, London.
17. J. Cockle, **A New Imaginary in Algebra.**1848 *London-Edinburgh-Dublin Philosophical Magazine*, **3(33)**: 345-349
18. Chester M: **Primer of Quantum mechanics.** 1987, John Wiley and Sons, New Jersey (2003, Dover Publications)
19. Fano RM: **Transmission of Information: A Statistical Theory of Communications.** (1961), MIT Press, Cambridge.
20. Robson B, Balis UGJ, Caruso, TP: **Considerations for a Universal Exchange Langue in Healthcare.** 2011, *e-Health Networking Applications and Services (Healthcom)*, 173-176
21. Robson B: **Towards intelligent Internet-roaming agents for mining and inference from medical data.** *Stud Health Technol Inform* 2009; **149**:(157-77

Citation:

Robson B: **Towards Automated Reasoning for Drug Discovery and Pharmaceutical Business Intelligence.** *Journal of Pharmaceutical Technology and Drug Research* 2012, **1**:3. http://dx.doi.org/10.7243/2050-120X-1-3