A quantum framework for likelihood ratios

Rachael Bond

December 12th, 2015

University of Sussex

The annual scientific meeting of the
Mathematical, Statistical, & Computing Psychology Section
of the British Psychological Society

r.l.bond@sussex.ac.uk
@rachael_bond
www.rachaelbond.com
rlb.me/pdf1215

University of Sussex

Contents

1. Pseudodiagnosticity

Pseudo-diagnosticity

Doherty, Mynatt, Tweney, & Schiavo [1]

"An undersea explorer has found a pot with a square base that has been made from smooth clay.

Using the information below, you must decide from which of two nearby islands it came. You may select one more piece of information to help you make your decision."

Pseudo-diagnosticity

	Shell Is.	Coral Is.
# Finds	10	10
% Smooth	80	?
% Sq. base	?	?

Doherty, Mynatt, Tweney, & Schiavo [1]

Pseudo-diagnosticity

	Shell Is.	Coral Is.
# Finds	10	10
% Smooth	80	✓
% Sq. base	✗	✗

Doherty, Mynatt, Tweney, & Schiavo [1]

Doherty et al. expected their participants to select the paired datum to the given “anchor information” in order to calculate a Bayes' ratio. The majority didn't.

“Pseudodiagnosticity is clearly disfunctional.”

~ Doherty, Mynatt, Tweney, & Schiavo (1979) [1], p. 121

What if all the data are known?

	Shell Is. (H_1)	Coral Is. (H_2)
# Finds (*Base rate*)	10	10
# Smooth clay (D_1)	8	7
# Square base (D_2)	6	5

What if all the data are known?

	H_1	H_2
Base	10	10
D_1	8	7
D_2	6	5

To calculate the value $P(H_1)$ using Bayes' theorem, this expression must be solved

$$P(H_1|D_1 \cap D_2) = \frac{P(H_1)P(D_1 \cap D_2|H_1)}{P(H_1)P(D_1 \cap D_2|H_1) + P(H_2)P(D_1 \cap D_2|H_2)}$$

However, the measures of covariate intersection, ie. $P(D_1 \cap D_2|H_x)$, are unknowns.

What if all the data are known?

	H_1	H_2
Base	10	10
D_1	8	7
D_2	6	5

Doherty et al. suggest that the data should be treated as conditionally independent. This allows for a simple estimation of $P(H_1)$ from the multiplication of marginal probabilities

$$P(H_1|D_1 \cap D_2) = \frac{0.5\times0.8\times0.6}{(0.5\times0.8\times0.6)+(0.5\times0.7\times0.5)}$$

$$\approx 0.578$$

What if all the data are known?

However, it would also be reasonable to note that the covariate intersections form ranges:

$$n(D_1 \cap D_2|H_i) \in \begin{cases} \Big[n(D_1|H_i) + n(D_2|H_i) - n(H_i) \ , \ldots , \min(n(D_1|H_i), n(D_2|H_i)) \Big] \\ \qquad \text{if } \ n(D_1|H_i) + n(D_2|H_i) > n(H_i) \ , \qquad \text{or} \\ \Big[0 \ , \ldots , \min(n(D_1|H_i), n(D_2|H_i)) \Big] \\ \qquad \text{if } \ n(D_1|H_i) + n(D_2|H_i) \leq n(H_i) \ . \end{cases}$$

ie.,

$$n(D_1 \cap D_2|H_1) \in \{4, \ 5, \ 6\} \ ,$$

$$n(D_1 \cap D_2|H_2) \in \{2, \ 3, \ 4, \ 5\} \ .$$

What if all the data are known?

	H_1	H_2
Base	10	10
D_1	8	7
D_2	6	5

This means that it is also possible to calculate a probability from the mean value of these ranges:

$$P(\mu[n(D_1 \cap D_2|H_1)]) = \frac{1}{10} \times \frac{1}{3}(4+5+6) = 0.5 \ ,$$

$$P(\mu[n(D_1 \cap D_2|H_2)]) = \frac{1}{10} \times \frac{1}{4}(2+3+4+5) = 0.35$$

$$\Rightarrow \quad P(H_1|\mu D_1 \cap D_2) \approx 0.588$$

What if all the data are known?

	H_1	H_2
Base	10	10
D_1	8	7
D_2	6	5

Or, to take the mean value of the minimum→maximum probability range:

$$\min P(H_1|D_1 \cap D_2) = \frac{4}{4+5} \ ,$$

$$\max P(H_1|D_1 \cap D_2) = \frac{6}{6+2}$$

$$\Rightarrow \ \mu[P(H_1|D_1 \cap D_2)] \approx 0.597 \ .$$

What if all the data are known?

	H_1	H_2
Base	10	10
D_1	8	7
D_2	6	5

Other possible approaches include regression analysis, which would assume a low level of co-linearity, or using an expectation-maximisation algorithm (eg., see Dempster, Laird, & Rubin, 1977) [2]

2. Is probability subjective?

Is probability subjective?

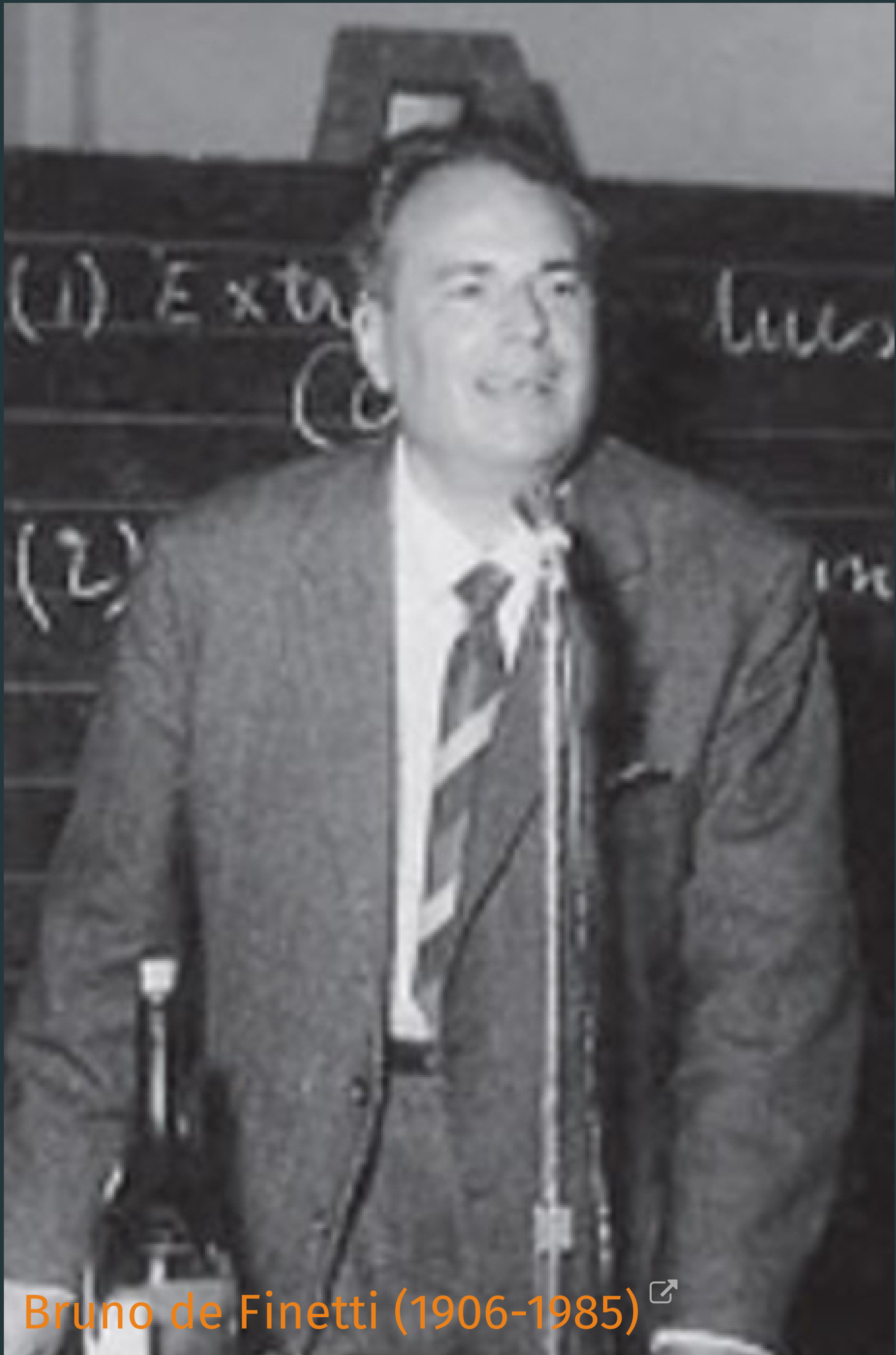
Given the variety of probability values which may be reasonably calculated, one may conclude that there is no objectively correct likelihood ratio.

Is probability subjective?

Given the variety of probability values which may be reasonably calculated, one may conclude that there is no objectively correct likelihood ratio.

The subjective nature of probability has moved to the centre of statistical research since Bruno de Finetti claimed that “probability does not exist”.

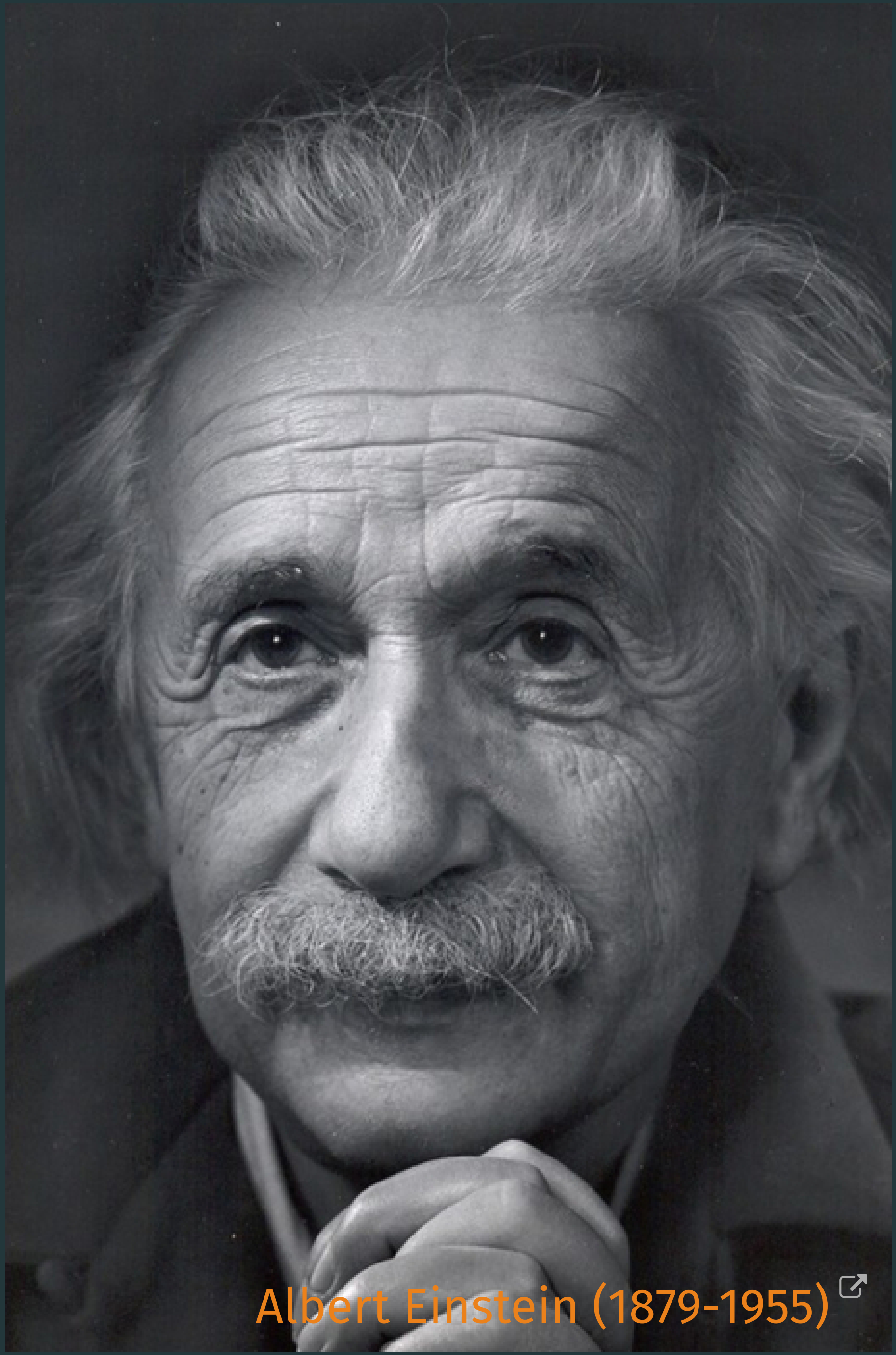
(de Finetti, 1974) [3]



Bruno de Finetti (1906-1985)

de Finetti's subjective view of probability may be found in epistemological research, and modern statistics, eg., the “quantum Bayesian” work of Caves, Fuchs, & Schack (2002) [4]

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.”
(Geometry & Experience, 1921)



3. Describing an objective reality

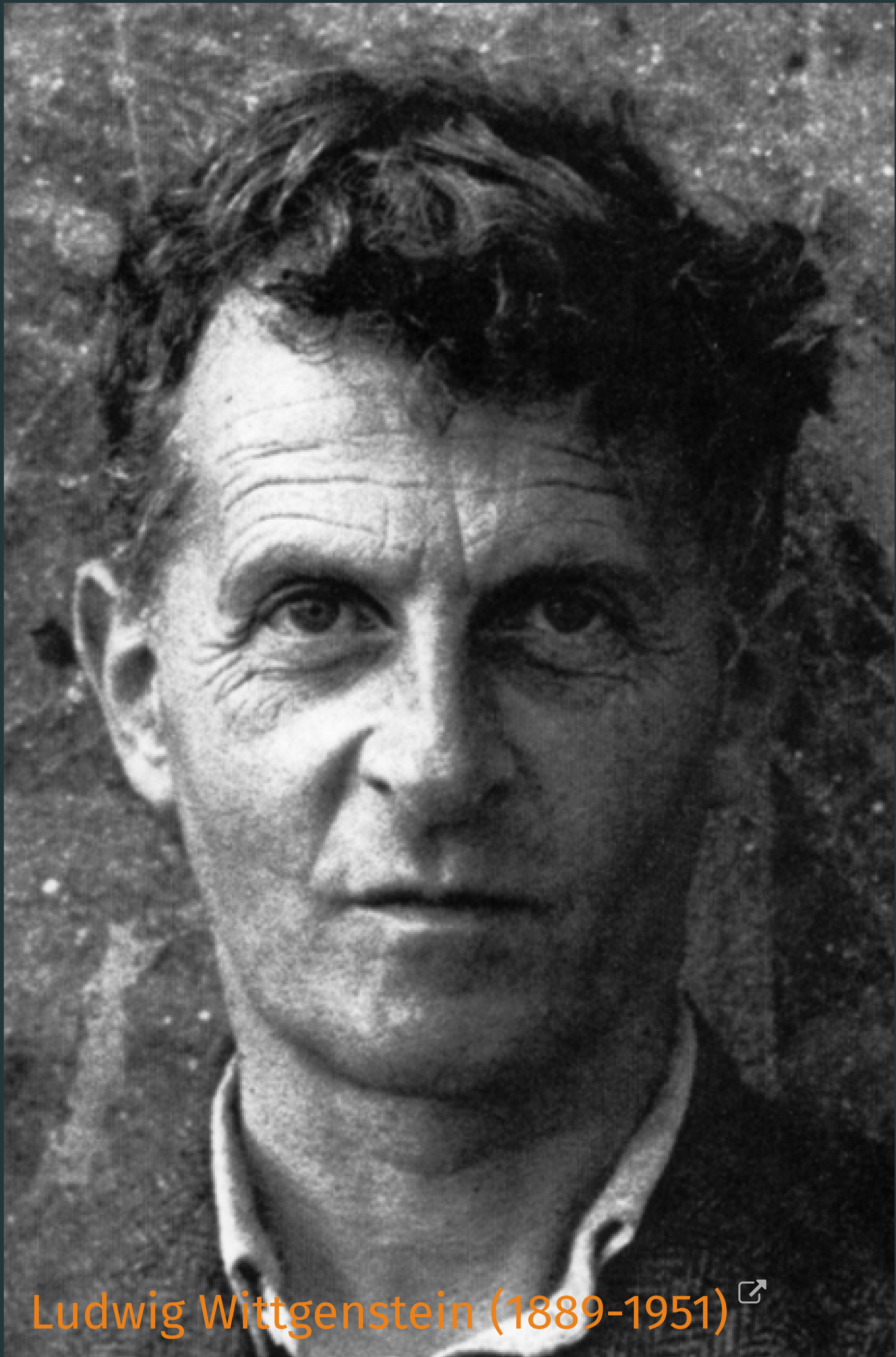
Describing an objective reality

Aristotle (384-322 BCE) argued that "reality" is described by the unity of form and substance: "substance" being what something is made from, and "form" being its innate characteristics.

Describing an objective reality

Aristotle (384-322 BCE) argued that "reality" is described by the unity of form and substance: "substance" being what something is made from, and "form" being its innate characteristics.

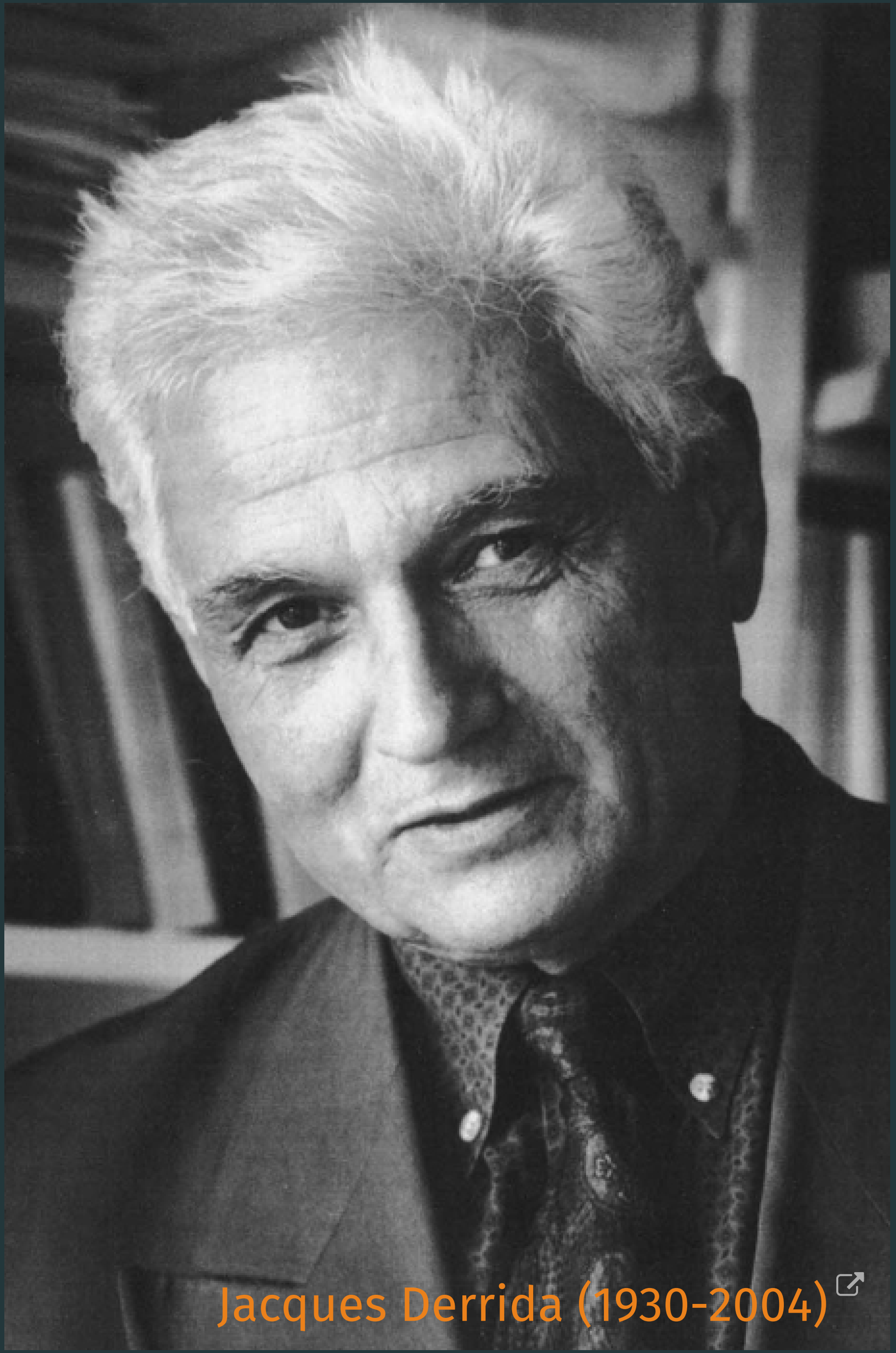
In the contingency table, the "substances" (ie., the differentiating characteristics), and their "forms" (ie., their values), are known. Yet an objective probability value cannot be calculated from this description of the table's reality.



Ludwig Wittgenstein (1889-1951)

In the “Tractatus” (1922) Wittgenstein said that “the world is the totality of facts”, and that “ it is the relationship between facts and there being all the facts”.

Jacques Derrida believed that the relationships between facts can only be discovered through a process of "deconstruction".



4. Deconstructing the contingency table

Deconstructing the contingency table

Assuming, for the moment, the case of even base rates, the contingency table may be deconstructed into 4 sub-contingency tables ...

	H_1	H_2
D_1	8	7
D_2	6	5

	H_1
D_1	8

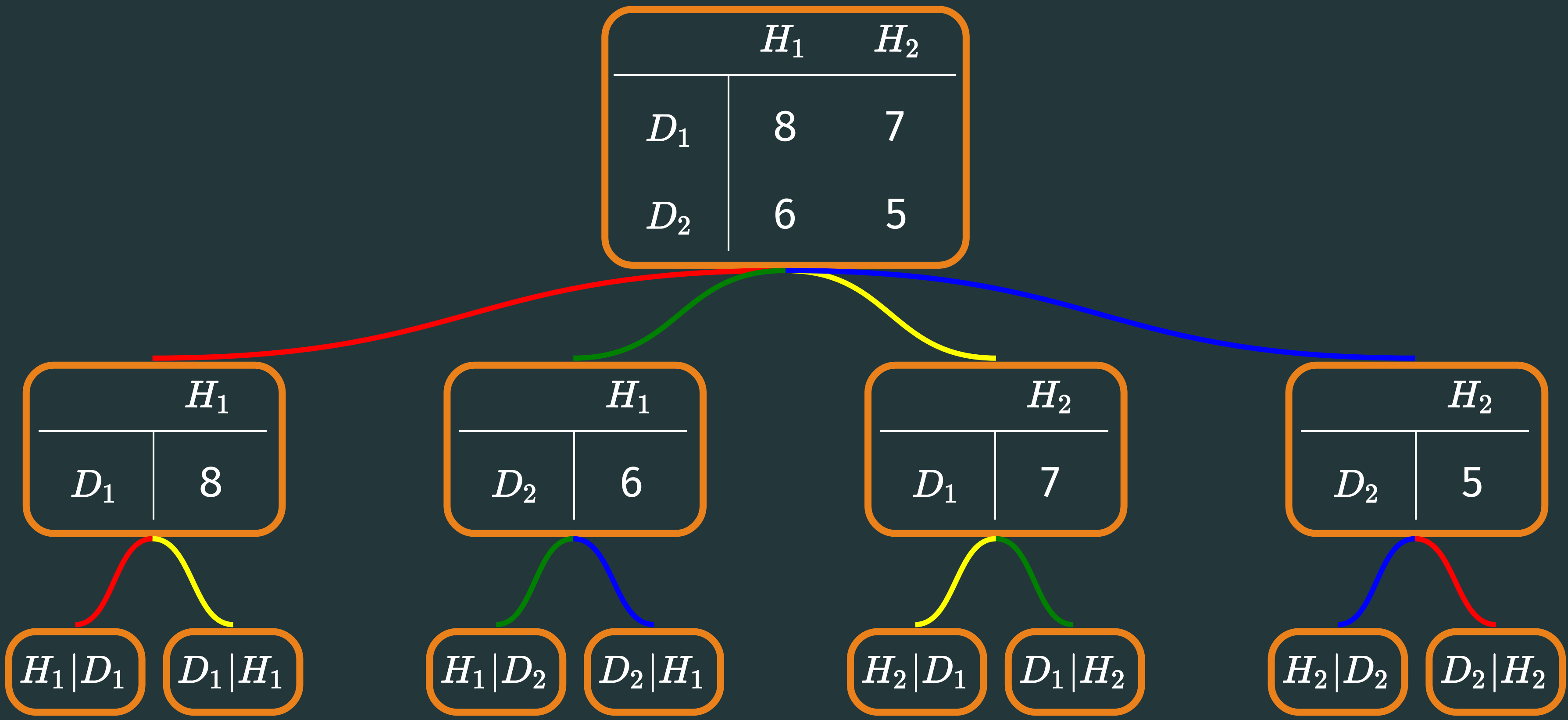
	H_1
D_2	6

	H_2
D_1	7

	H_2
D_2	5

Deconstructing the contingency table

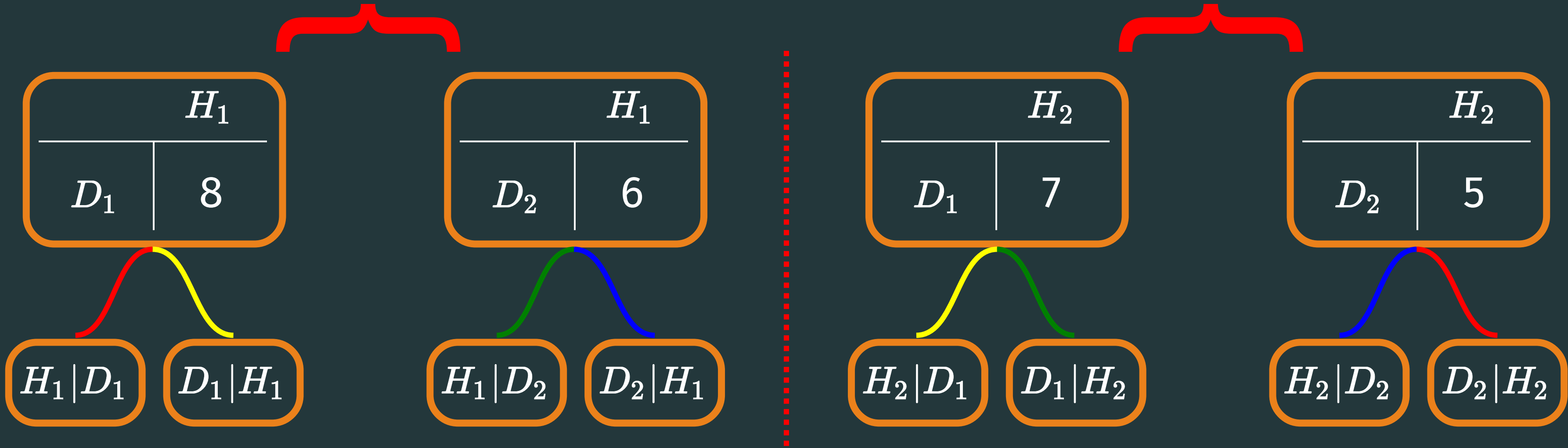
... each of which provides two pieces of “pure” information generated from the facts of H_x and D_x. These are not logically separable.



Deconstructing the contingency table

While the relationships between $D_x|H_1$ and $D_x|H_2$ are known (they are mutually exclusive), the relationships between $D_1|H_x$ and $D_2|H_x$ cannot be stated.

$D_1|H_1 \cap D_2|H_1 = ?$ $D_1|H_2 \cap D_2|H_2 = ?$



Deconstructing the contingency table

What is needed is a mathematical approach which allows the covariate intersections to be directly mapped to $D_1|H_x$ and $D_2|H_x$.

Deconstructing the contingency table

What is needed is a mathematical approach which allows the covariate intersections to be directly mapped to $D_1|H_x$ and $D_2|H_x$.

In other words, the contingency table's internal relationships must be rewritten in a way that includes the covariate intersections, but does not make any structural changes. This can only be achieved by using the mathematics of quantum mechanics.

5. Quantum mechanics 101

There are many competing models of quantum mechanics.

Multiverse theory ✗

String theory ✗

Decoherence theory ✗

The Copenhagen interpretation ✓

Niels Bohr (1885-1962)

In 1935 Niels Bohr suggested that psychology & quantum mechanics might be linked, but it is only recently that research has been conducted in this field.

Quantum mechanics 101

Instead of the joint probability spaces used in classical statistics, quantum mechanics works in vector spaces.

Quantum mechanics 101

Instead of the joint probability spaces used in classical statistics, quantum mechanics works in vector spaces.

The vectors are normalised wave functions which are orthogonal to each other in n-dimensions.

Quantum mechanics 101

Instead of the joint probability spaces used in classical statistics, quantum mechanics works in vector spaces.

The vectors are normalised wave functions which are orthogonal to each other in n-dimensions.

In psychology these vectors could, for instance, represent attitudes, beliefs, or intent etc.

Quantum mechanics 101

Using the Dirac (1939) [5] “bra-ket” notation, the wave functions are described by horizontal matrices known as “kets”, written as $|\Psi\rangle$

Quantum mechanics 101

Using the Dirac (1939) [5] “bra-ket” notation, the wave functions are described by horizontal matrices known as “kets”, written as $|\Psi\rangle$

Their “complex conjugate transposes” form vertical matrix “bras”, written as $\langle\Psi|$

Quantum mechanics 101

Using the Dirac (1939) [5] “bra-ket” notation, the wave functions are described by horizontal matrices known as “kets”, written as $|\Psi\rangle$

Their “complex conjugate transposes” form vertical matrix “bras”, written as $\langle\Psi|$

Any ket multiplied by its own bra is “orthonormal”, meaning that

$$\langle\Psi|\Psi\rangle = 1$$

6. Describing the wave function

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$|H_1 \otimes D_1\rangle$$
$$|H_2 \otimes D_1\rangle$$
$$|H_1 \otimes D_2\rangle$$
$$|H_2 \otimes D_2\rangle$$

The four pieces of "pure" information may be written as kets. The tensor product acts as a logical "AND", re-enforcing the inseparability of H_x and D_x.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_1 \rangle = 1$$
$$\langle H_2 \otimes D_1 | H_2 \otimes D_1 \rangle = 1$$
$$\langle H_1 \otimes D_2 | H_1 \otimes D_2 \rangle = 1$$
$$\langle H_2 \otimes D_2 | H_2 \otimes D_2 \rangle = 1$$

Each of the kets is automatically orthonormal and forms an eigenstate basis of a Hilbert (vector) space.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

It is tempting to describe the covariate intersection as being the simple entanglement of $D_1|H_x$ and $D_2|H_x$. However, this would give an expression which would mix the whole of $D_1|H_x$ *and* the whole of $D_2|H_x$.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c$$

Instead we need to look at the “inner products”, which are usually interpreted as giving the probability amplitude of a ket collapsing into a bra.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c$$

The bra $|H_1 \otimes D_2\rangle$ can only collapse into the ket $\langle H_1 \otimes D_1|$ if the inner product contains both $D_1|H_1$ and $D_2|H_1$. As a consequence, the inner product is a measure of covariate overlap.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c$$
$$\langle H_1 \otimes D_2 | H_1 \otimes D_1 \rangle = c^*$$

The reverse, complex conjugate transposed, inner product is also true.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c$$

$$\langle H_1 \otimes D_2 | H_1 \otimes D_1 \rangle = c^*$$

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = \langle H_1 \otimes D_2 | H_1 \otimes D_1 \rangle$$

Because both inner products are real, and consistent with the conditional independence of H_1 and H_2, it follows that they also equal to each other.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_1 \rangle = \langle H_1 \otimes D_2 | H_1 \otimes D_2 \rangle = 1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_1 \rangle = \langle H_2 \otimes D_2 | H_2 \otimes D_2 \rangle = 1$$

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c_1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_2 1 \rangle = c_2$$

all other bra-kets $= 0$

Thus, the complete quantum contingency table consists of 4 orthonormal kets, and 2 inner products. It exactly matches the classical description.

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_1 \rangle = \langle H_1 \otimes D_2 | H_1 \otimes D_2 \rangle = 1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_1 \rangle = \langle H_2 \otimes D_2 | H_2 \otimes D_2 \rangle = 1$$

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c_1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_2 1 \rangle = c_2$$

$$\textit{all other bra-kets} = 0$$

To provide a full Hilbert space description, the inner products must be mapped to (ie., incorporated into) the base kets. This may be achieved using the Gram-Schmidt process (see, eg., Strang, 1980) [6].

Describing the wave function

	H_1	H_2
D_1	8	7
D_2	6	5

$$\langle H_1 \otimes D_1 | H_1 \otimes D_1 \rangle = \langle H_1 \otimes D_2 | H_1 \otimes D_2 \rangle = 1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_1 \rangle = \langle H_2 \otimes D_2 | H_2 \otimes D_2 \rangle = 1$$

$$\langle H_1 \otimes D_1 | H_1 \otimes D_2 \rangle = c_1$$

$$\langle H_2 \otimes D_1 | H_2 \otimes D_2 1 \rangle = c_2$$

$$\textit{all other bra-kets} = 0$$

The Gram-Schmidt process orthonormalizes the base kets with respect to the inner product, and acts as a unitary operator to generate a new isomorphic representation of the original Hilbert space.

Describing the wave function

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$|\Psi\rangle = \frac{\sqrt{x_1}|H_1 \otimes D_1\rangle + \sqrt{y_1}|H_1 \otimes D_2\rangle + \sqrt{x_2}|H_2 \otimes D_1\rangle + \sqrt{y_2}|H_2 \otimes D_2\rangle}{\sqrt{x_1 + x_2 + y_1 + y_2 + 2c_1\sqrt{x_1 y_1} + 2c_2\sqrt{x_2 y_2}}}$$

In doing so, it returns four base kets that give a full system description and includes the inner products. This allows the fully normalized system wave function to be described.

Describing the wave function

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$|\Psi\rangle = \frac{\sqrt{x_1}|H_1\otimes D_1\rangle+\sqrt{y_1}|H_1\otimes D_2\rangle+\sqrt{x_2}|H_2\otimes D_1\rangle+\sqrt{y_2}|H_2\otimes D_2\rangle}{\sqrt{x_1+x_2+y_1+y_2+2c_1\sqrt{x_1y_1}+2c_2\sqrt{x_2y_2}}}$$

$$P(H_1|D_1\cap D_2) = \frac{x_1+y+1+2c_1\sqrt{x_1y_1}}{x_1+x_2+y_1+y_2+2c_1\sqrt{x_1y_1}+2c_2\sqrt{x_2y_2}}$$

The correct expression for $P(H_1|D_1\cap D_2)$ may be found through rearrangement.

Describing the wave function

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$P(H_\alpha | D_1 \cap D_2 \ldots \cap D_m) = \frac{\sum_{i,j} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^{\alpha}}{\sum_{i,j,\alpha} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^{\alpha}}$$

This expression fully generalizes, and the individual elements may be weighted to incorporate the prior distributions.

7. Solving the $c_{1,2}$ functions

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$c_1 = c_1(x_1, y_1, x_2, y_2)$$
$$c_2 = c_2(x_1, y_1, x_2, y_2)$$

There are known features of $c_{1,2}$ which may be used to generate constraints. These include “data dependence”: $c_{1,2}$ must be, in some way, dependent upon the data in the table;

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$-1 < c_i(x_1, y_1, x_2, y_2) > 1$$

a “valid probability range”: the values of $P(H_x)$ must fall between 0 and 1;

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$P(H_1) + P(H_2) = 1$$

"complementarity": the law of total probability requires that the sum of all probabilities = 1;

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$c_i(x_1, y_1, x_2, y_2) = c_i(y_1, x_1, y_2, x_2)$$
$$c_1(x_1, y_1, x_2, y_2) = c_2(x_2, y_2, x_1, y_1)$$

“symmetry”: the exchanging of rows in the contingency table should not affect the calculated probability value, and if the columns are exchanged then the values should map;

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

	H_1	H_2	
D_1	1	1	$\rightarrow P(H_1) = \frac{m}{m+n}$
D_2	m	n	

	H_1	H_2	
D_1	m	n	$\rightarrow P(H_1) = \frac{m}{m+n}$
D_2	1	1	

	H_1	H_2	
D_1	m	1	$\rightarrow P(H_1) = \frac{m}{m+n}$
D_2	1	n	

	H_1	H_2	
D_1	n	m	$\rightarrow P(H_1) = 0.5$
D_2	m	n	

	H_1	H_2	
D_1	m	m	$\rightarrow P(H_1) = 0.5$
D_2	n	n	

	H_1	H_2	
D_1	m	m	$\rightarrow P(H_1) = 0.5$
D_2	m	m	

“known probabilities”: there are certain contingency table structures which must return specific probabilities.

Solving the $c_{1,2}$ functions

	H_1	H_2
D_1	x_1	x_2
D_2	y_1	y_2

$$c_1(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_1 y_1}}{2 x_2 y_2}$$

$$c_2(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_2 y_2}}{2 x_1 y_1}$$

Using these principles and constraints demonstrates that $c_{1,2}$ are anti-symmetric bivariate functional equations, to which only one solution exists.

8. The objective covariate probability

The objective covariate probability

	H_1	H_2
D_1	8	7
D_2	6	5

$$P(H_1|D_1 \cap D_2) = \frac{\frac{x_1 y_1}{x_2 y_2} + x_1 + y_1}{\frac{x_1 y_1}{x_2 y_2} + x_1 + y_1 + \frac{x_2 y_2}{x_1 y_1} + x_2 + y_2}$$

$$\approx 0.5896$$

Substituting in the derived $c_{1,2}$ functional expressions allows for a final probability to be calculated.

9. The implications for psychology

The implications for psychology

"Calculating probabilities for predicting performance"

With only 10 data points in the "pot" example, there is not much difference between 0.5896 (QT) and 0.578 (classical Bayes' theorem) and is unlikely to affect ordinal predictions. However, in modelling phenomena based on thousands, or millions, of data points (eg., in perception, memory, social learning etc.) this difference will matter a lot more.

The implications for psychology

“Predicting new phenomena”

Bayesian learning lends itself to modelling systems that develop linearly. However, humans often show nonlinear, sometimes seemingly nondeterministic, behaviours, such as sudden switches in strategy that don't necessarily accord with the available data.

10. The relational information seeker

The relational information seeker

We conducted an experiment with a larger, 3x4, contingency table, giving the participants (n=150) 5 degrees of freedom in their selections.

For the first 4 selections, the choices made followed an information gain model, based on Shannon's entropy, with a significance of $p < 0.0001$ for each choice (using a Chi-squared test of predicted selection against random).

The relational information seeker

However, the final selection demonstrated a strategy change towards “weak” information. This suggests that the search process only follows information theory in-so-far as it is required to identify the diagnostically important relationships.

The relational information seeker

However, the final selection demonstrated a strategy change towards “weak” information. This suggests that the search processonly follows information theory in-so-far as it is required to identify the diagnostically important relationships.

This is not the same as mental model building. Rather, information search refines the mental representation created by the question.

The relational information seeker

It is unclear as to whether these relationships are classical, or quantum, in nature.

11. Conclusions

Conclusions

Any full description of objective reality may have to include mathematical concepts that only exist in quantum mechanics.

Conclusions

Any full description of objective reality may have to include mathematical concepts that only exist in quantum mechanics.

Quantum mechanics can describe models, and provide solutions to them, which lie beyond the scope of classical mathematics.

Conclusions

Any full description of objective reality may have to include mathematical concepts that only exist in quantum mechanics.

Quantum mechanics can describe models, and provide solutions to them, which lie beyond the scope of classical mathematics.

Bayes' theorem is a special case of a more general, quantum mechanical expression.

Download this presentation from
`http://rlb.me/pdf1215`

Rachael Bond

University of Sussex

Professor Tom Ormerod

University of Sussex

Professor Yang-Hui He

City University; Nankai University;
Merton college, Oxford University

References

[1] Doherty, M.E., Mynatt, C.R., Tweney, R.D., & Schiavo, M.D. (1979). Pseudodiagnosticity. *Acta Psychologica*, vol. 43(2), pp. 111-121. doi: 10.1016/0001-6918(79)90017-9

[2] Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 39(1), pp. 1-38.

[3] de Finetti, B. (1974). *Theory of probability: A critical introductory treatment*. New York, New York: Wiley.

[4] Caves, C.M., Fuchs, C.A., & Schack, R. (2002). Unknown quantum states: The quantum de Finetti representation. *Journal of Mathematical Physics*, vol. 43(9), pp. 4537-4559. doi: 10.1063/1.1494475

[5] Dirac, P.A.M. (1939). A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 35(03), pp. 416-418. doi: 10.1017/S0305004100021162

[6] Strang, G. (1980). *Linear algebra and its applications* (2nd ed.). New York, New York: Academic Press.