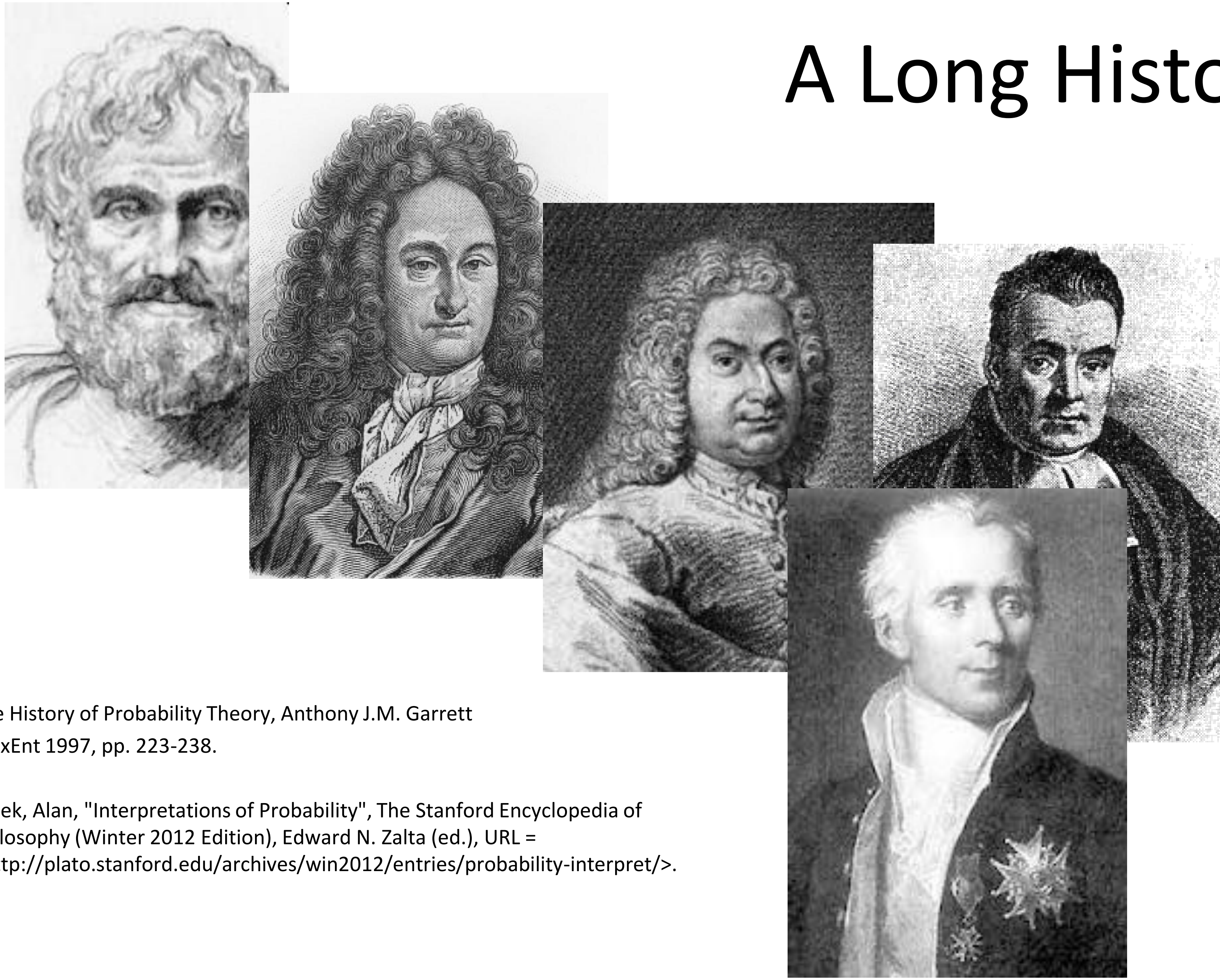
A Modern History of Probability Theory

Kevin H. Knuth
Depts. of Physics and Informatics
University at Albany (SUNY)
Albany NY USA

A Modern History of Probability Theory

Kevin H. Knuth
Depts. of Physics and Informatics
University at Albany (SUNY)
Albany NY USA

A Long History



The History of Probability Theory, Anthony J.M. Garrett
MaxEnt 1997, pp. 223-238.

Hájek, Alan, "Interpretations of Probability", The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.

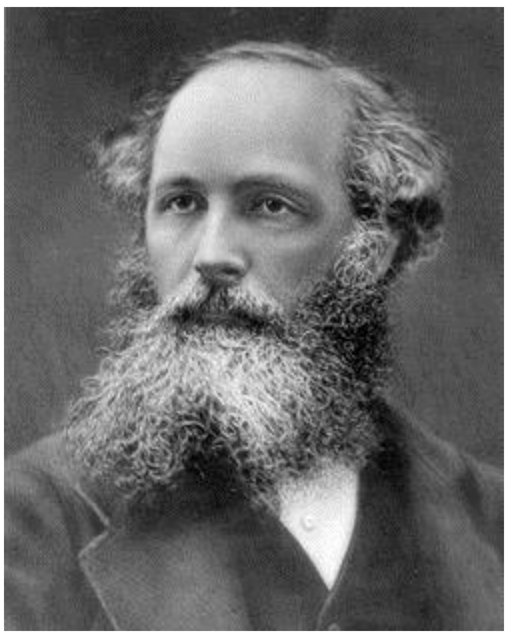
… la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul …

… the theory of probabilities is basically just common sense reduced to calculation …

Pierre Simon de Laplace
Théorie Analytique des Probabilités

They say that Understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic; but the actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

J. CLERK MAXWELL



Taken from Harold Jeffreys "Theory of Probability"

The terms certain and probable describe the various degrees of rational belief about a proposition which different amounts of knowledge authorise us to entertain. All propositions are true or false, but the knowledge we have of them depends on our circumstances; and while it is often convenient to speak of propositions as certain or probable, this expresses strictly a relationship in which they stand to a corpus of knowledge, actual or hypothetical, and not a characteristic of the propositions in themselves. A proposition is capable at the same time of varying degrees of this relationship, depending upon the knowledge to which it is related, so that it is without significance to call a proposition probable unless we specify the knowledge to which we are relating it.

John Maynard Keynes

To this extent, therefore, probability may be called subjective. But in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is rational to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.

"In deriving the laws of probability from more fundamental ideas, one has to engage with what 'probability' means.

This is a notoriously contentious issue; fortunately, if you disagree with the definition that is proposed, there will be a get-out that allows other definitions to be preserved."

– Anthony J.M. Garrett,
"Whence the Laws of Probability", MaxEnt 1997

The function $p(x|y)$ is often read as 'the probability of x given y'

This is most commonly interpreted as the probability that the proposition x is true given that the proposition y is true.

This concept can be summarized as a **degree of truth**

Concepts of Probability:

- degree of truth

Laplace, Maxwell, Keynes, Jeffreys and Cox all presented a concept of probability based on a **degree of rational belief**.

As Keynes points out, this is not to be thought of as subject to human capriciousness, but rather what an ideally rational agent ought to believe.

Concepts of Probability:

- **degree of truth**
- **degree of rational belief**

Anton Garrett discusses Keynes as conceiving of probability as a **degree of implication**. I don't get that impression reading Keynes. Instead, it seems to me that this is the concept that Garrett had (at the time) adopted.

Garrett uses the word *implicability*.

Concepts of Probability:

- **degree of truth**
- **degree of rational belief**
- **degree of implication**

Concepts of Probability:

~~- degree of truth~~

- degree of rational belief

- degree of implication

John Skilling argued against relying on the concept of truth thusly:

"You wouldn't know the truth if I told it to you!"

Meaning of Probability

Concepts of Probability:

- ~~**degree of truth**~~
- ~~**degree of rational belief**~~
- **degree of implication**

Jeffrey Scargle once pointed out that if probability quantifies truth or degrees of belief, one cannot assign a non-zero probability to a model that is known to be an approximation.

One cannot claim to be making inferences with any honesty or consistency while entertaining a concept of probability based on a degree of truth or a degree of rational belief.

Concepts of Probability:

degree of truth

Jeffre ies
truth
proba on.

Can I give you a “Get-Out”
like Anton did?

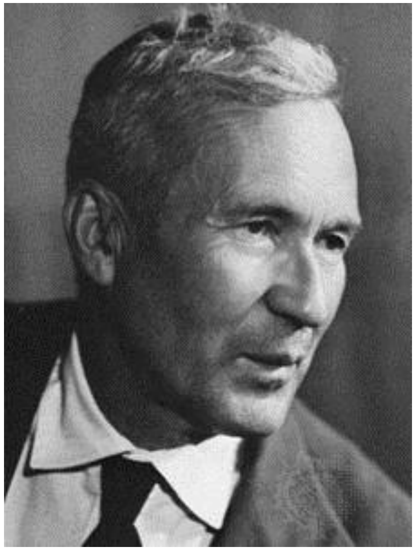
One c sty or
consi ased
on a degree of truth or a degree of rational belief.

Three Foundations of Probability Theory

Bruno de Finetti - 1931

Foundation Based on Consistent Betting

Unfortunately, the most commonly presented foundation of probability theory in modern quantum foundations



Andrey Kolmogorov - 1933

Foundation Based on Measures on Sets of Events

Perhaps the most widely accepted foundation by modern Bayesians

Richard Threlkeld Cox - 1946

Foundation Based on Generalizing Boolean Implication to Degrees

The foundation which has inspired the most investigation and development

Bruno de Finetti - 1931

Foundation Based on Consistent Betting

Unfortunately, the most commonly presented foundation of probability theory in modern quantum foundations

Subjective Bayesianism and the Dutch Book Argument

De Finetti conceived of probabilities as a degree of belief which could be quantified by considering how much one would be willing to bet on a proposition.

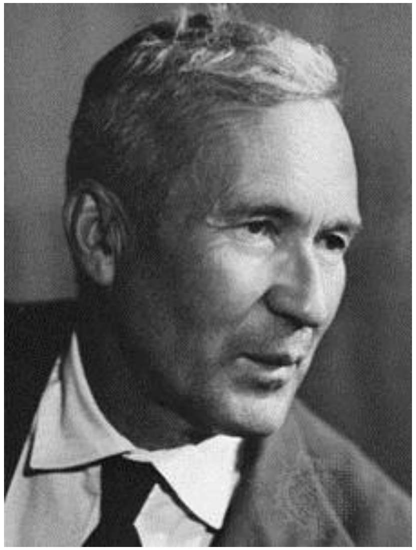
Consistency in betting is central to the foundation.

A **Dutch Book** is a series of bets which guarantees that one person will profit over another regardless of the outcome.

One can show that if one's subjective degree of belief does not obey the probability calculus, then one is susceptible to a Dutch Book.

Moreover, one can avoid a Dutch Book by ensuring that one's subjective degree of belief is in agreement with the probability calculus.

Important due to its reliance on consistency.



Andrey Kolmogorov - 1933

Foundation Based on
Measures on Sets
of Events

Perhaps the most widely
accepted foundation
by modern Bayesians

Kolmogorov's Probability Calculus

Axiom I (Non-Negativity)
Probability is quantified by a non-negative real number.

Axiom II (Normalization)
Probability has a maximum value $\Pr(e) \leq 1$ such that the probability that an event in the set *E* will occur is unity.

Axiom III (Finite Additivity)
Probability is σ-additive, such that the probability of any countable union of disjoint events $e_1, e_2, \cdots \epsilon E$ is given by $\Pr(e_1 \cup e_2 \cup \cdots) = \sum_i^\infty \Pr(e_i)$.

It is perhaps the both the conventional nature of his approach and the simplicity of the axioms that has led to such wide acceptance of his foundation.

Richard Threlkeld Cox - 1946

Foundation Based on Generalizing Boolean Implication to Degrees

The foundation which has inspired the most investigation and development

Generalizing Boolean Logic to Degrees of Belief

Axiom 0

Probability quantifies the reasonable credibility of a proposition when another proposition is known to be true

Axiom I

The likelihood $c \cdot b \mid a$ is a function of $b|a$ and $c| \; b \cdot a$

$c \cdot b \mid a = \mathrm{F}(b|a, c \mid b \cdot a)$

Axiom II

There is a relation between the likelihood of a proposition and its contradictory

$\sim b|a = S(b \mid a)$

In Physics we have a saying,

“The greatness of a scientist is measured by how long he/she retards progress in the field.”

Both de Finetti and Kolmogorov considered a well-defined domain, left few loose ends, and no noticeable conceptual glitches to give their disciples sufficient reason or concern to keep investigating.

Cox, on the other hand, proposed a radical approach that raised concerns about how belief could be quantified as well as whether one could improve upon his axioms despite justification by common-sense.

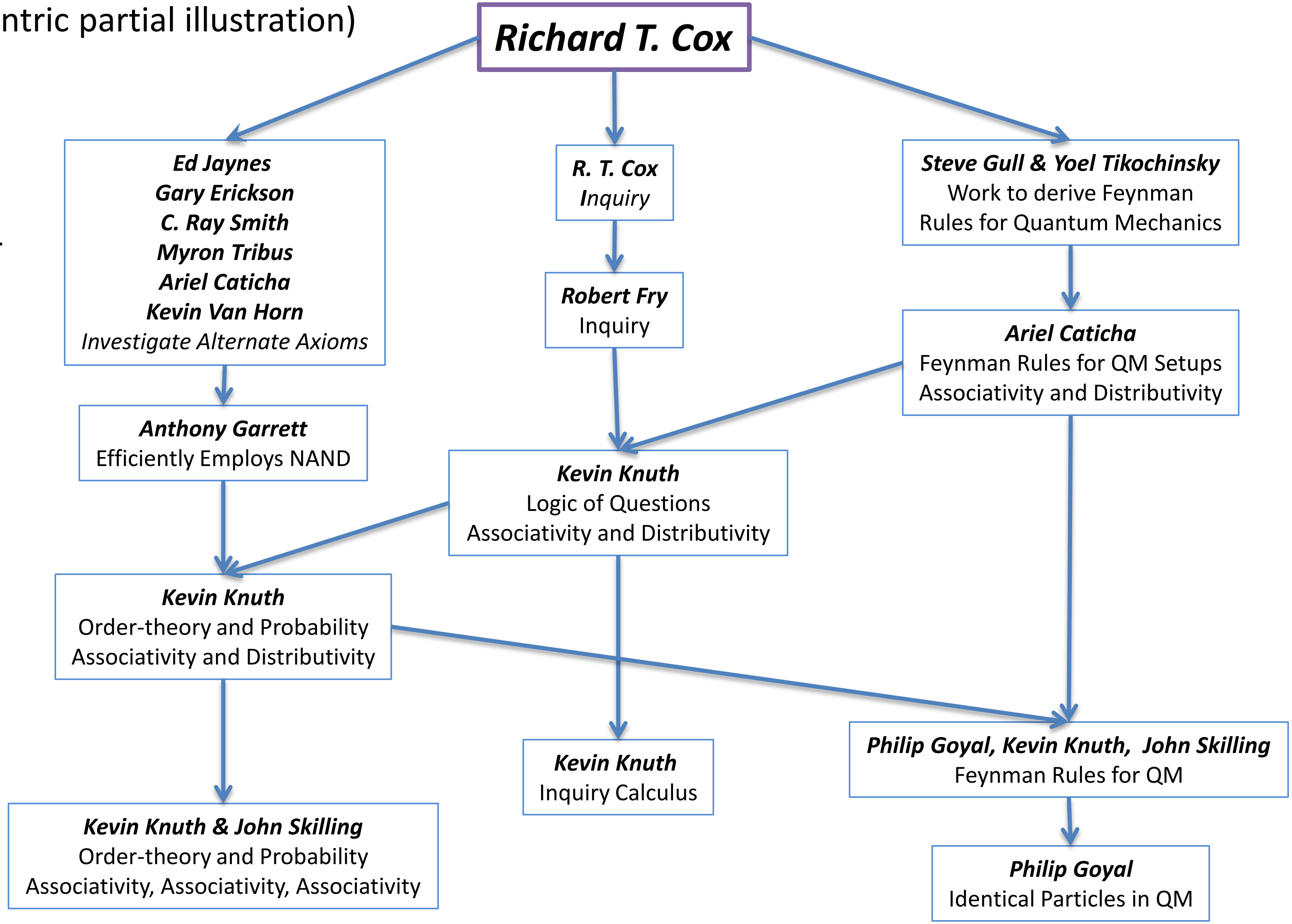
His work was just the right balance between

- **Pushing it far enough to be interesting**
- **Getting it right enough to be compelling**
- **Leaving it rough enough for there to be remaining work to be done**

And Work Was Done!

(Knuth-centric partial illustration)

Jos Uffink
Imre Czisar



Probability Theory Timeline

- 1920
- John Maynard Keynes - 1921
- 1930
- Bruno de Finetti - 1931
- Andrey Kolmogorov - 1933
- Sir Harold Jeffreys - 1939
- 1940
- Richard Threlkeld Cox - 1946
- Claude Shannon - 1948
- 1950
- Edwin Thompson Jaynes - 1957
- 1960

Probability Theory Timeline

1920

John Maynard Keynes - 1921

1930

Bruno de Finetti - 1931
Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939

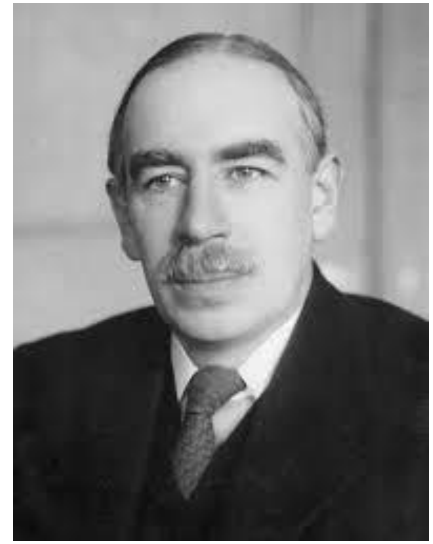
1940

Richard Threlkeld Cox - 1946
Claude Shannon - 1948

1950

Edwin Thompson Jaynes - 1957

1960



Probability Theory Timeline

- 1920
- John Maynard Keynes - 1921
- 1930
- **Bruno de Finetti - 1931**
- Andrey Kolmogorov - 1933
- Sir Harold Jeffreys - 1939
- 1940
- Richard Threlkeld Cox - 1946
- Claude Shannon - 1948
- 1950
- Edwin Thompson Jaynes - 1957
- 1960

Probability Theory Timeline

1920

John Maynard Keynes - 1921

1930

Bruno de Finetti - 1931

Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939

1940

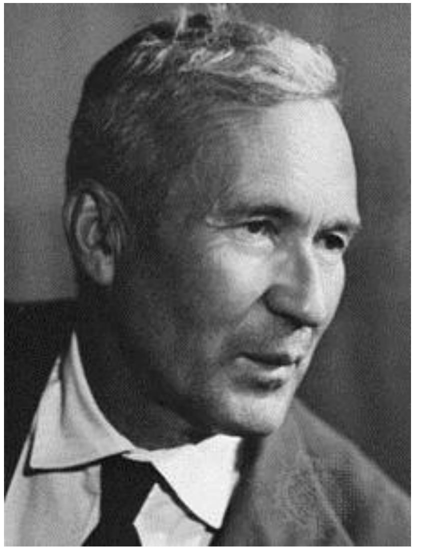
Richard Threlkeld Cox - 1946

Claude Shannon - 1948

1950

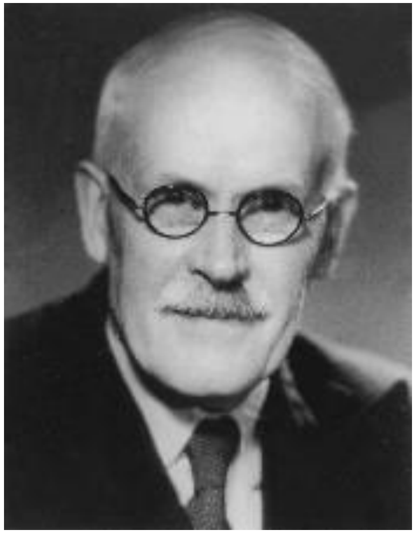
Edwin Thompson Jaynes - 1957

1960



Probability Theory Timeline

- 1920
  - John Maynard Keynes - 1921
- 1930
  - Bruno de Finetti - 1931
  - Andrey Kolmogorov - 1933
  - **Sir Harold Jeffreys - 1939**
- 1940
  - Richard Threlkeld Cox - 1946
  - Claude Shannon - 1948
- 1950
  - Edwin Thompson Jaynes - 1957
- 1960



Probability Theory Timeline

1920
John Maynard Keynes - 1921

1930
Bruno de Finetti - 1931
Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939
1940

Richard Threlkeld Cox - 1946
Claude Shannon - 1948
1950

Edwin Thompson Jaynes - 1957
1960

Probability Theory Timeline

1920

John Maynard Keynes - 1921

1930

Bruno de Finetti - 1931
Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939

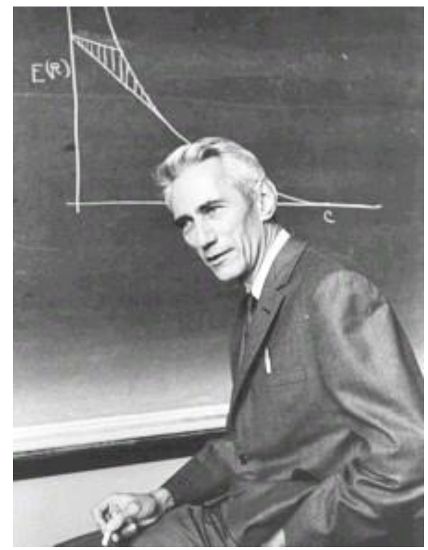
1940

Richard Threlkeld Cox - 1946
Claude Shannon - 1948

1950

Edwin Thompson Jaynes - 1957

1960



Probability Theory Timeline

1920

John Maynard Keynes - 1921

1930

Bruno de Finetti - 1931
Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939

1940

Richard Threlkeld Cox - 1946
Claude Shannon - 1948

1950

Edwin Thompson Jaynes - 1957

1960

Probability Theory Timeline

1920

John Maynard Keynes - 1921

1930

Bruno de Finetti - 1931
Andrey Kolmogorov - 1933

Sir Harold Jeffreys - 1939

1940

Richard Threlkeld Cox - 1946
Claude Shannon - 1948

1950

Edwin Thompson Jaynes - 1957

1960

Quantum Mechanics Timeline

1920

Niels Bohr – 1922 (NP)

Erwin Schrödinger - 1926

1930

Werner Heisenberg – 1932 (NP)

John Von Neumann - 1936

1940

Richard Feynman - 1948

1950

1960

Familiarity breeds the illusion of understanding

Anonymous

In graduate school I asked:

why

V

=

results in

$$1 + 2 = 3$$

$\vee$

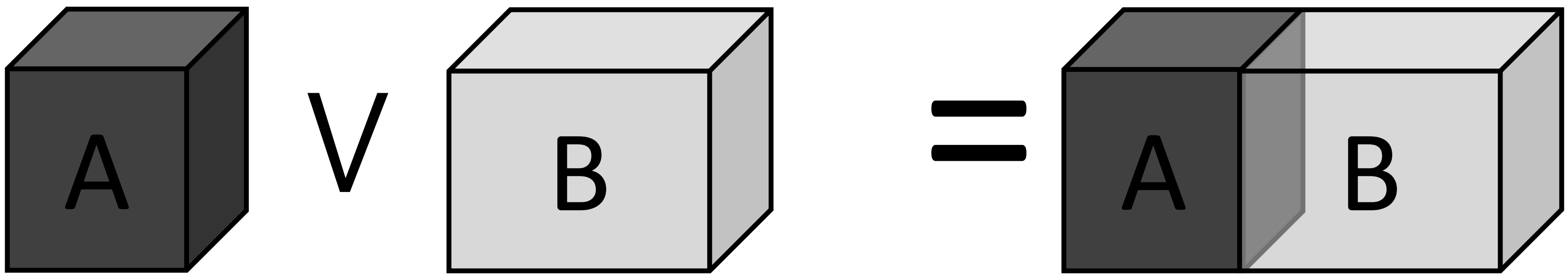
=

$$1 + 2 = 3$$

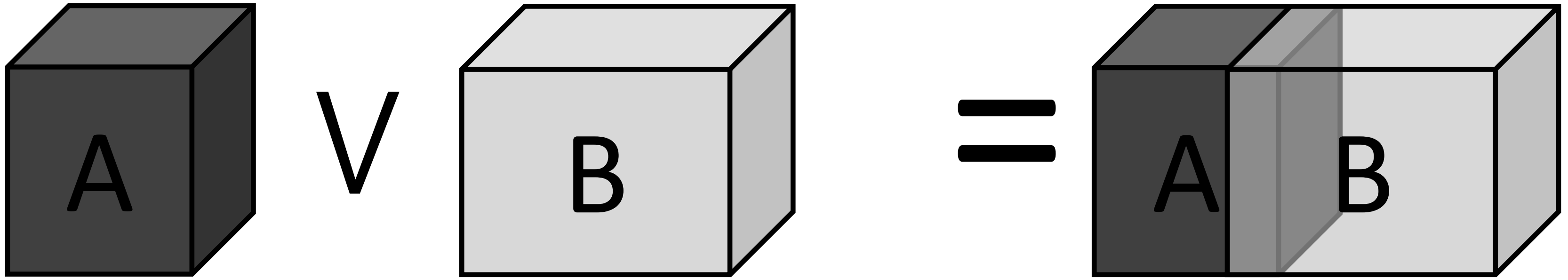
$\vee$

=

$$1 + 2 = 3$$



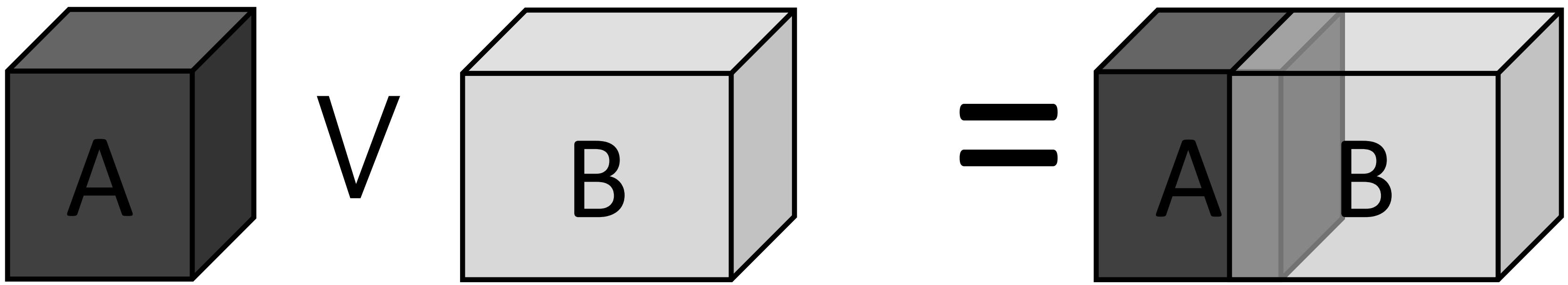
$$v(A \cup B) = v(A) + v(B)$$



$$v(A \cup B) = v(A) + v(B) - v(A \cap B)$$

volume

Knuth, MaxEnt 2003



$$s(A \cup B) = s(A) + s(B) - s(A \cap B)$$

surface area

Knuth, MaxEnt 2003

$$\Pr(A \vee B \mid I) = \Pr(A \mid I) + \Pr(B \mid I) - \Pr(A \wedge B \mid I)$$

sum rule of probability

Knuth, MaxEnt 2003

$$I(A;B) = H(A) + H(B) - H(A,B)$$

mutual information

Knuth, MaxEnt 2003

$$max(a, b) = a + b - min(a, b)$$

polya's min-max rule

Knuth, MaxEnt 2003

$$\log\left(LCM(a,b)\right) = \log(a) + \log(b) - \log(GCD(a,b))$$

number theory identity

Knuth, MaxEnt 2009

Clearly, my original question:

why

V

=

results in

$$1 + 2 = 3$$

Is related to:

why the disjunction of A and B results in

$$\Pr(\mathrm{A} \vee B \mid I) = \Pr(\mathrm{A} \mid \mathrm{I}) + \Pr(\mathrm{B} \mid \mathrm{I}) - \Pr(\mathrm{A} \wedge \mathrm{B} \mid \mathrm{I})$$

the essential content of both statistical mechanics and communication theory, of course, does not lie in the equations; it lies in the ideas that lead to those equations

E. T. Jaynes

the essential content of both statistical mechanics and communication theory, of course, does not lie in the equations; it lies in the ideas that lead to those equations

E. T. Jaynes

the essential content of both statistical mechanics and communication theory, of course, does not lie in the equations; it lies in the ideas that lead to those equations

E. T. Jaynes

A MODERN PERSPECTIVE

Measure what is measurable,

and make measurable that which is not so.

Galileo Galilei

ideas that lead to equations

Laws of the Universe
Reflect
An Underlying Order

↑

Laws are fundamental
and are dictated by
God or Mother Nature

Underlying Order
Constrains
Quantification

↑

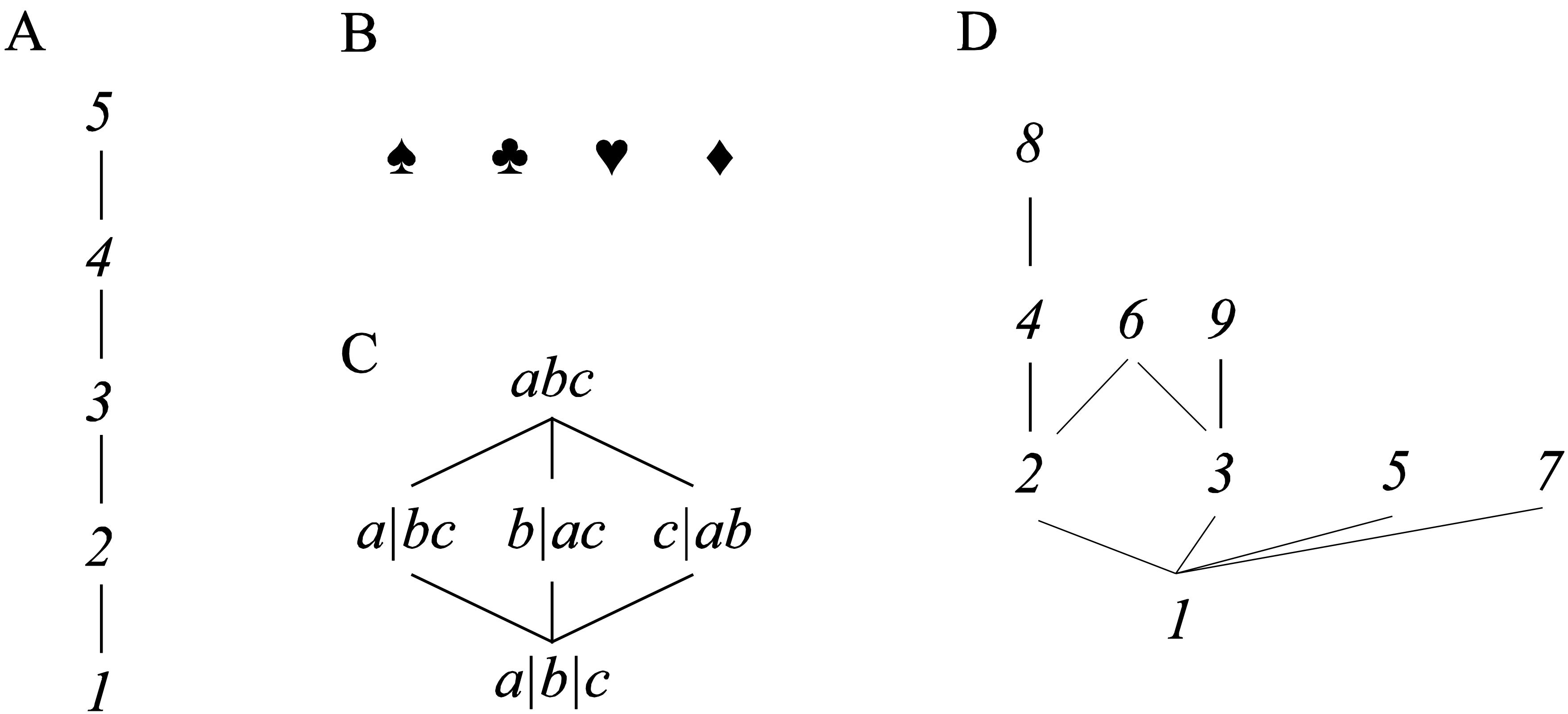
Order and symmetries
are fundamental
Laws are constraints
on quantification

When hypothesizing Laws,
one can be right or wrong

whereas

Applying consistent quantification
can only be useful or not useful

Lattices are partially ordered sets where each pair of elements has a least upper bound and a greatest lower bound



Lattices are Algebras

Structural Viewpoint

Operational Viewpoint

$$a \le b \quad \Leftrightarrow \quad \begin{array}{l} a \vee b = b \\ a \wedge b = a \end{array}$$

Structural Viewpoint	Operational Viewpoint
$a \le b \quad \Leftrightarrow$	$a \vee b = b$ $a \wedge b = a$

Sets, Is a subset of

$$a \subseteq b \quad \Leftrightarrow \quad \begin{array}{c} a \cup b = b \\ a \cap b = a \end{array}$$

Positive Integers, Divides

$$a \mid b \quad \Leftrightarrow \quad \begin{array}{c} \mathrm{lcm}(a,b) = b \\ \gcd(a,b) = a \end{array}$$

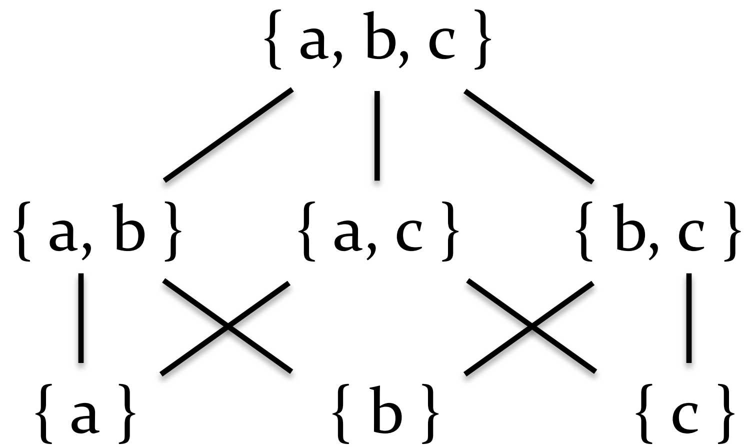
Assertions, Implies

$$a \rightarrow b \quad \Leftrightarrow \quad \begin{array}{c} a \vee b = b \\ a \wedge b = a \end{array}$$

Integers, Is less than or equal to

$$a \le b \quad \Leftrightarrow \quad \begin{array}{c} \max(a,b) = b \\ \min(a,b) = a \end{array}$$

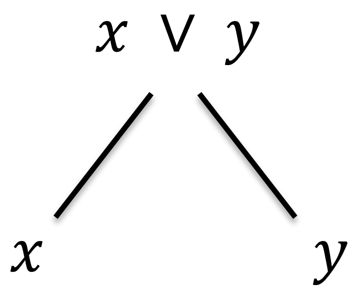
quantify the partial order $\equiv$ assign real numbers to the elements



$$f : x \in L \rightarrow \mathbb{R}$$

Require that quantification be consistent with the structure.
Otherwise, information about the partial order is lost.

Any general rule must hold for special cases
Look at special cases to constrain general rule



$$f : x \in L \ \rightarrow \ \mathbb{R}$$

Enforce local consistency

$$f(x \vee y) = f(x) \ \oplus \ f(y)$$

where $\oplus$ is an unknown operator to be determined.

Write the same element two different ways

$$x \vee (y \vee z) = (x \vee y) \vee z$$

which implies

$$f(x) \oplus \big(f(y) \oplus f(z)\big) = \big(f(x) \oplus f(y)\big) \oplus f(z)$$

Note that the unknown operator $\oplus$ is nested in two distinct ways, which reflects associativity

This is a functional equation known as the
Associativity Equation

$$f(x) \oplus \big(f(y) \oplus f(z)\big) = \big(f(x) \oplus f(y)\big) \oplus f(z)$$

where the aim is to find all the possible operators $\oplus$ that satisfy the equation above.

We require that the join operations are closed,
That the valuations respect ranking, i.e. $x \geq y \Rightarrow f(x) \geq f(y)$
And that $\oplus$ is commutative and associative.

The general solution to the Associativity Equation

$$f(x) \oplus \big(f(y) \oplus f(z)\big) = \big(f(x) \oplus f(y)\big) \oplus f(z)$$

is (Aczel 1966; Craigen and Pales 1989; Knuth and Skilling 2012):

$$F\big(f(x) \oplus f(y)\big) = F\big(f(x)\big) + F\big(f(y)\big)$$

where F is an arbitrary invertible function.

$$F\big(f(x)\ \oplus f(y)\big) = F\big(f(x)\big) + F\big(f(y)\big)$$

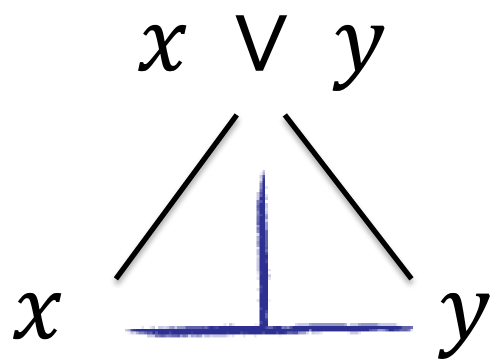
Since the function F is arbitrary and invertible, we can define a new quantification $v(x) = F\big(f(x)\big)$ so that the combination is always additive.

Thus we can always write

$$v(x \vee y) = v(x) + v(y)$$

In essence, we have **derived measure theory** from algebraic symmetries.

Additivity



$$v(x \vee y) = v(x) + v(y)$$

Knuth, MaxEnt 2009

Epiphany!

$\oplus$

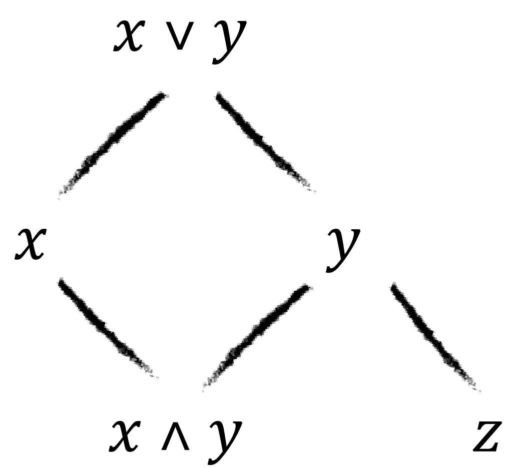
$=$

always results in

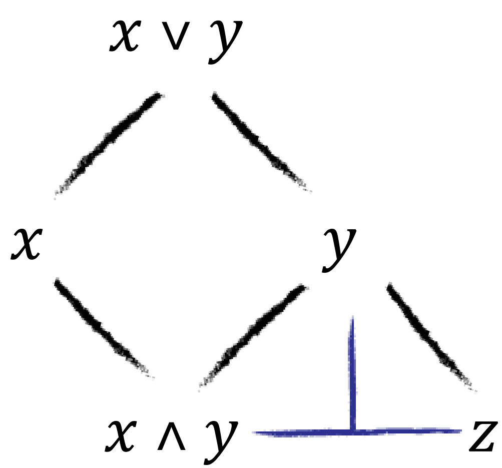
$$1 + 2 = 3$$

because combining crayons in this way is
closed, commutative, associative,
and I can order sets of crayons.

General Case

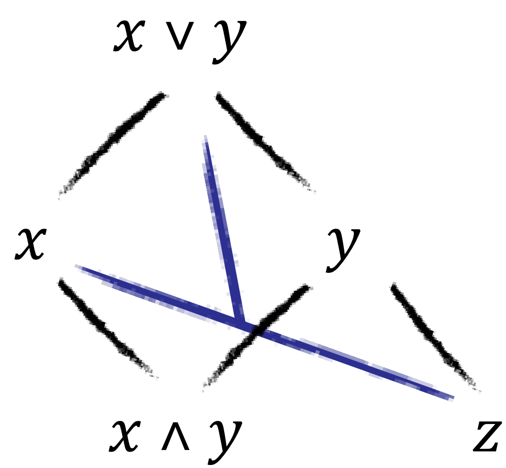


General Case



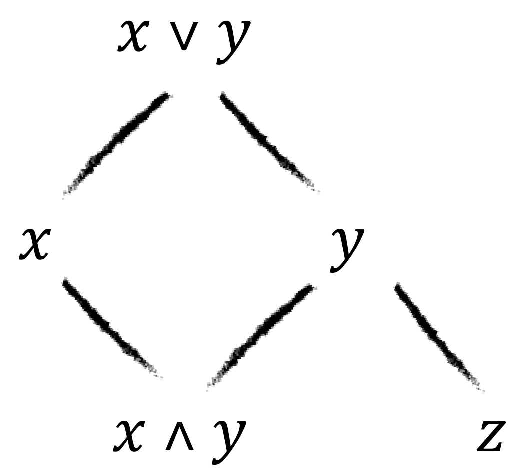
$$v(y) = v(x \wedge y) + v(z)$$

General Case



$$v(y) = v(x \wedge y) + v(z) \qquad v(x \vee y) = v(x) + v(z)$$

General Case

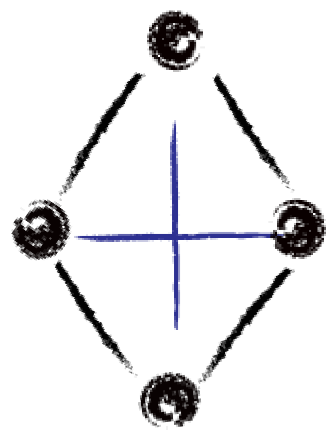


$$v(y) = v(x \wedge y) + v(z) \qquad v(x \vee y) = v(x) + v(z)$$

$$v(x \vee y) = v(x) + v(y) - v(x \wedge y)$$

Sum Rule

$$v(x \vee y) = v(x) + v(y) - v(x \wedge y)$$



$$v(x \vee y) + v(x \wedge y) = v(x) + v(y)$$

symmetric form (self-dual)

Fundamental symmetries are why the Sum Rule is ubiquitous

Ubiquity (inclusion-exclusion)

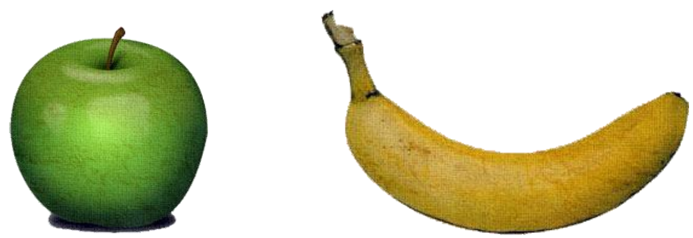
$\Pr(A \vee B \mid C) = \Pr(A \mid C) + \Pr(B \mid C) - \Pr(A \wedge B \mid C)$	Probability
$I(A;B) = H(A) + H(B) - H(A,B)$	Mutual Information
$Area(A \cup B) = Area(A) + Area(B) - Area(A \cap B)$	Areas of Sets
$\max(A,B) = A + B - \min(A,B)$	Polya's Min-Max Rule
$\log LCM(A,B) = \log A + \log B - \log GCD(A,B)$	Integral Divisors
$I_3(A,B,C) = \lvert A \sqcup B \sqcup C \rvert - \lvert A \sqcup B \rvert - \lvert A \sqcup C \rvert - \lvert B \sqcup C \rvert + \lvert A \rvert + \lvert B \rvert + \lvert C \rvert$	Amplitudes from three-slits (Sorkin arXiv:\\gr-qc/9401003)

The relations above are constraint equations ensuring consistent quantification in the face of certain symmetries

Knuth, 2003. Deriving Laws, arXiv:physics/0403031 [physics.data-an]
Knuth, 2009. Measuring on Lattices, arXiv:0909.3684 [math.GM]
Knuth, 2015. The Deeper Roles of Mathematics in Physical Laws, arXiv:1504.06686 [math.HO]

INFERENCE

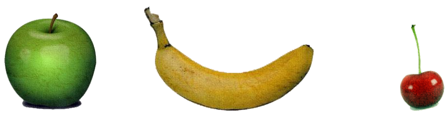
states



apple banana cherry

states of the contents of my grocery basket

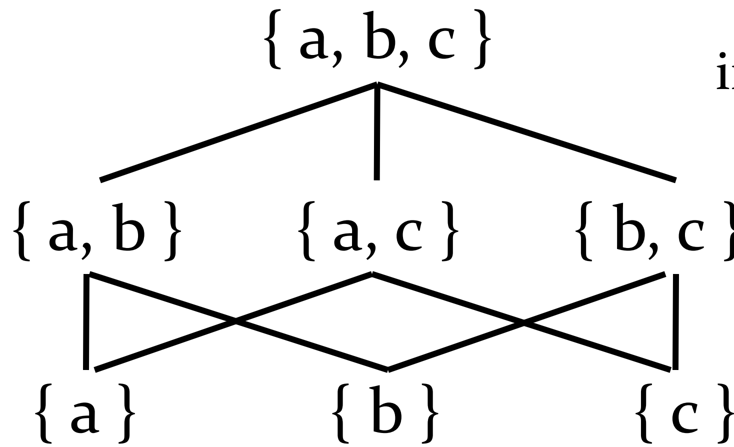
crudely describe knowledge by listing a set of potential states



a b c

states of the contents of
my grocery basket

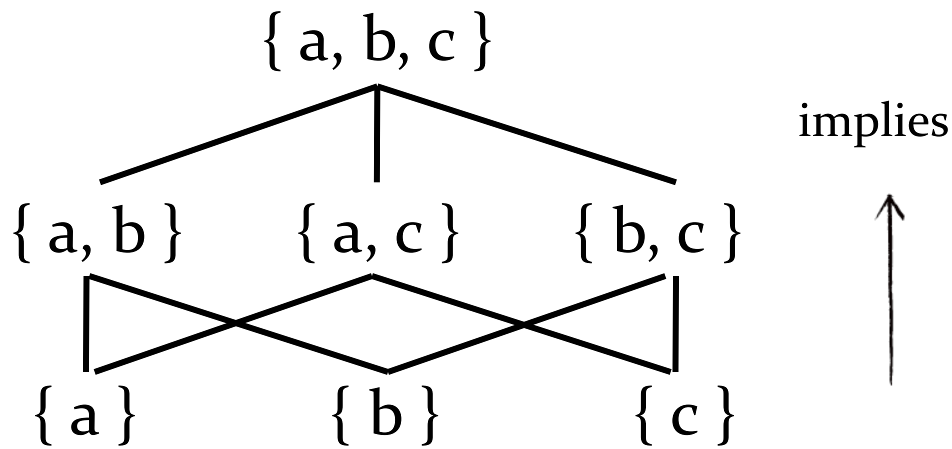
powerset



subset
inclusion

statements
about the contents of
my grocery basket

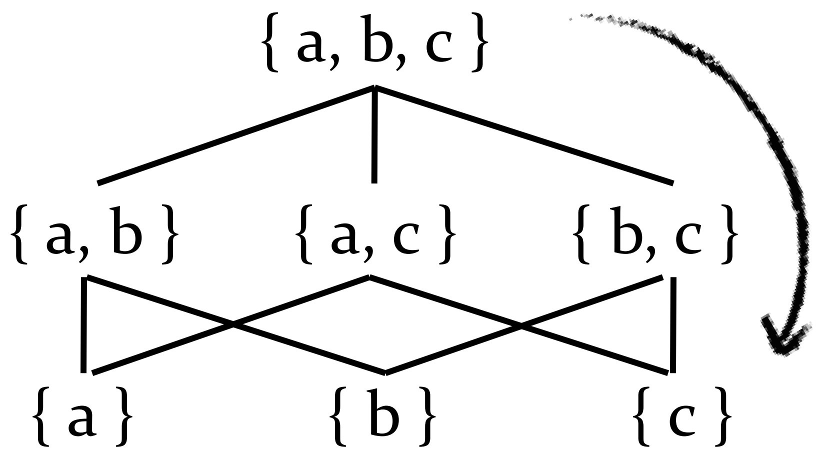
What can be said about a system?



statements
about the contents of
my grocery basket

ordering encodes implication
DEDUCTION

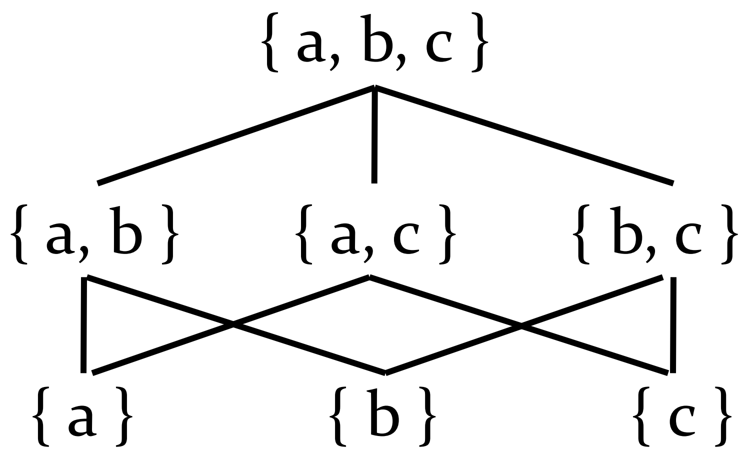
What can be said about a system?



statements
about the contents of
my grocery basket

Quantify to what degree
the statement that the system is in
one of three states {a, b, c}
implies knowing that it is
in some other set of states

inference works backwards



The Zeta function encodes inclusion on the lattice.

$$\zeta(x, y) = \begin{cases} 1 & if \;\; x \leq y \\ 0 & if \;\; x \not\leq y \end{cases}$$

Context and Bi-Valuations

BI-VALUATION $p : x, i \in L \rightarrow R$

Bi-Valuation				Valuation
$p(x \mid i)$	$\longrightarrow$	$v_i(x)$	$\longrightarrow$	$v(x)$
Context *i* is explicit		Measure of *x* with respect to Context *i*		Context *i* is implicit

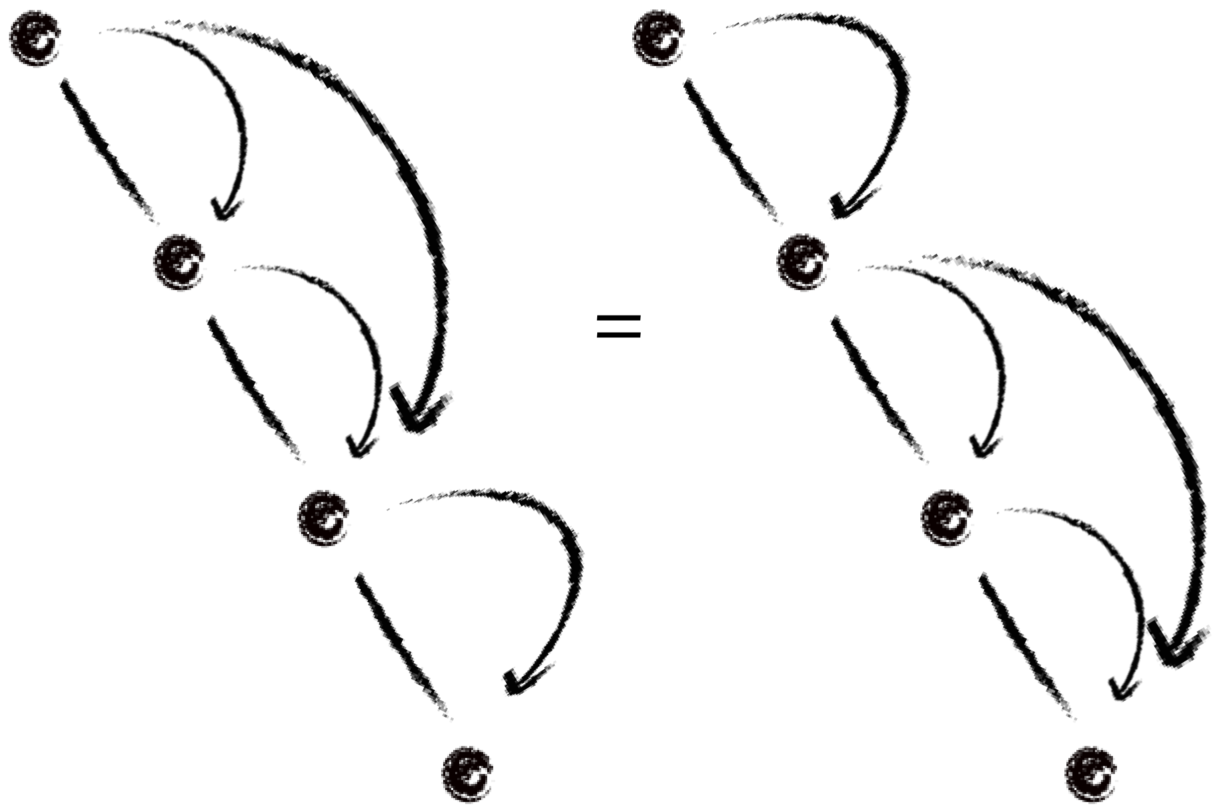
Bi-valuations generalize lattice inclusion to degrees of inclusion

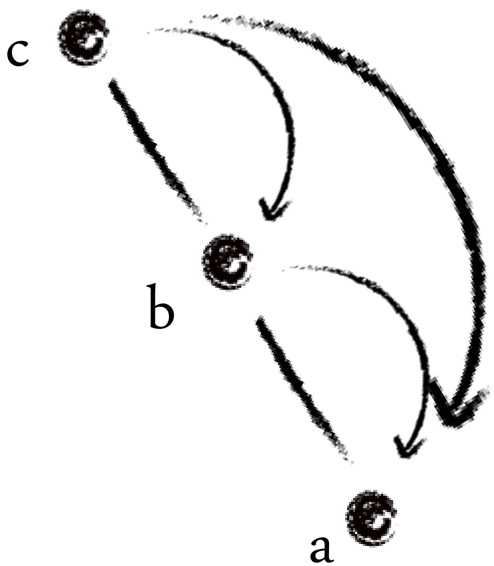
Context is Explicit

Sum Rule

$$p(x \mid i) + p(y \mid i) = p(x \vee y \mid i) + p(x \wedge y \mid i)$$

Associativity of Context





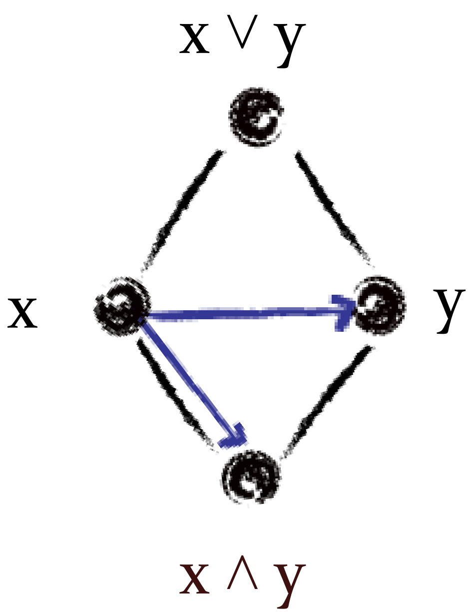
Chain Rule

$$p(a \mid c) = p(a \mid b)\, p(b \mid c)$$

Lemma

$$p(x \mid x) + p(y \mid x) = p(x \vee y \mid x) + p(x \wedge y \mid x)$$

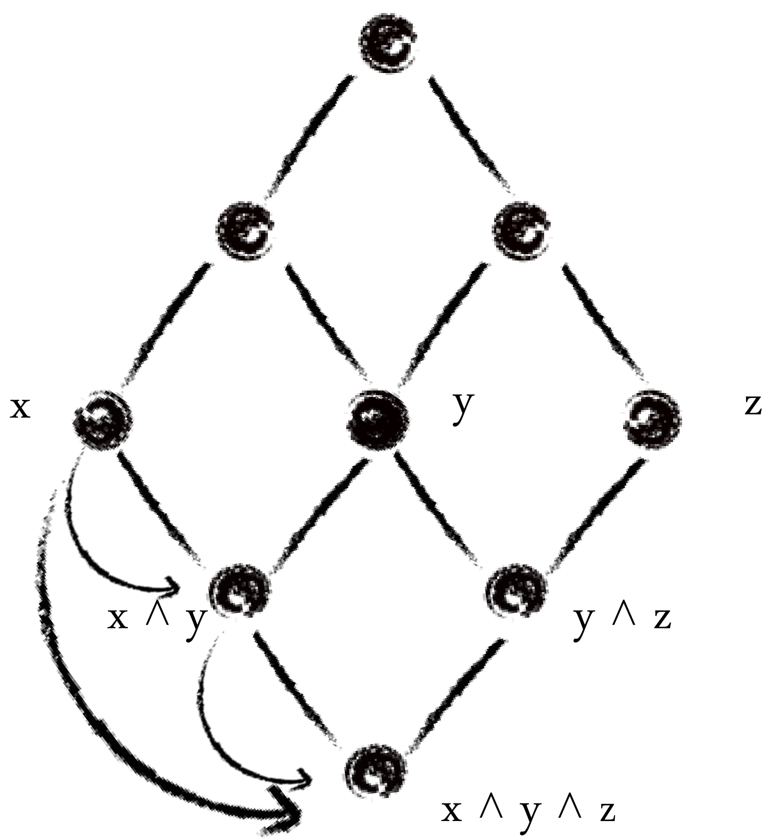
Since $x \le x$ and $x \le x \vee y$, $p(x \mid x) = 1$ and $p(x \vee y \mid x) = 1$



$$p(y \mid x) = p(x \wedge y \mid x)$$

Extending the Chain Rule

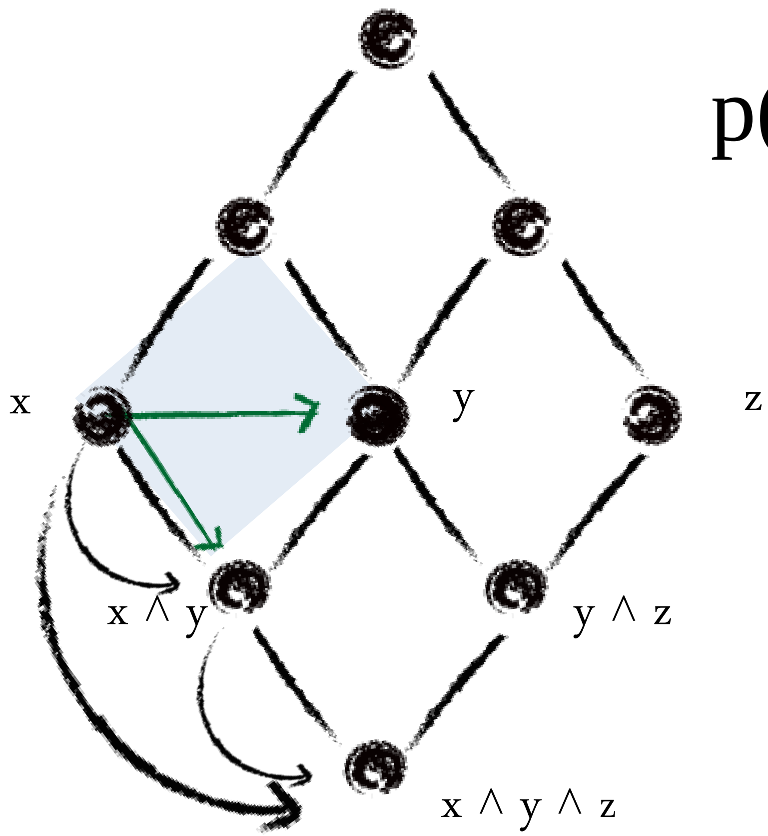
$$p(x \wedge y \wedge z \mid x) = p(x \wedge y \mid x)\, p(x \wedge y \wedge z \mid x \wedge y)$$



Extending the Chain Rule

$$p(x \wedge y \wedge z \,|\, x) = p(x \wedge y \,|\, x)\, p(x \wedge y \wedge z \,|\, x \wedge y)$$

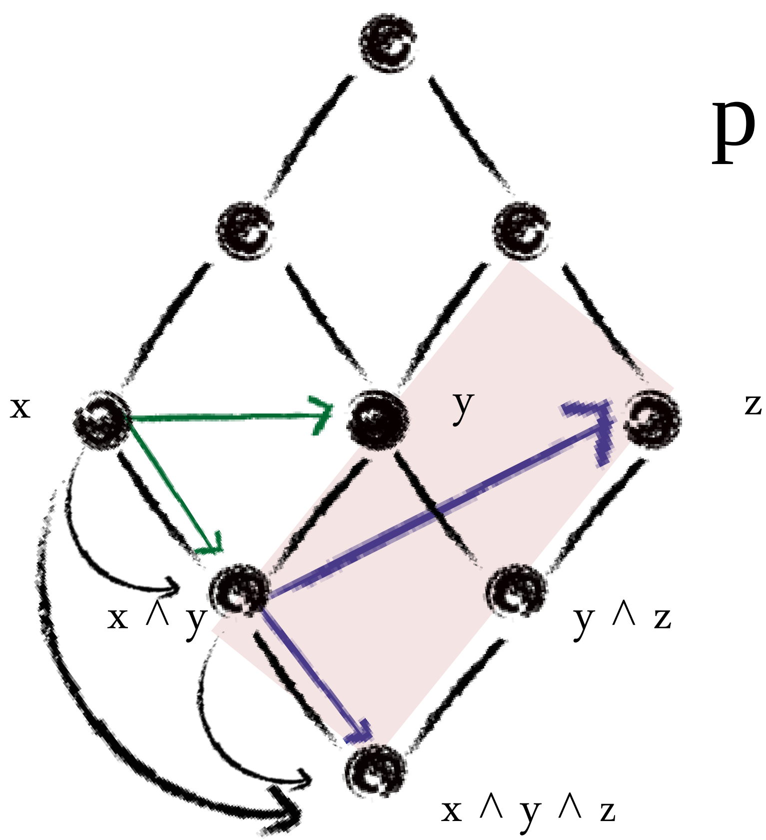
$$p(y \wedge z \,|\, x) = p(y \,|\, x)\, p(z \,|\, x \wedge y)$$



Extending the Chain Rule

$$p(x \wedge y \wedge z \mid x) = p(x \wedge y \mid x)\, p(x \wedge y \wedge z \mid x \wedge y)$$

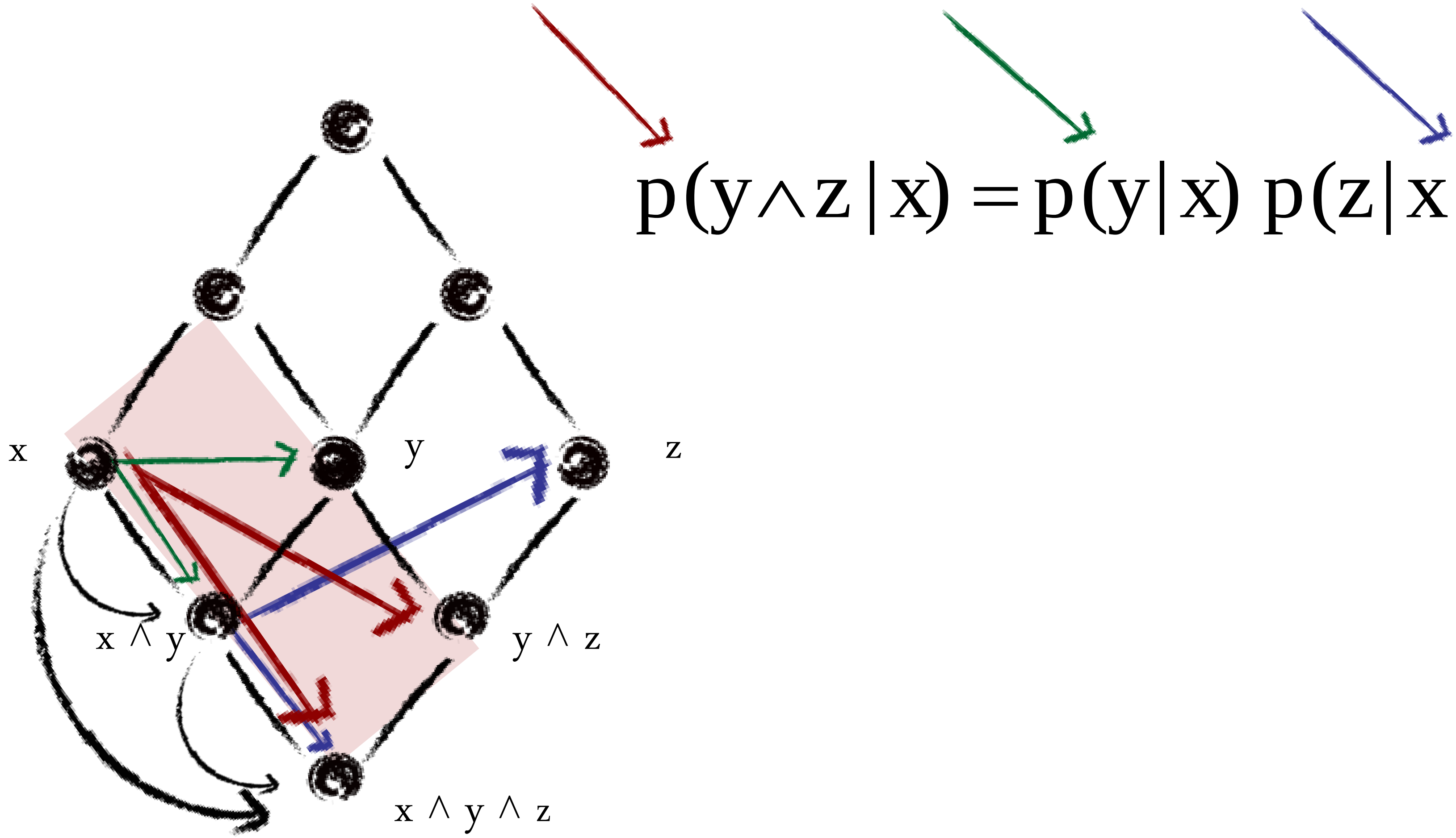
$$p(y \wedge z \mid x) = p(y \mid x)\, p(z \mid x \wedge y)$$



Extending the Chain Rule

$$p(x \wedge y \wedge z \mid x) = p(x \wedge y \mid x)\, p(x \wedge y \wedge z \mid x \wedge y)$$

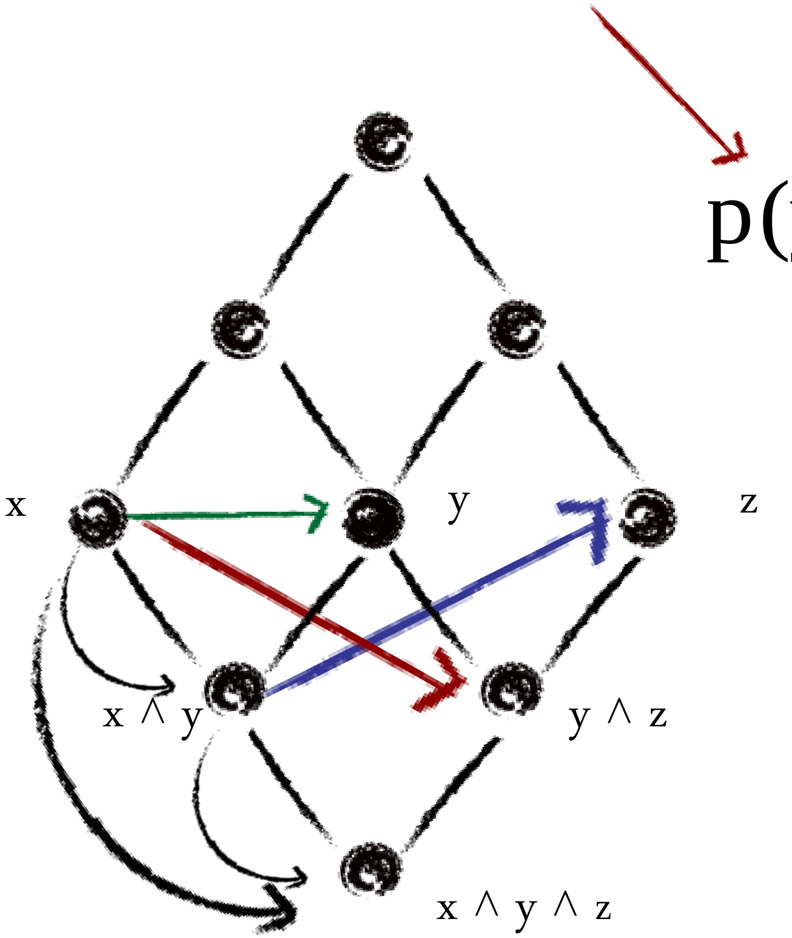
$$p(y \wedge z \mid x) = p(y \mid x)\, p(z \mid x \wedge y)$$



Extending the Chain Rule

$$p(x \wedge y \wedge z \mid x) = p(x \wedge y \mid x)\, p(x \wedge y \wedge z \mid x \wedge y)$$

$$p(y \wedge z \mid x) = p(y \mid x)\, p(z \mid x \wedge y)$$



Commutativity of the product
leads to **Bayes Theorem…**

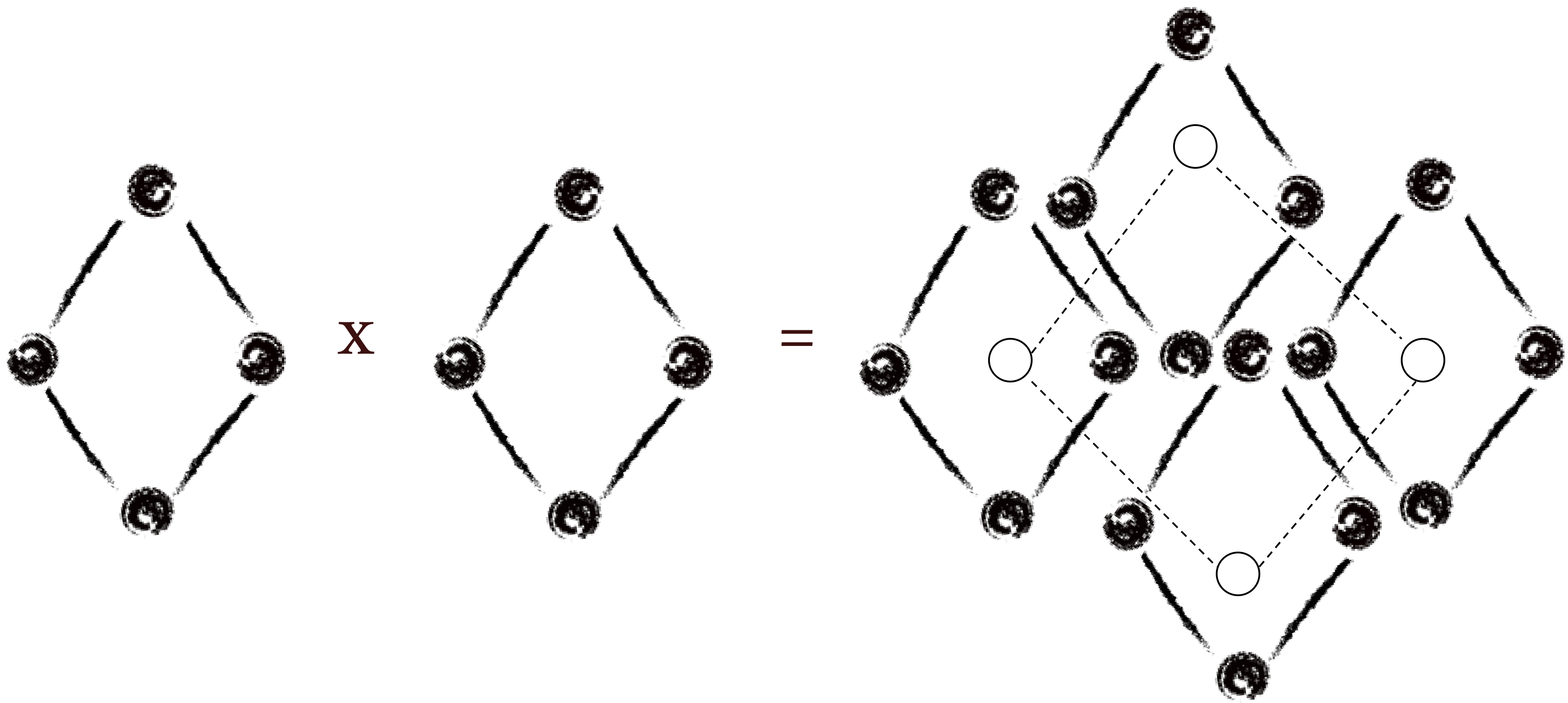
$$p(x \mid y \wedge i) = p(y \mid x \wedge i)\,\frac{p(x \mid i)}{p(y \mid i)}$$

↓

$$p(x \mid y) = p(y \mid x)\,\frac{p(x \mid i)}{p(y \mid i)}$$

Bayes Theorem involves a change of context.

Lattice Products



Direct (Cartesian) product of two spaces

Direct Product Rule

The lattice product is also associative

$$A \times (B \times C) \quad = \quad (A \times B) \times C$$

After the sum rule, the only freedom left is rescaling

$$p(a,b \mid i, j) \quad = \quad p(a \mid i)\, p(b \mid j)$$

which is again summation (after taking the logarithm)

Bayesian Probability Theory = Constraint Equations

Sum Rule

$$p(x \vee y \mid i) = p(x \mid i) + p(y \mid i) - p(x \wedge y \mid i)$$

Direct Product Rule

$$p(a, b \mid i, j) \quad = \quad p(a \mid i)\, p(b \mid j)$$

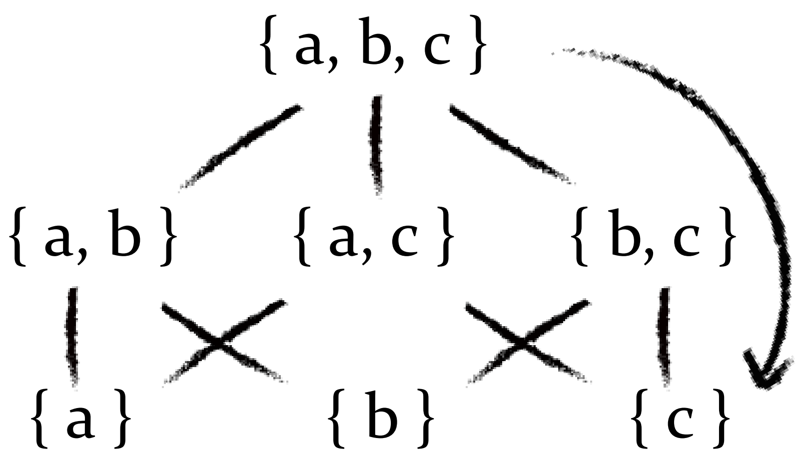
Product Rule

$$p(y \wedge z \mid x) = p(y \mid x)\, p(z \mid x \wedge y)$$

Bayes Theorem

$$p(x \mid y) = p(y \mid x) \frac{p(x \mid i)}{p(y \mid i)}$$

Inference



statements

Given a quantification of the join-irreducible elements, one uses the constraint equations to consistently assign any desired bi-valuations (probability)

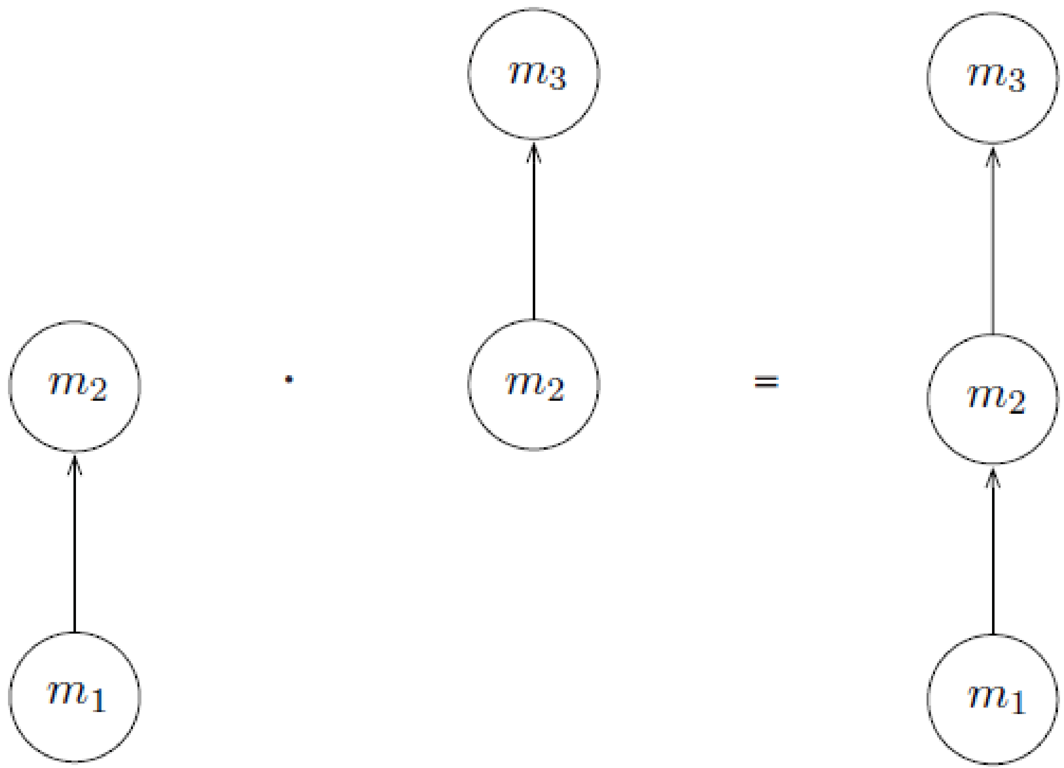
How far can we take these ideas?

Quantum Mechanics!

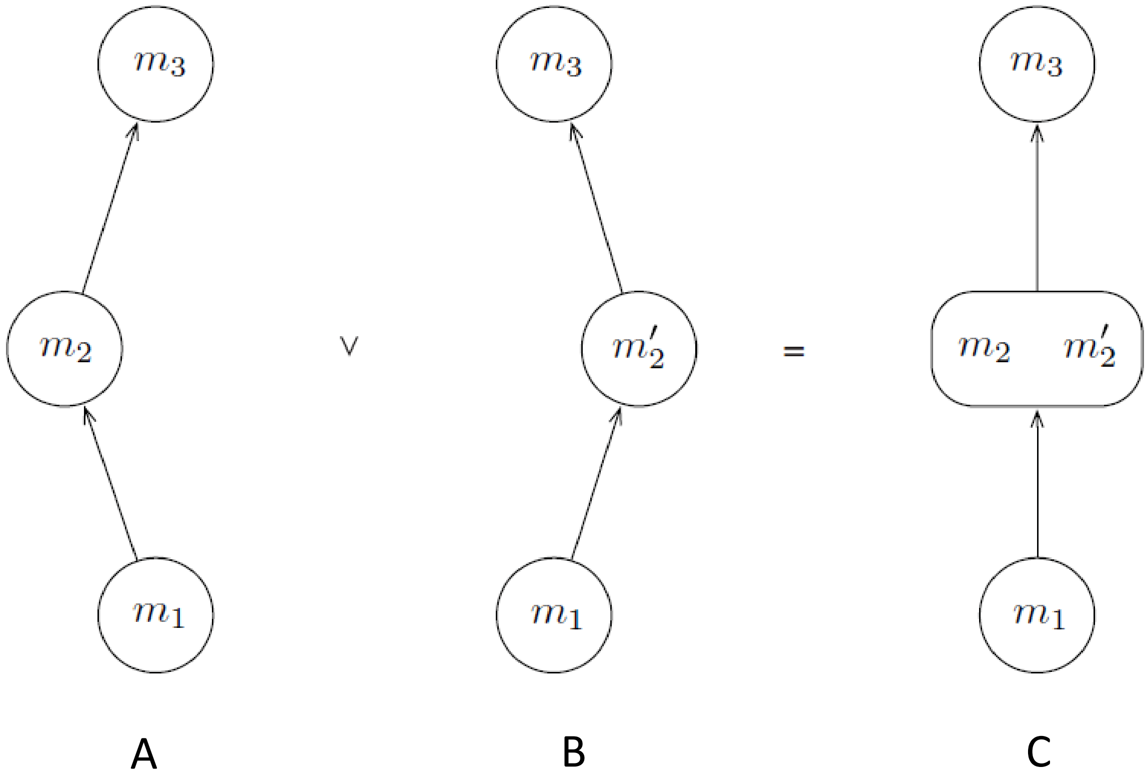
Quantum Measurements in Series

Quantum measurements can be performed in series.

Series combinations of measurement sequences are associative.



Quantum Measurements in Parallel

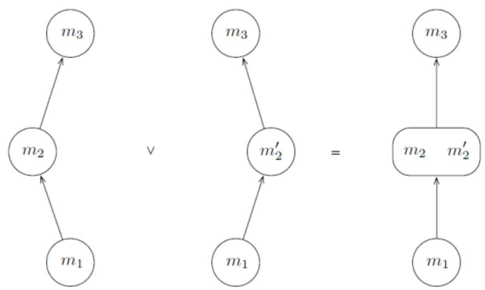


Quantum measurements can be performed in parallel (coarse graining).

Parallel combinations of measurement sequences are commutative and associative.

By quantifying a measurement sequence with a pair of numbers, $a = \binom{a_1}{a_2}$,

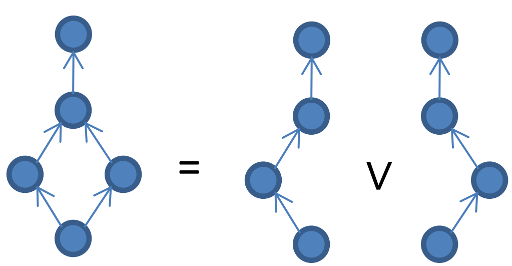
Associativity and Commutativity of Parallel combinations of measurements results in **component-wise additivity** of the pairs:



A B C

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

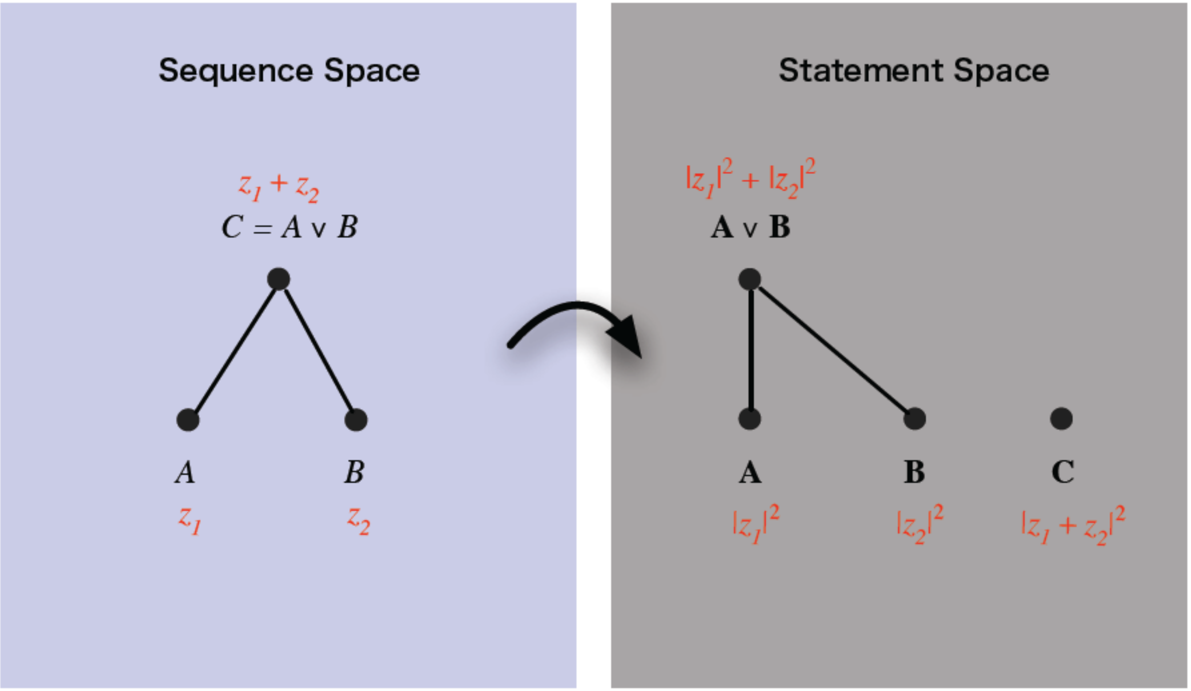
Distributivity of Series over Parallel combinations of measurements results in a **bilinear multiplicative form** for combining the pairs:

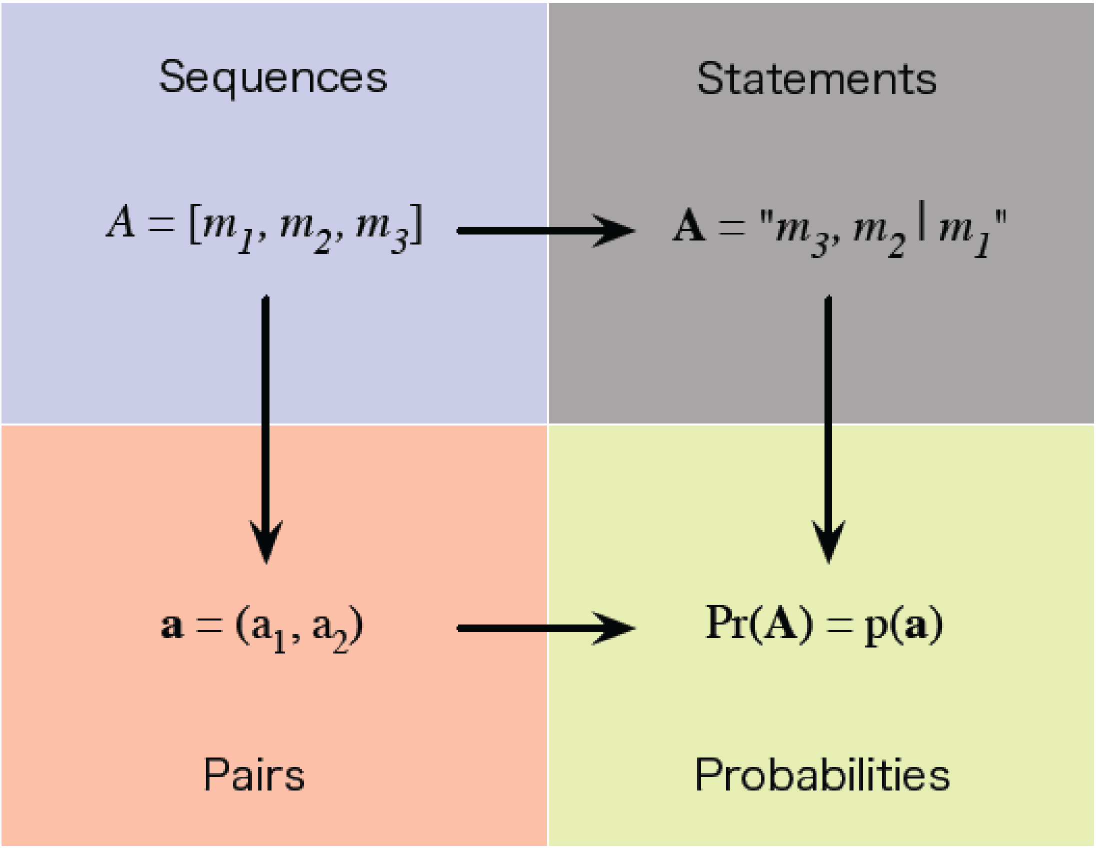


C A B

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 a_1 b_1 + \gamma_2 a_1 b_2 + \gamma_3 a_2 b_1 + \gamma_4 a_2 b_2 \\ \gamma_5 a_1 b_1 + \gamma_6 a_1 b_2 + \gamma_7 a_2 b_1 + \gamma_8 a_2 b_2 \end{pmatrix}$$

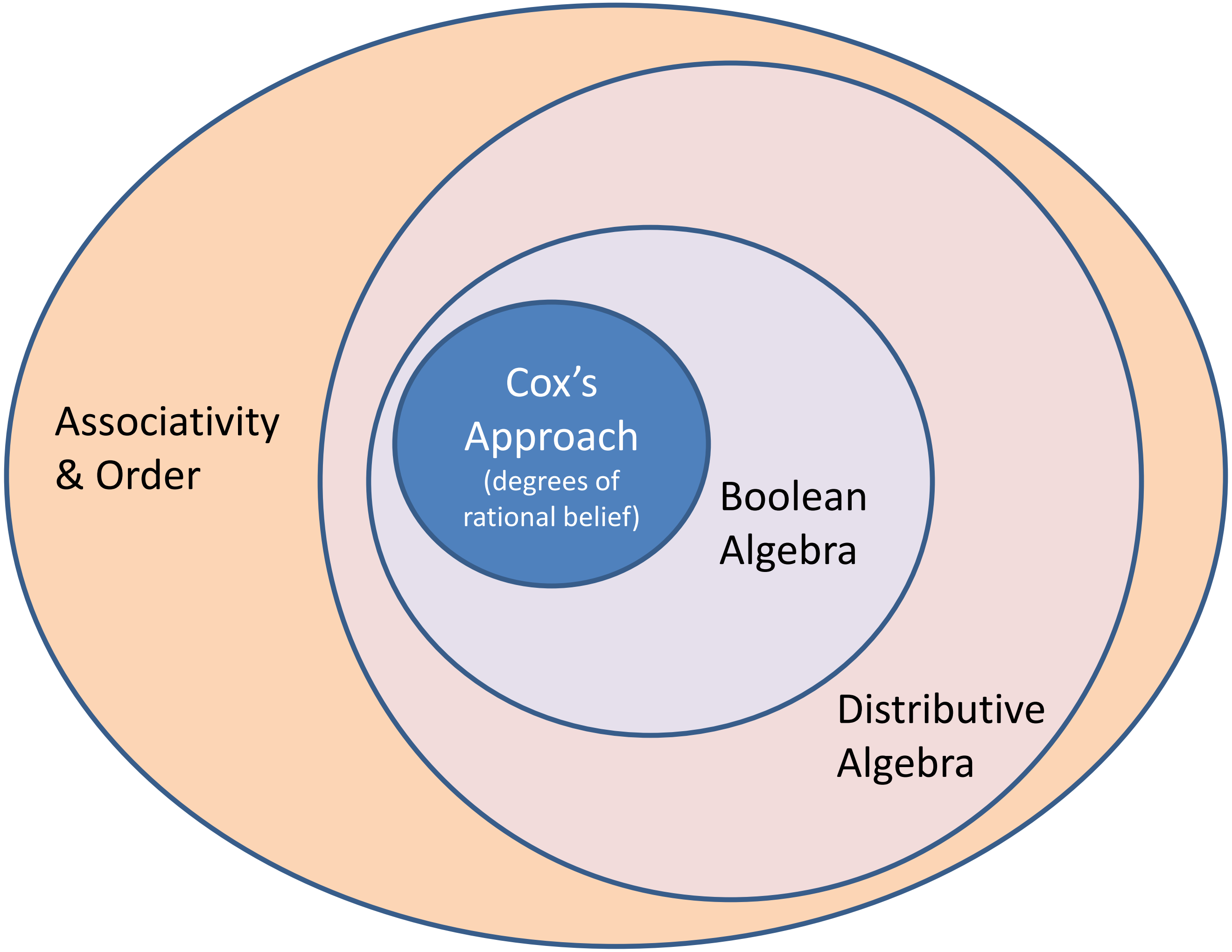
One can then show that the probabilities of measurement sequences are given by the Born Rule, where for $a = \binom{a_1}{a_2}, \mathbf{P(A)} = \mathbf{p(a)} = |a_1|^2 + |a_2|^2$




Sequences
Statements
$A = [m_1, m_2, m_3]$
$\mathbf{A} = "m_3, m_2 \mid m_1"$
$\mathbf{a} = (a_1, a_2)$
$\Pr(\mathbf{A}) = p(\mathbf{a})$
Pairs
Probabilities

Foundations are Important.

A solid foundation acts as a broad base on which theories can be constructed to unify seemingly disparate phenomena.



THANK YOU

Knuth K.H., Skilling J. 2012. Foundations of Inference. *Axioms* 1:38-73. arXiv:1008.4831 [math.PR]

Goyal P., Knuth K.H., Skilling J. 2010. Origin of complex quantum amplitudes and Feynman's rules, *Physical Review A* 81, 022109. arXiv:0907.0909v3 [quant-ph]