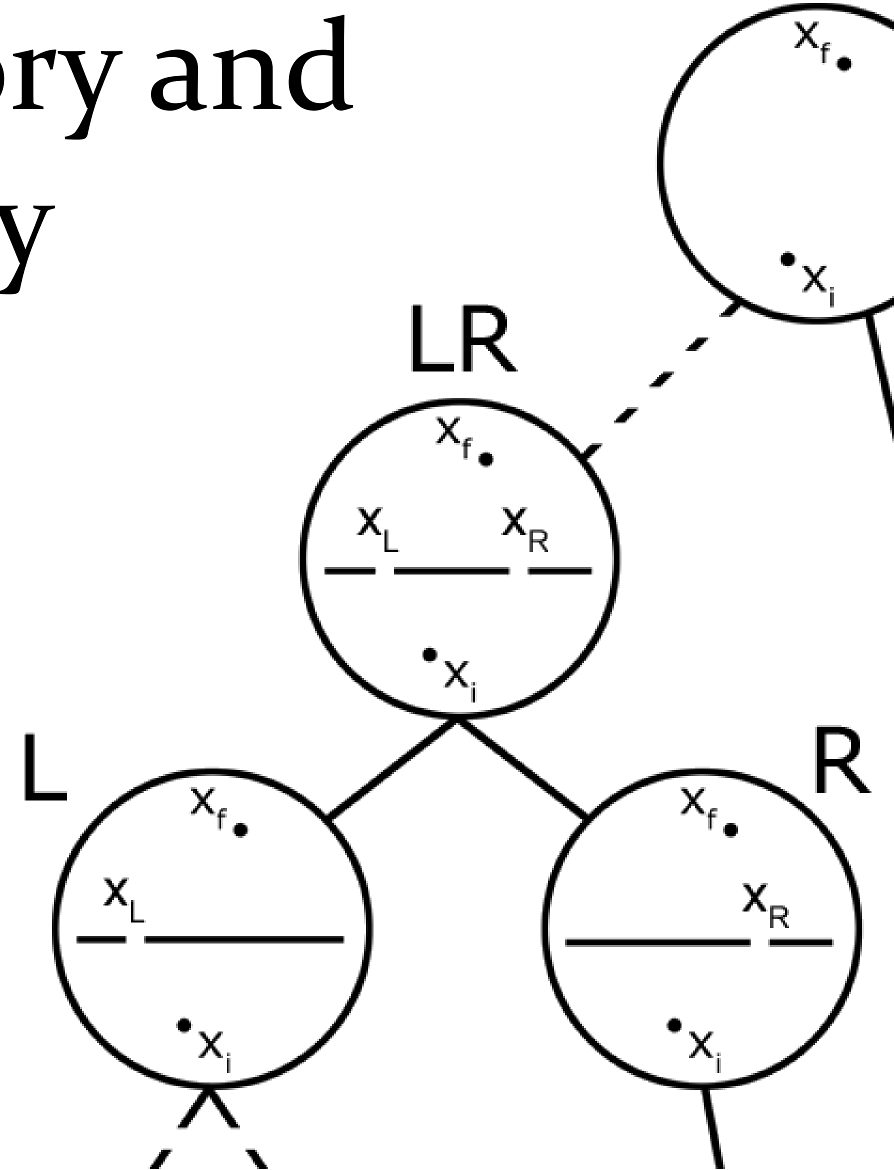
The Foundations of Probability Theory and Quantum Theory



Kevin H. Knuth

Departments of Physics and Informatics
University at Albany (SUNY)

Partially Ordered Sets

A partially ordered set is a set along with a binary ordering relation, eg. $S \geq L$

S

L R

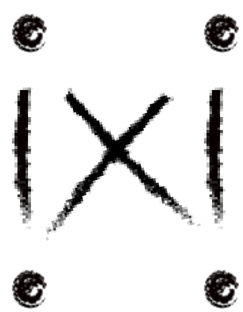
Parts of a Bridge

Photograph by Barbara Maddrell, National Geographic Image Collection

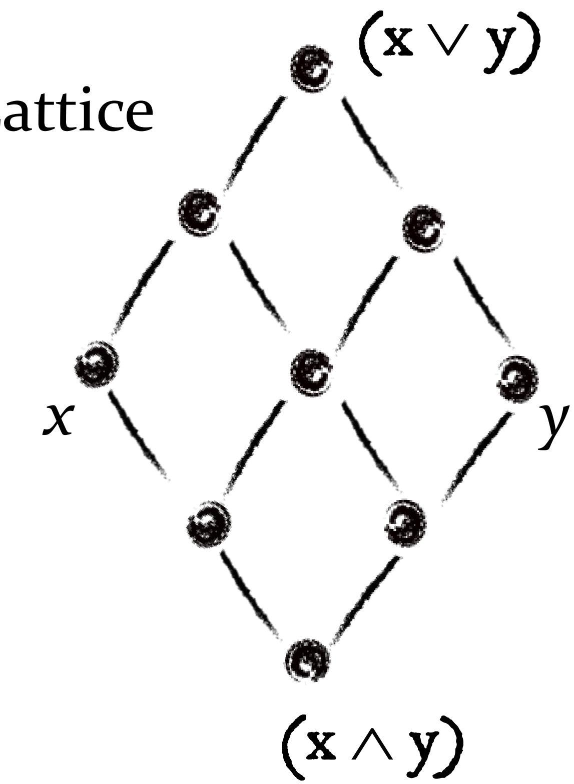
Posets versus Lattices

Lattices are posets where every join and meet is unique.

Poset



Lattice

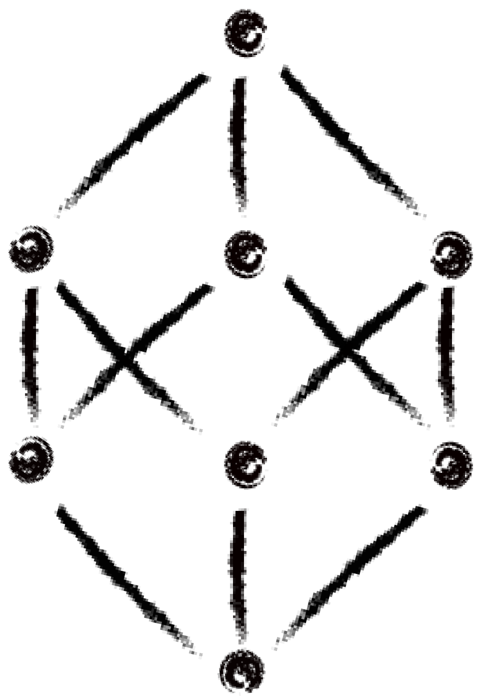


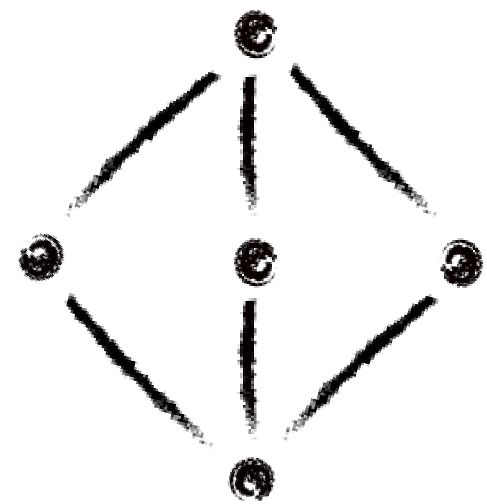
Chains and Antichains

Chains are totally ordered

Antichains are unordered

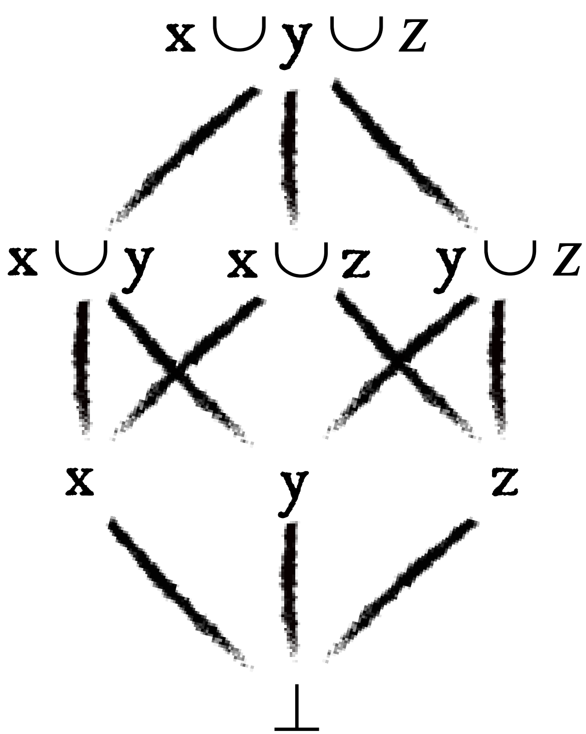
Examples



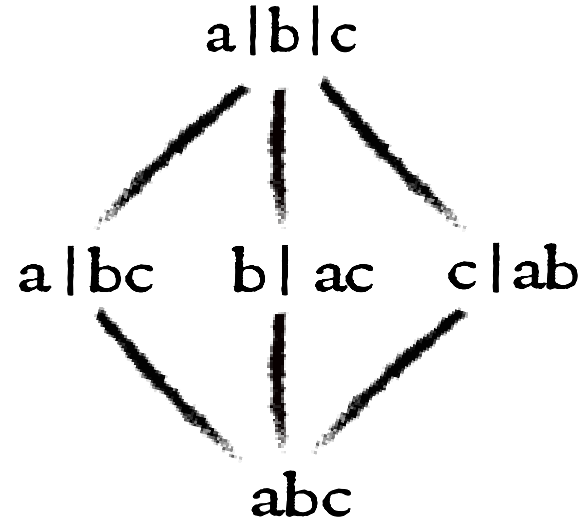


More Complex Examples

Subset Inclusion



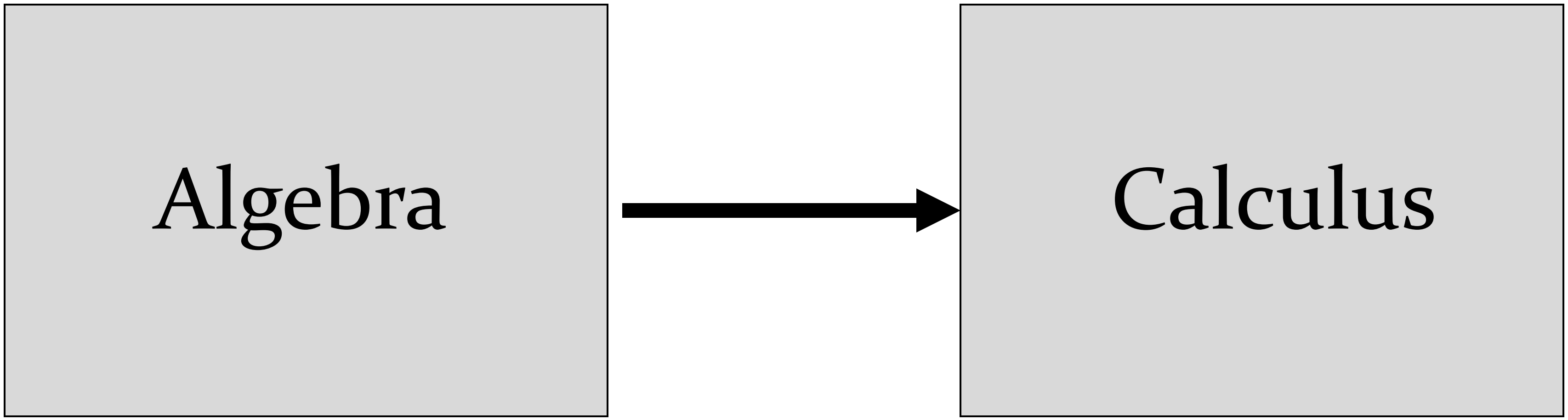
Partitions



Lattices are Algebras

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

Qualitative to Quantitative



Classical Inference

States

apple

banana

cherry

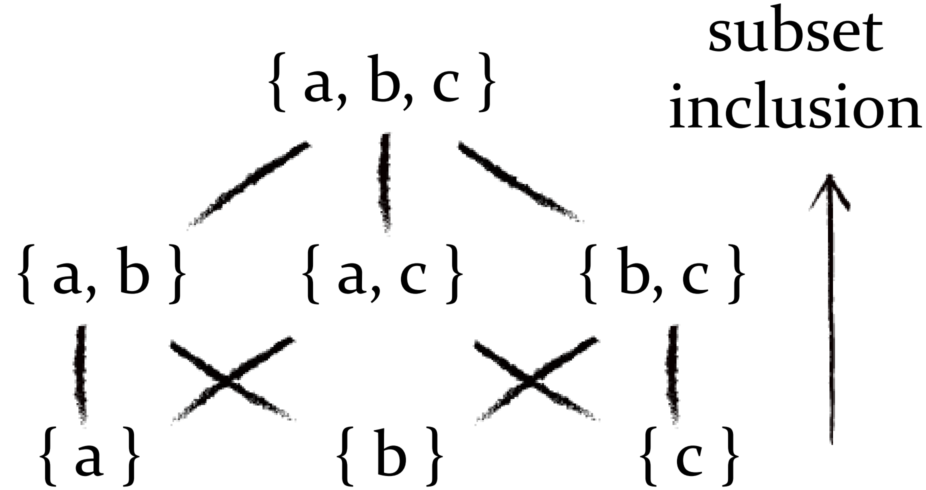
states of the contents of
my grocery basket

Statements

a b c

states of the contents of
my grocery basket

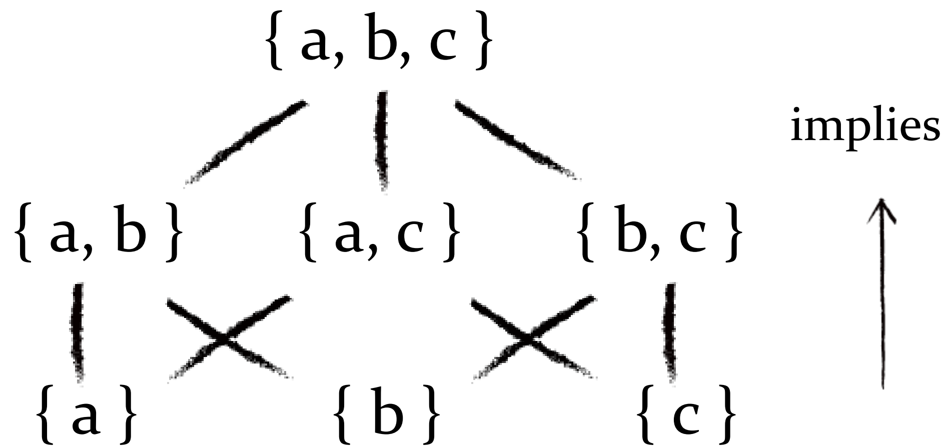
powerset



statements
about the contents of
my grocery basket

statements describe potential states

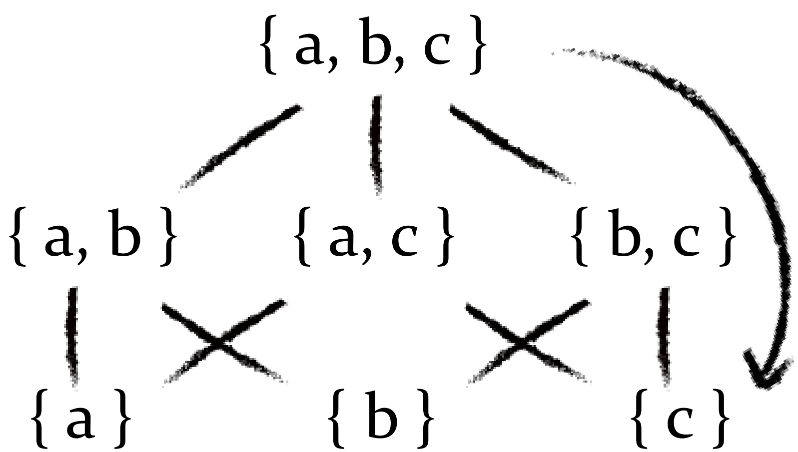
Implication



statements
about the contents of
my grocery basket

ordering encodes implication

Inference



statements
about the contents of
my grocery basket

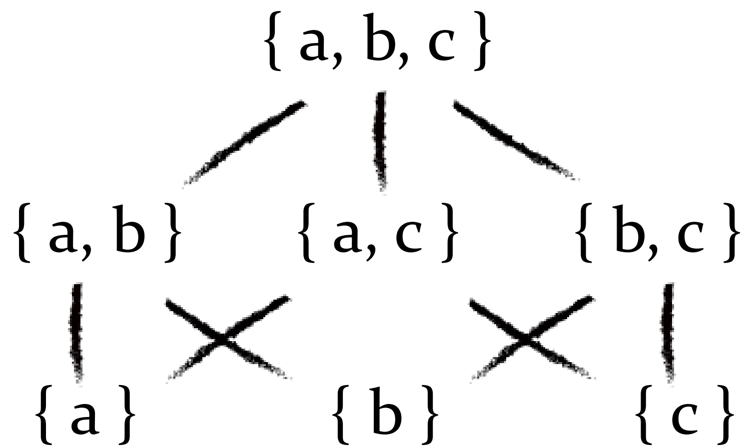
Quantify to what degree knowing that the system is in one of three states {a, b, c} implies knowing that it is in some other set of states

inference works backwards

Quantification

Quantification

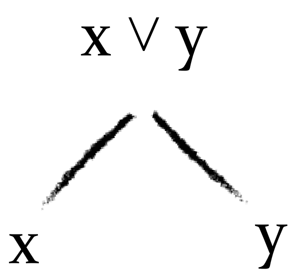
Quantify the partial order by assigning real numbers to the elements



Any quantification must be consistent with the lattice structure. Otherwise, information about the partial order is lost.

Local Consistency

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

$$f(x \vee y) \leftrightarrow f(x) \text{ and } f(y)$$

This implies that:

$$f(x \vee y) = S[f(x), f(y)]$$

where S is an unknown function
to be determined.

Associativity of Join

Write the same element two different ways

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

which implies

$$S[f(x), S[f(y), f(z)]] \quad = \quad S[S[f(x), f(y)], f(z)]$$

Note that the unknown function S is nested in two distinct ways, which reflects associativity

Associativity Equation

$$S[f(x), S[f(y), f(z)]] = S[S[f(x), f(y)], f(z)]$$

The general solution (Aczel 1966; Knuth & Skilling 2012) is:

$$F(S[f(x), f(y)]) = F(f(x)) + F(f(y))$$

where F is an arbitrary function.
Define $v(x) = F(f(x))$ so that we have straightforward summation.

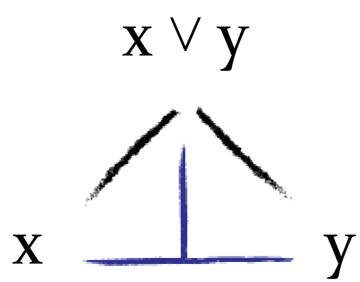
$$v(x \vee y) = v(x) + v(y)$$

Derivation of the Summation Axiom in Measure Theory

Valuation

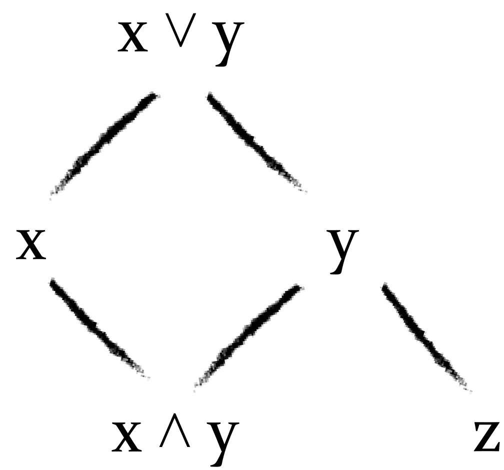
VALUATION $v : x \in L \quad \to \quad R$

If $y \geq x$ then $v(y) \geq v(x)$

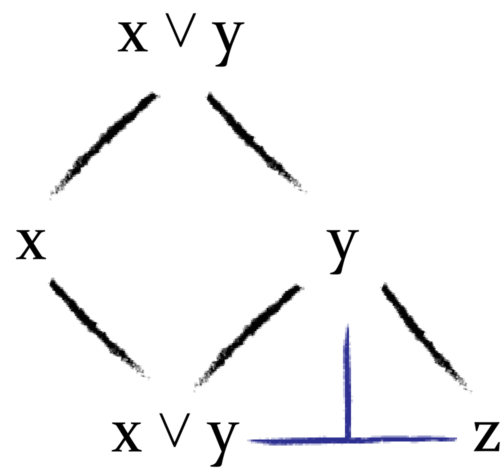


$$v(x \vee y) \quad = \quad v(x) + v(y)$$

General Case

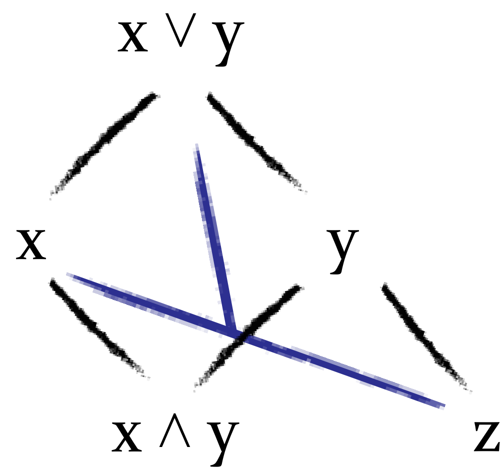


General Case



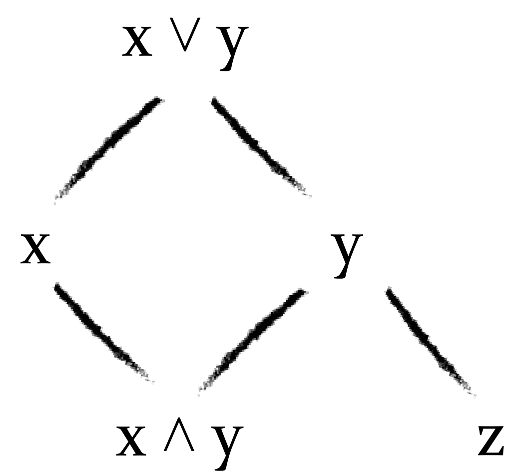
$$v(y) \quad = \quad v(x \wedge y) + v(z)$$

General Case



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

General Case

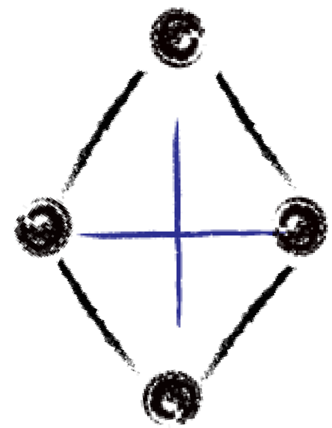


$$v(y) = v(x \wedge y) + v(z) \qquad v(x \vee y) = v(x) + v(z)$$

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

Sum Rule

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



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

symmetric form (self-dual)

Sum Rule

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

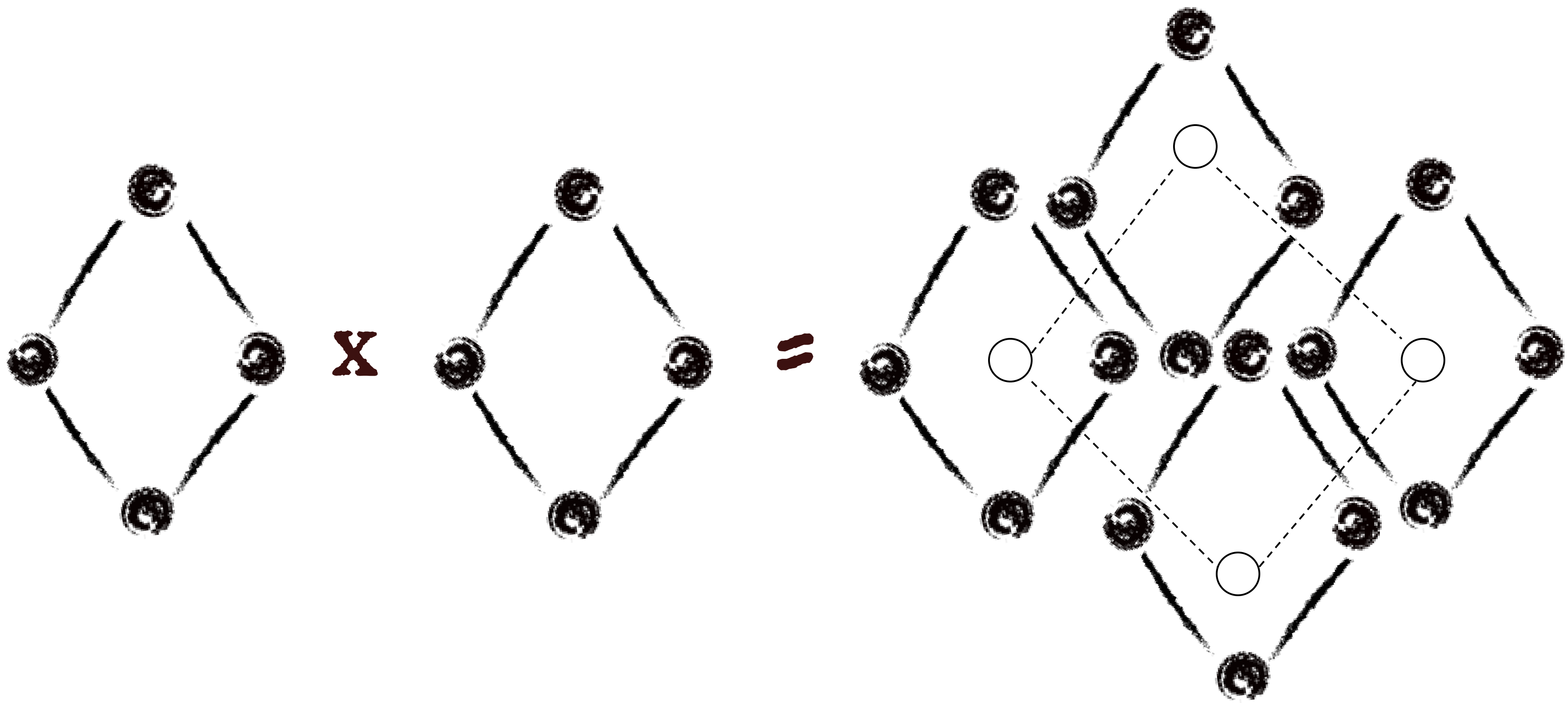
$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$\max(x,y) = x + y - \min(x,y)$$

$$\chi = V - E + F$$

$$\log(\gcd(x,y)) = \log(x) + \log(y) - \log(\mathrm{lcm}(x,y))$$

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

$$v((a, b)) \quad = \quad v(a)\, v(b)$$

which is again summation (after taking the logarithm)

Context and Bi-Valuations

BI-VALUATION $w : x, i \in L \rightarrow \mathbb{R}$

Bi-Valuation				Valuation
$w(x \mid i)$	$\longrightarrow$	$v_i(x)$	$\longrightarrow$	$v(x)$
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 Explicit

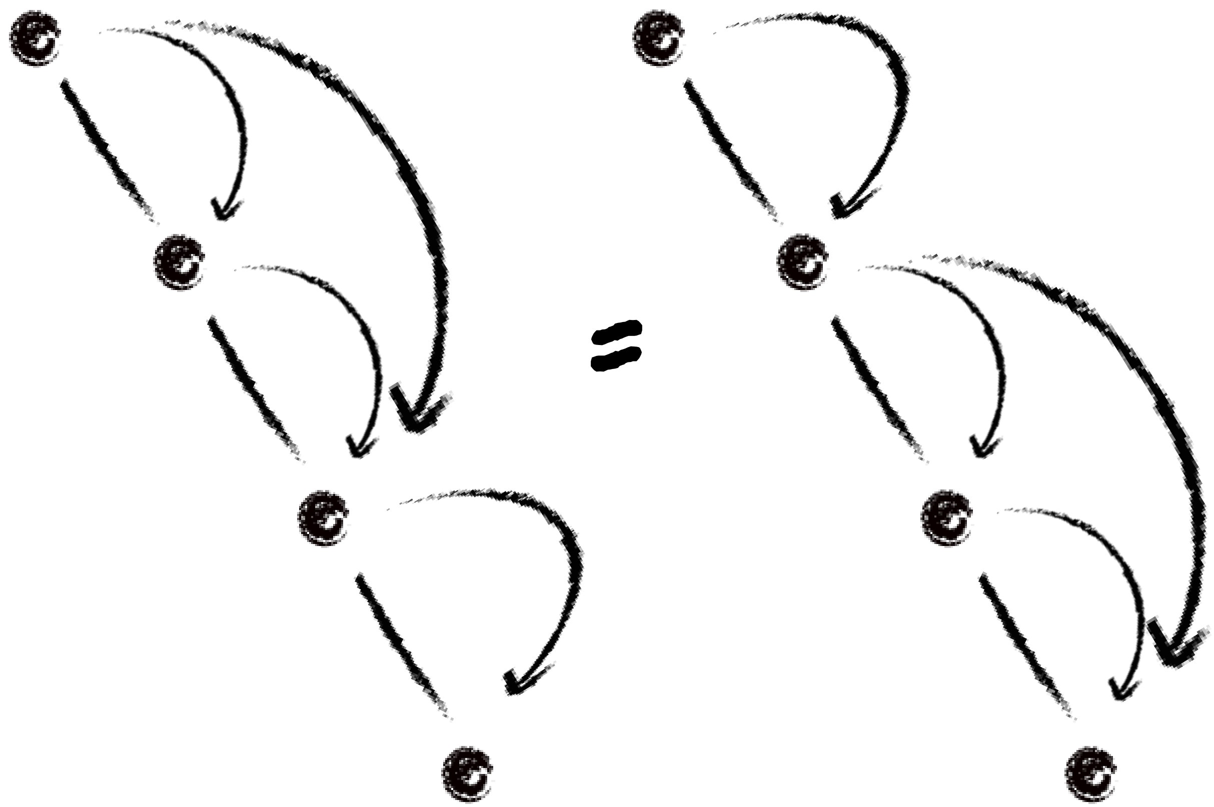
Sum Rule

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

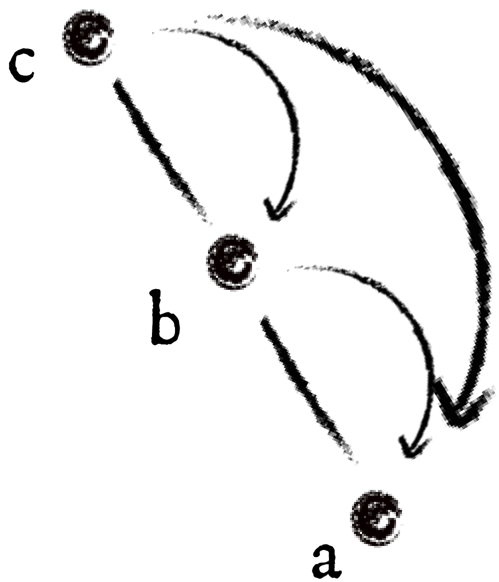
Direct Product Rule

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

Associativity of Context



Chain Rule

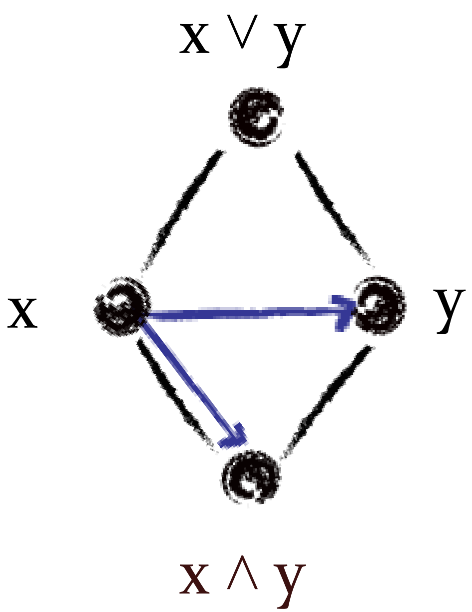


$$w(a|c) \quad = \quad w(a|b)\, w(b|c)$$

Lemma

$$w(x|x) + w(y|x) = w(x \vee y|x) + w(x \wedge y|x)$$

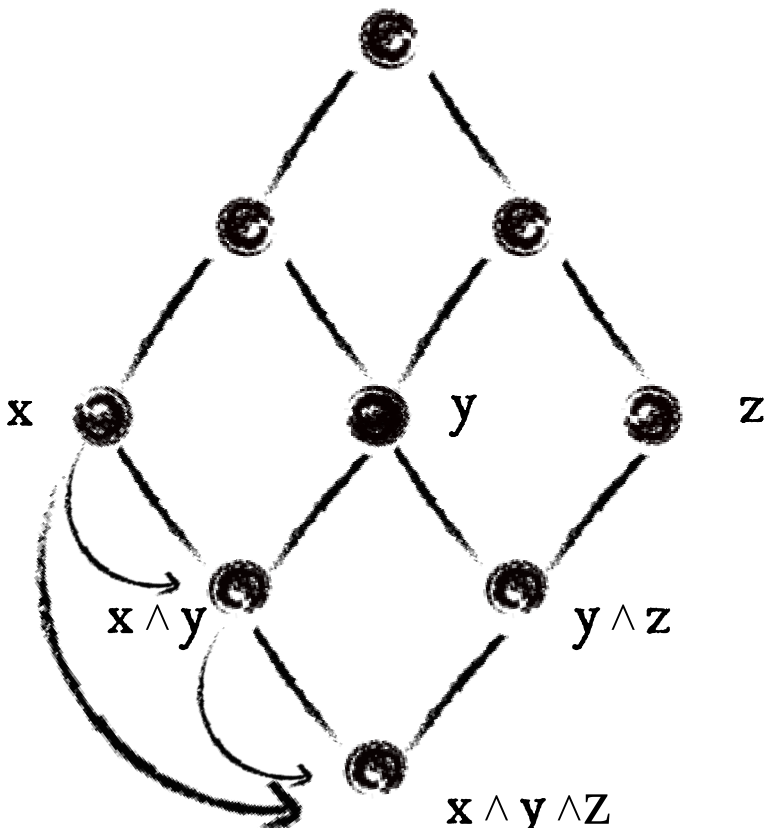
Since $x \leq x$ and $x \leq x \vee y$, $w(x \mid x) = 1$ and $w(x \vee y \mid x) = 1$



$$w(y|x) = w(x \wedge y|x)$$

Extending the Chain Rule

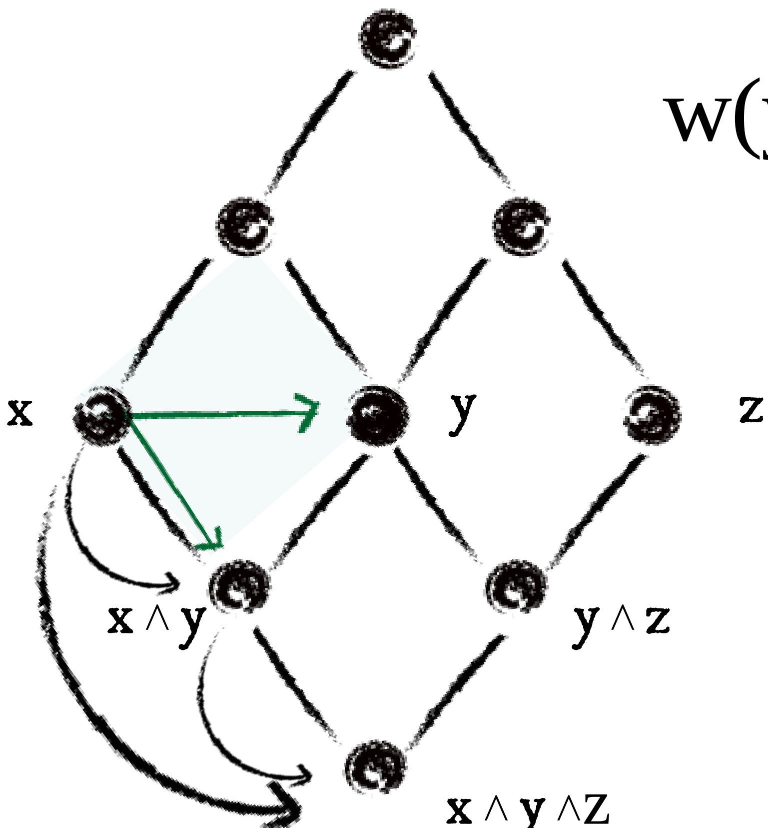
$$w(x \wedge y \wedge z \mid x) = w(x \wedge y \mid x)\,w(x \wedge y \wedge z \mid x \wedge y)$$



Extending the Chain Rule

$$w(x \wedge y \wedge z \mid x) = w(x \wedge y \mid x) w(x \wedge y \wedge z \mid x \wedge y)$$

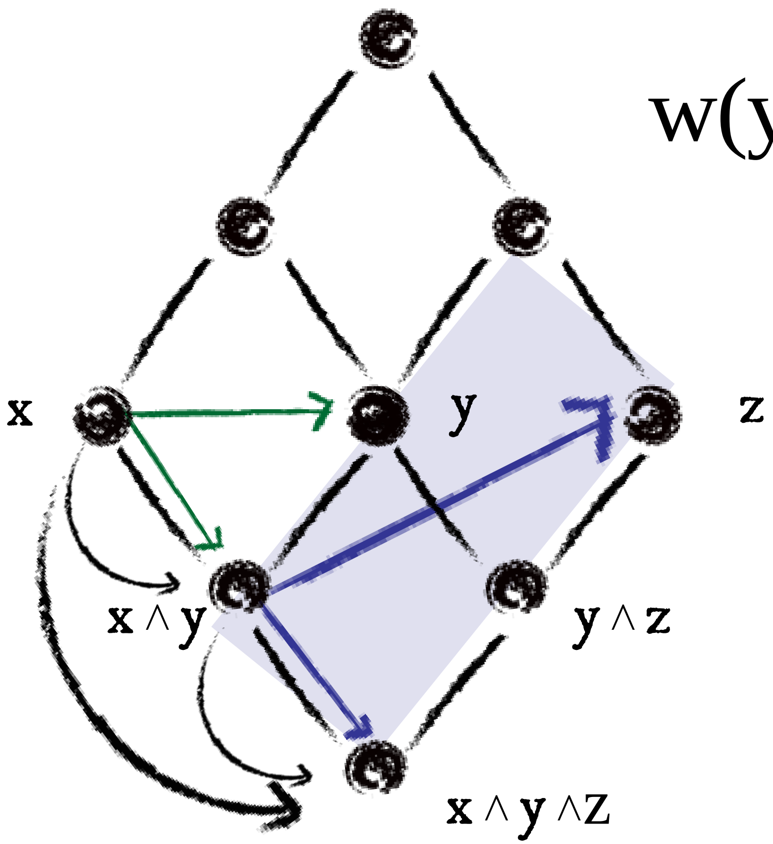
$$w(y \wedge z \mid x) = w(y \mid x) w(z \mid x \wedge y)$$



Extending the Chain Rule

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

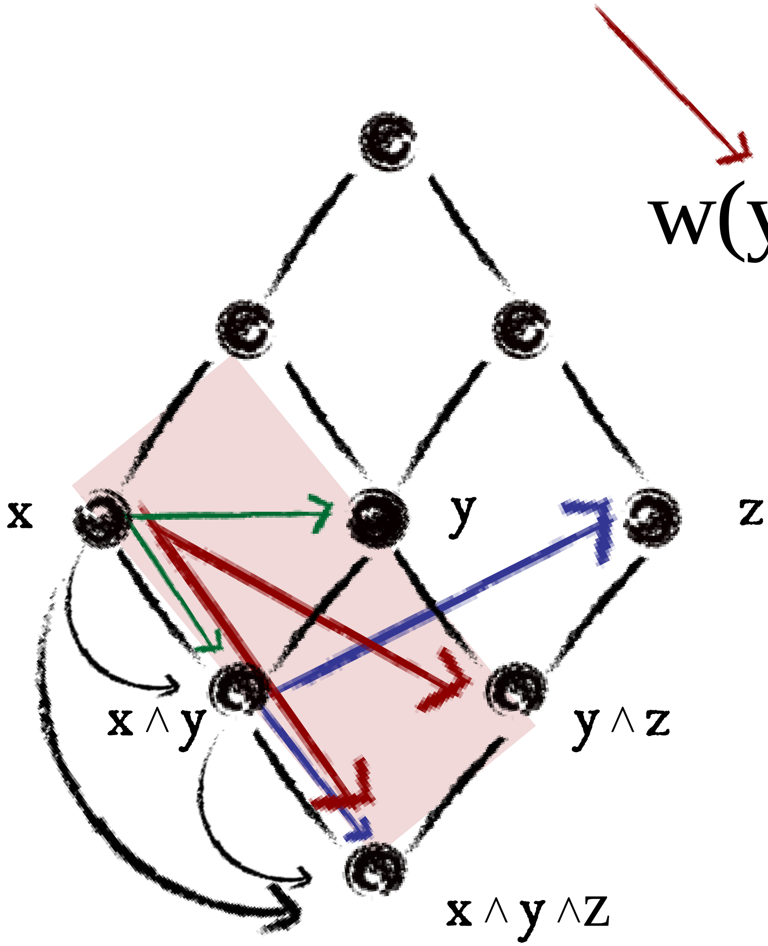
$$w(y \wedge z \mid x) = w(y \mid x)\, w(z \mid x \wedge y)$$



Extending the Chain Rule

$$w(x \wedge y \wedge z \mid x) = w(x \wedge y \mid x) w(x \wedge y \wedge z \mid x \wedge y)$$

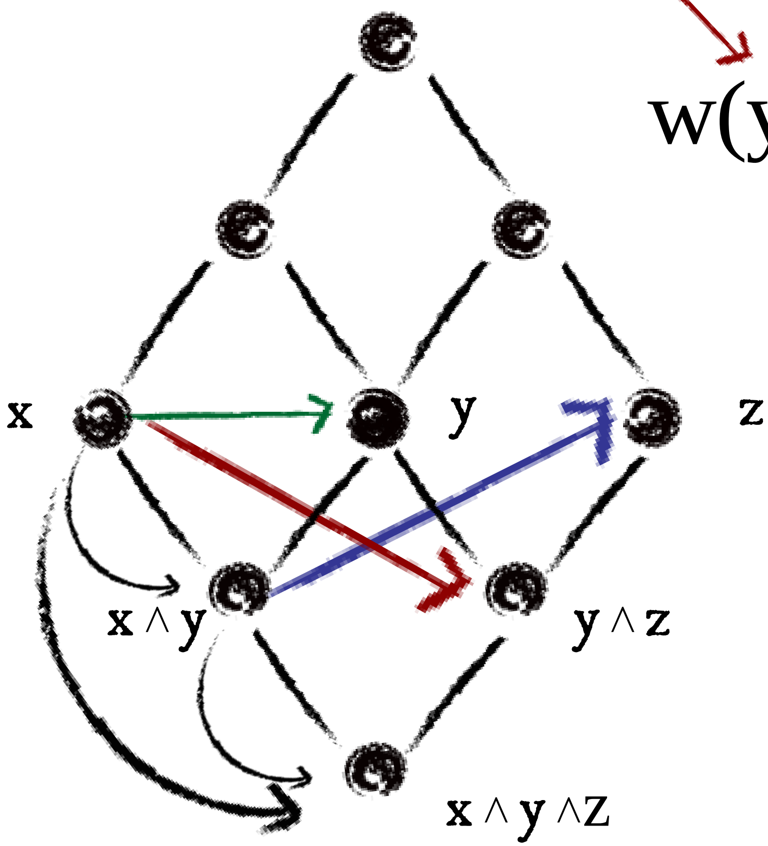
$$w(y \wedge z \mid x) = w(y \mid x) w(z \mid x \wedge y)$$



Extending the Chain Rule

$$w(x \wedge y \wedge z \mid x) = w(x \wedge y \mid x)w(x \wedge y \wedge z \mid x \wedge y)$$

$$w(y \wedge z \mid x) = w(y \mid x)w(z \mid x \wedge y)$$



Constraint Equations

Sum Rule

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

Direct Product Rule

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

Product Rule

$$w(y \wedge z \mid x) = w(y \mid x) w(z \mid x \wedge y)$$

Bayes Theorem

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

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

$$\downarrow$$

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

Bayes Theorem involves a change of context.

Bayesian Probability Theory

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

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

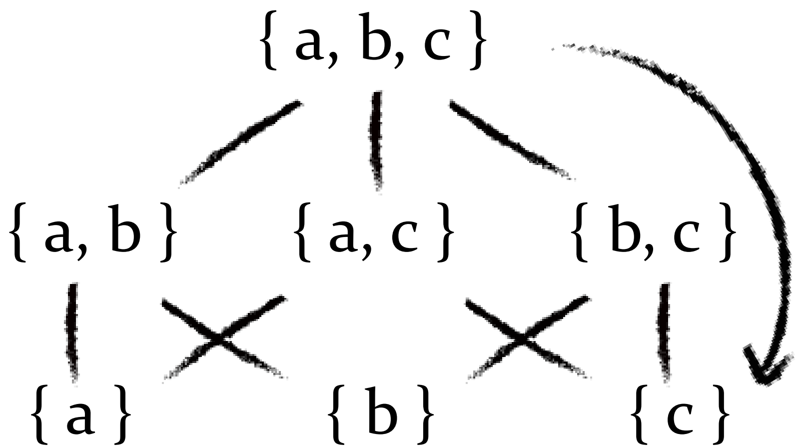
Product Rule

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

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)

Quantum Inference

The Goal of QM

Here we focus on experimental setups
that lead to measurement sequences

The ultimate goal is to compute the probability of
statements about these measurement sequences

The QMical Question

To what degree does

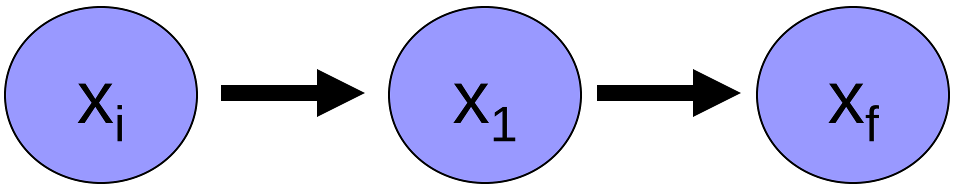
“knowing that we observe any one of a large set of possible measurement sequences”

imply that

“we know that we observe a particular measurement sequence”?

Measurement Sequences

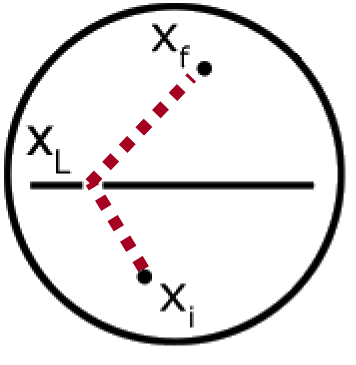
Three measurements are made in succession with outcomes x_i, x_1, x_f, respectively



$$A = [x_i, x_1, x_f]$$

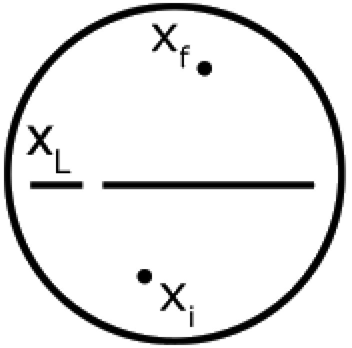
There is no explicit notion of time,
only an notion of an order in which events occur.

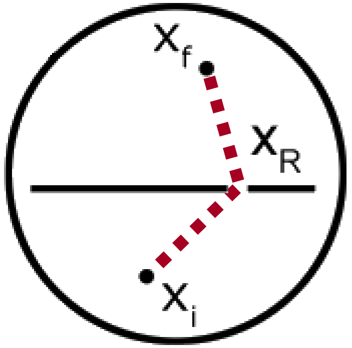
Slit Experiment



$A = [x_i, x_L, x_f]$

Slit Experiment

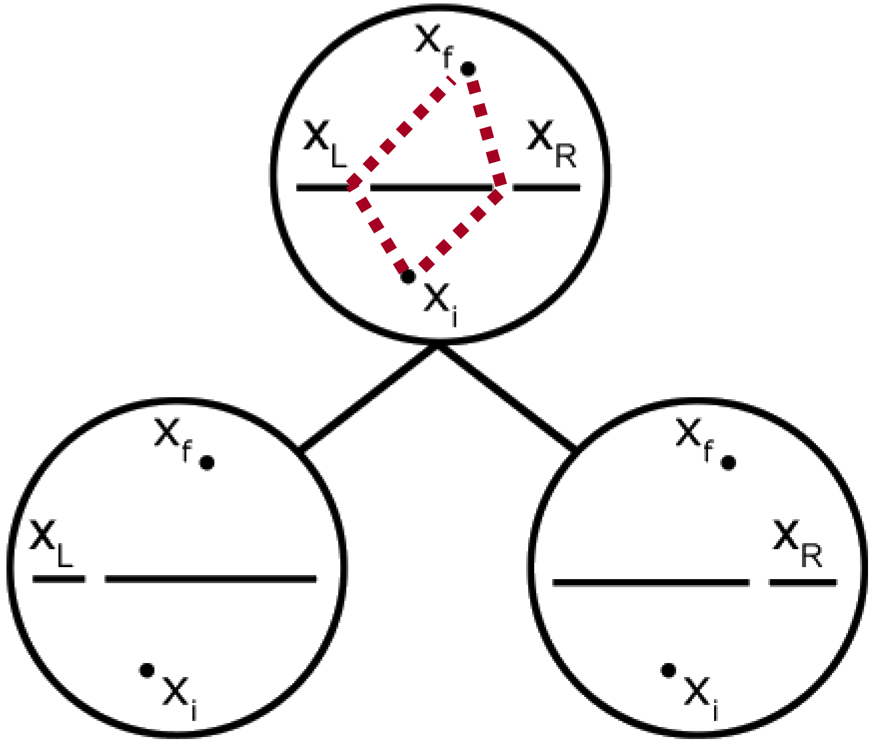




$A = [x_i, x_L, x_f]$ $B = [x_i, x_R, x_f]$

Relating Measurement Sequences

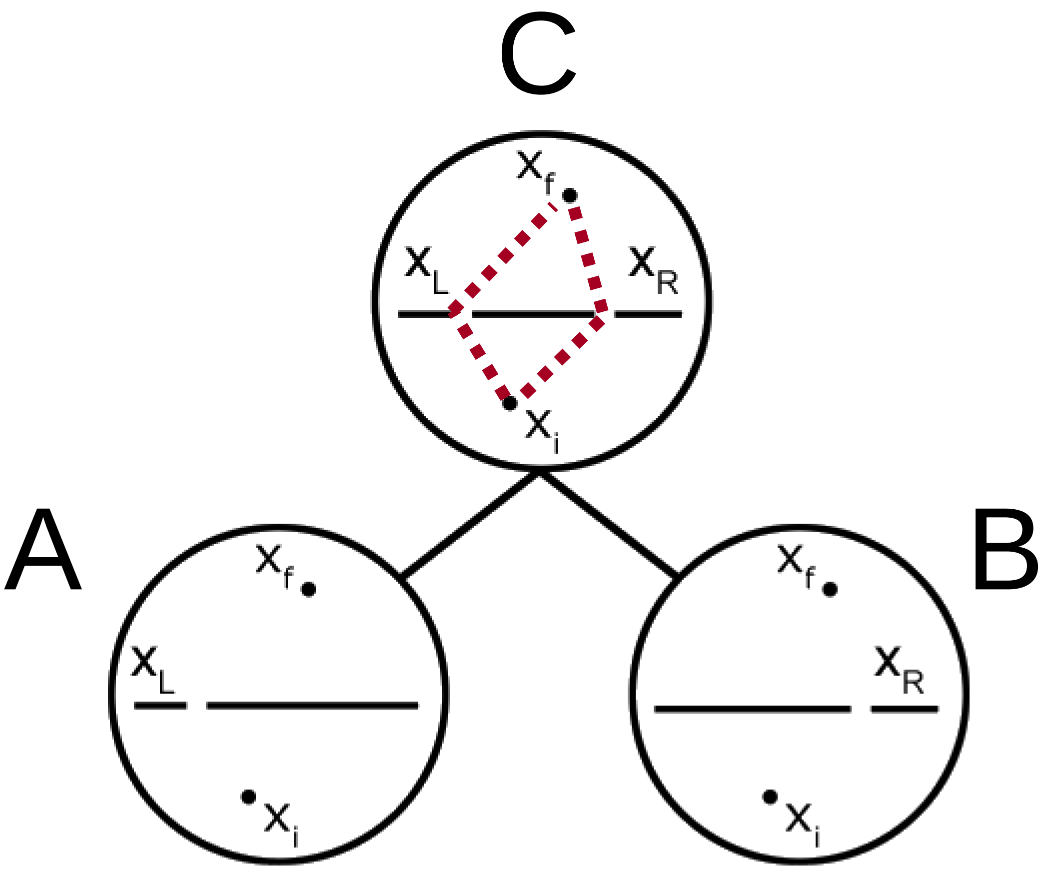
$C = [x_i, (x_L, x_R), x_f]$



$A = [x_i, x_L, x_f]$

$B = [x_i, x_R, x_f]$

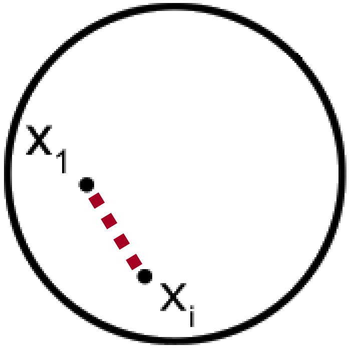
Parallel Combination



$A \vee B = C$

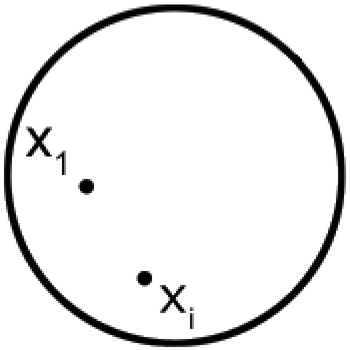
Combining in Series

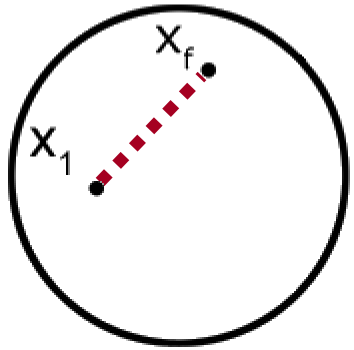
$A = [x_i, x_1]$



Combining in Series

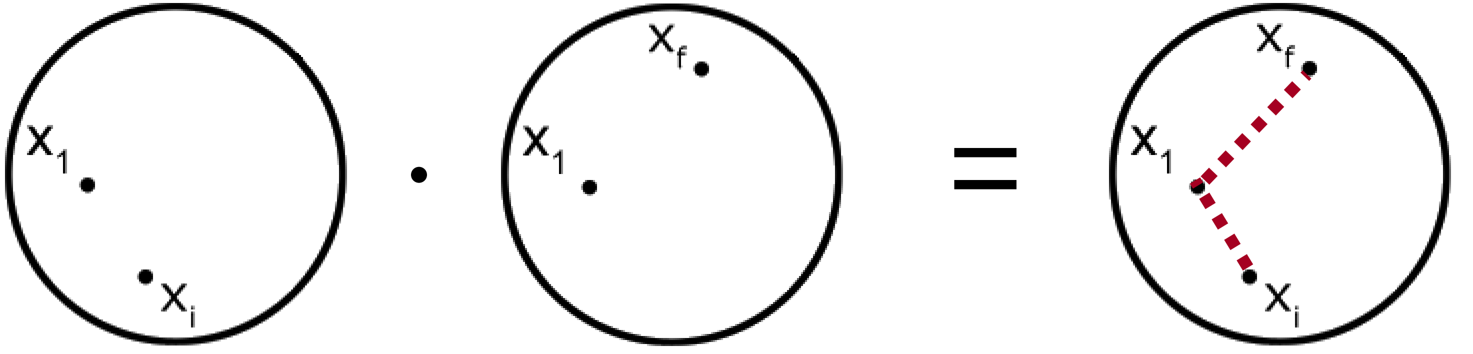
$A = [x_i, x_1]$ $B = [x_1, x_f]$



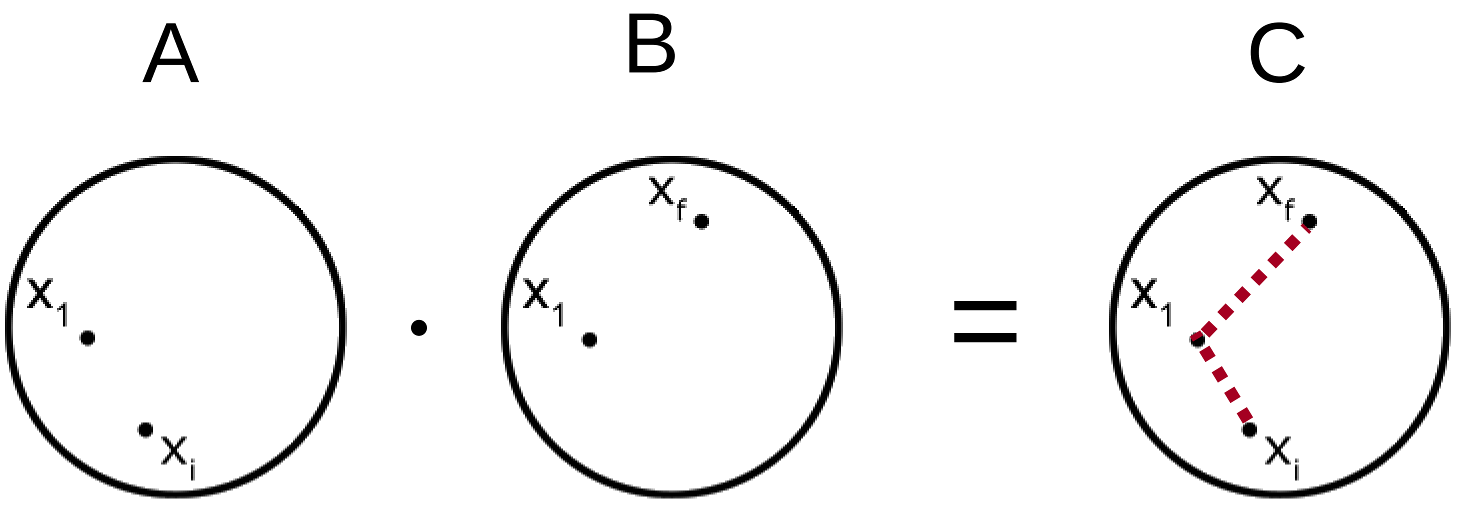


Combining in Series

$A = [x_i, x_1]$ $B = [x_1, x_f]$ $C = [x_i, x_1, x_f]$



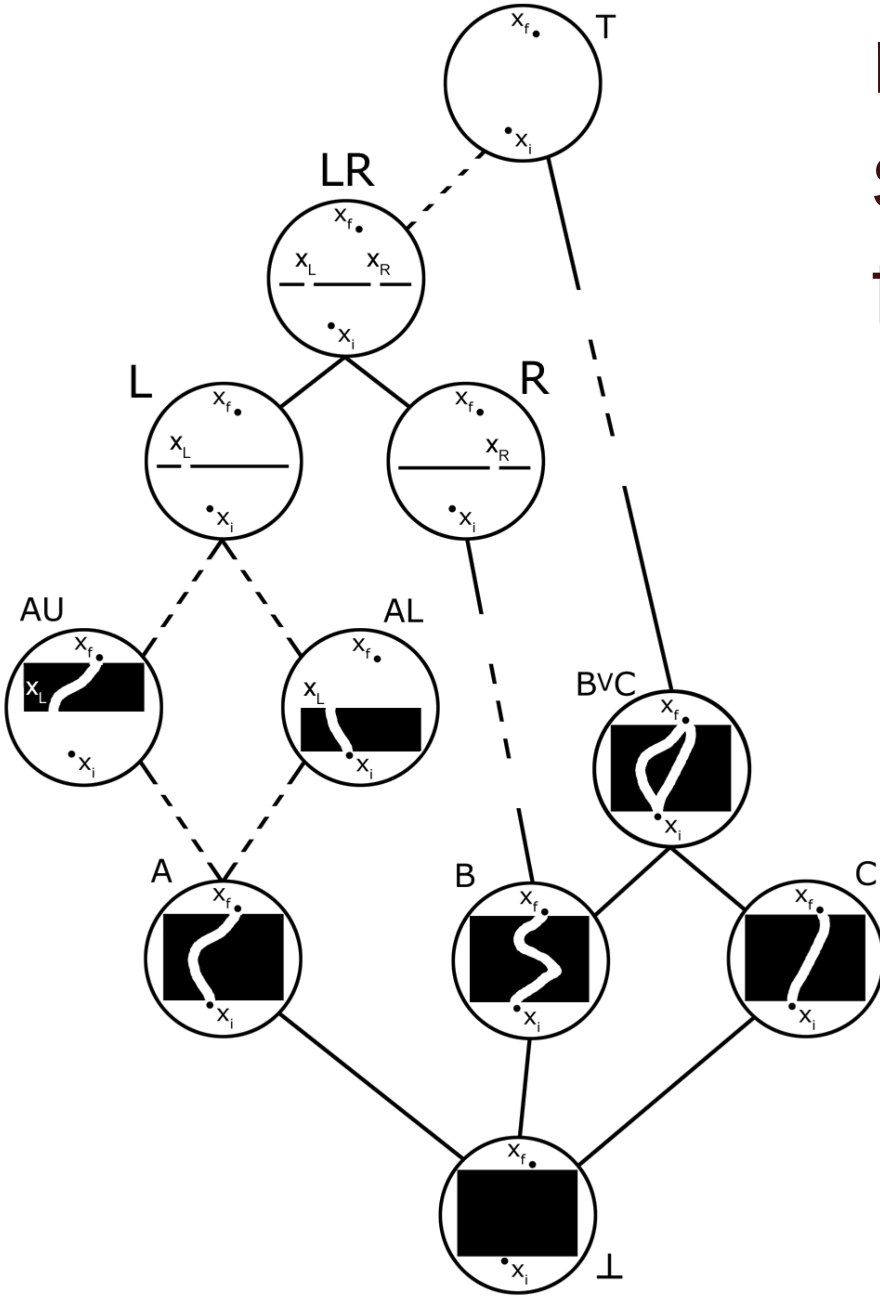
Series Combination



$$A \cdot B = C$$

Algebraic Relations

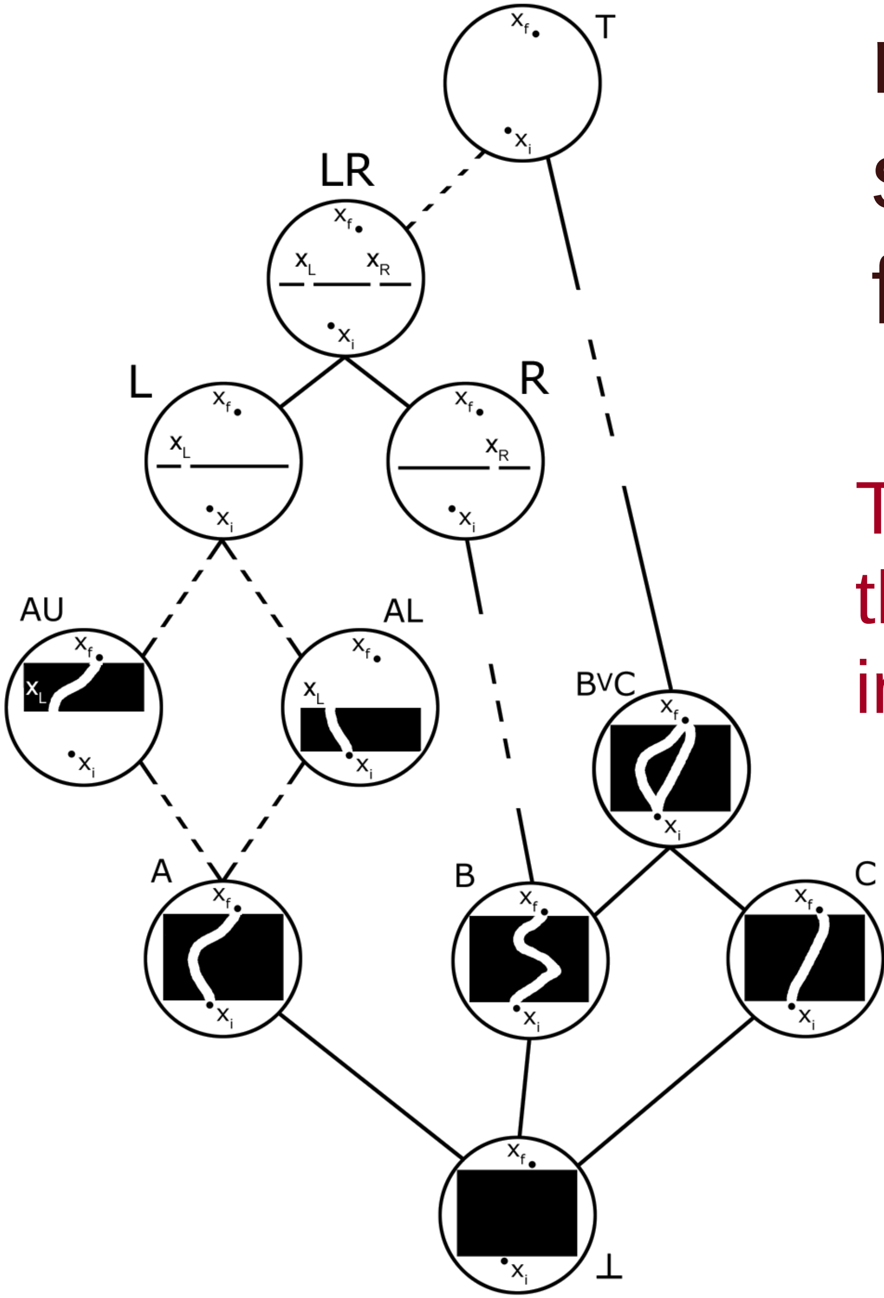
$A \vee B = B \vee A$	Commutativity of $\vee$
$A \vee (B \vee C) = (A \vee B) \vee C$	Associativity of $\vee$
$A \cdot (B \cdot C) = (A \cdot B) \cdot C$	Associativity of $\cdot$
$A \cdot (B \vee C) = (A \cdot B) \vee (A \cdot C)$ $(B \vee C) \cdot A = (B \cdot A) \vee (C \cdot A)$	Distributivity $\cdot$ over $\vee$



measurement sequences form a poset

Think of slits as filters

Knuth 2003

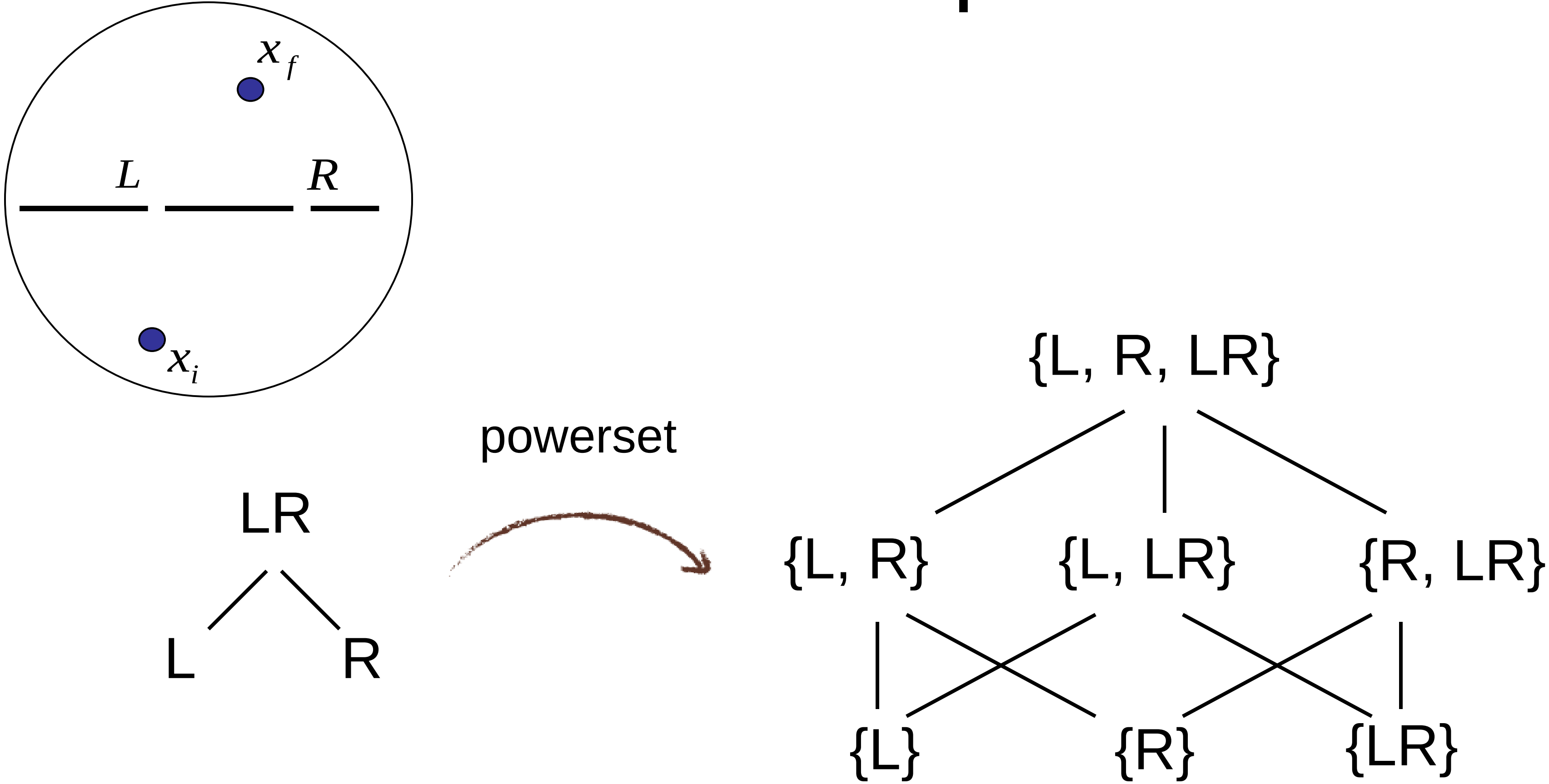


measurement sequences form a poset

This is in contrast to the antichain of states in classical inference

Think of slits as filters
Knuth 2003

Double-Slit Experiment



Quantification

The Pair Postulate

Each sequence of measurement outcomes obtained in a given experiment is represented by a pair of real numbers, where the probability associated with this sequence is a continuous, nontrivial function of both components of this real number pair.

$$A \quad \rightarrow \quad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Why Pairs?

The goal is to quantify the poset of measurement sequences.

A scalar quantification simply ranks elements on a one-dimensional scale.

This will be done with statements when we compute probabilities. But for now, we wish to maintain some of the richness of the poset structure.

Pairs Represent Sequences

Sequences

$A \vee B = B \vee A$

$A \vee (B \vee C) = (A \vee B) \vee C$

$A \cdot (B \cdot C) = (A \cdot B) \cdot C$

$A \cdot (B \vee C) = (A \cdot B) \vee (A \cdot C)$

$(B \vee C) \cdot A = (B \cdot A) \vee (C \cdot A)$

Pairs

$a \oplus b = b \oplus a$

$a \oplus (b \oplus c) = (a \oplus b) \oplus c$

$a \odot (b \odot c) = (a \odot b) \odot c$

$a \odot (b \oplus c) = (a \odot b) \oplus (a \odot c)$

$(b \oplus c) \odot a = (b \odot a) \oplus (c \odot a)$

Associativity of ⊕

$$a \oplus (b \oplus c) = (a \oplus b) \oplus c$$

↓ Aczél and Hosszú 1956

$$F(a \oplus b) = F(a) + F(b)$$

Sum Rule for Pairs

Without any loss of generality

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \oplus \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

The only freedom left is a real invertible linear transform

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} S & T \\ U & V \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

with $SV - TU \neq 0$

Distributivity of $\odot$ over $\oplus$

Using the sum rule and distributivity

$(a + b) \odot c = (a \odot c) + (b \odot c)$

$a \odot (b + c) = (a \odot b) + (a \odot c)$

$a \odot b$ has a bilinear multiplicative form

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

Associativity of $\odot$

Using the associativity
along with our freedom to apply a linear transform

C1 $$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$$

C2 $$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$$

C3 $$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix}$$

N1 $$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \end{pmatrix}$$

N2 $$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_1 \end{pmatrix}$$

Probability of a Sequence

We will map these sequences to statements about sequences, and in doing so, will assign a probability to each sequence of measurements.

$$P(A) = \Pr(m_n, m_{n-1}, \ldots, m_2 \mid m_1)$$

Our Pair Postulate dictates that this be a continuous real-valued function of the pair that depends non-trivially on both components of the pair.

$$P(A) = p(a)$$

Furthermore, since we do not at the outset indicate which component of the pair is which, there must exist a representation such that the probability is symmetric with respect to pair-interchange

$$p(a_1, a_2) = p(a_2, a_1)$$

Direct Product

Consider $A=[m_1,m_2]$ and $B=[m_2,m_3]$ so that $C=[m_1,m_2,m_3]$ and

$$P(C) = \Pr(m_3, m_2 \mid m_1)$$

By the product rule of probability

$$P(C) = \Pr(m_3 \mid m_2, m_1)\,\Pr(m_2 \mid m_1)$$

Measurement m_2 overrides all information obtained from m_1 so that

$$P(C) = \Pr(m_3 \mid m_2)\,\Pr(m_2 \mid m_1)$$
$$= P(B)\,P(A)$$

Therefore $p(a \odot b) = p(a)\, p(b)$

Results from Probability

C1 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$ $p(a) = ({a_1}^2 + {a_2}^2)^{\frac{\alpha}{2}}$

C2 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$ $p(a) = |a_1|^\alpha e^{\beta \frac{a_2}{a_1}}$

C3 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix}$ $p(a) = |a_1|^\alpha |a_2|^\beta$

N1 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \end{pmatrix}$ $p(a) = |a_1|^\alpha$

N2 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_1 \end{pmatrix}$ $p(a) = |a_1|^\alpha$

Results from Probability

C1 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$ $p(a) = ({a_1}^2 + {a_2}^2)^{\frac{\alpha}{2}}$

C2 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$ $p(a) = |a_1|^\alpha e^{\beta \frac{a_2}{a_1}}$

C3 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix}$ $p(a) = |a_1|^\alpha |a_2|^\beta$

N1 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \end{pmatrix}$ $p(a) = |a_1|^\alpha$

N2 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_1 \end{pmatrix}$ $p(a) = |a_1|^\alpha$

symmetry rules out three cases

Summation

Clearly, when performing measurements in parallel, probabilities do not sum in general. However, if these measurements are sufficiently disjoint, one would expect that **at least sometimes**, probabilities should sum.

This rules out C3 and fixes α=2 in case C1 so that

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix} \qquad p(a) = {a_1}^2 + {a_2}^2$$

Manipulating Amplitude Pairs

Given that quantum measurement sequences are quantified by pairs of real numbers, obey the requisite symmetries of combination in series and parallel and are consistent with probabilities of statements about sequences. we have derived

Feynman Rules

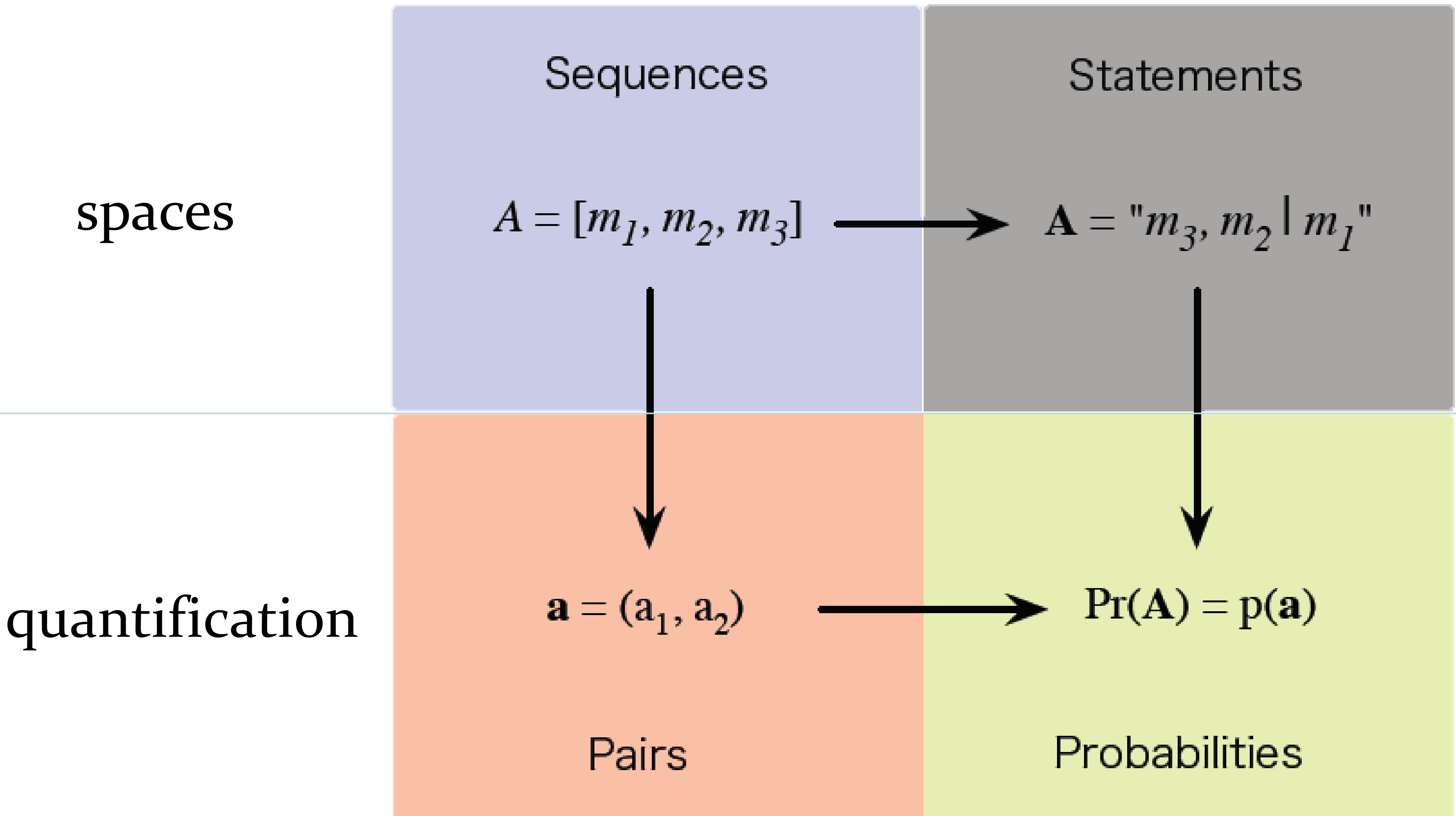
$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \oplus \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$$

Born Rule

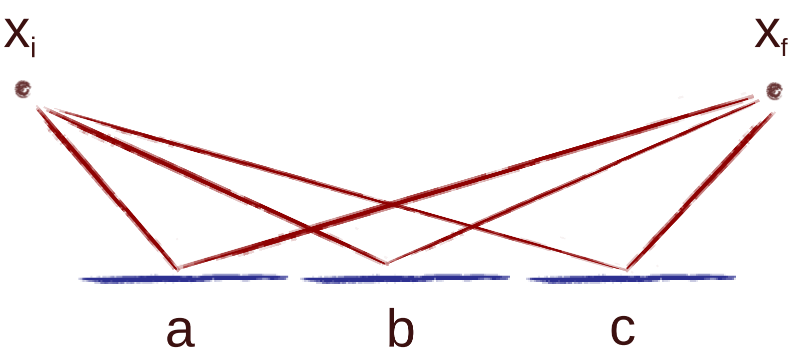
$$p(a) = {a_1}^2 + {a_2}^2$$

Relationships

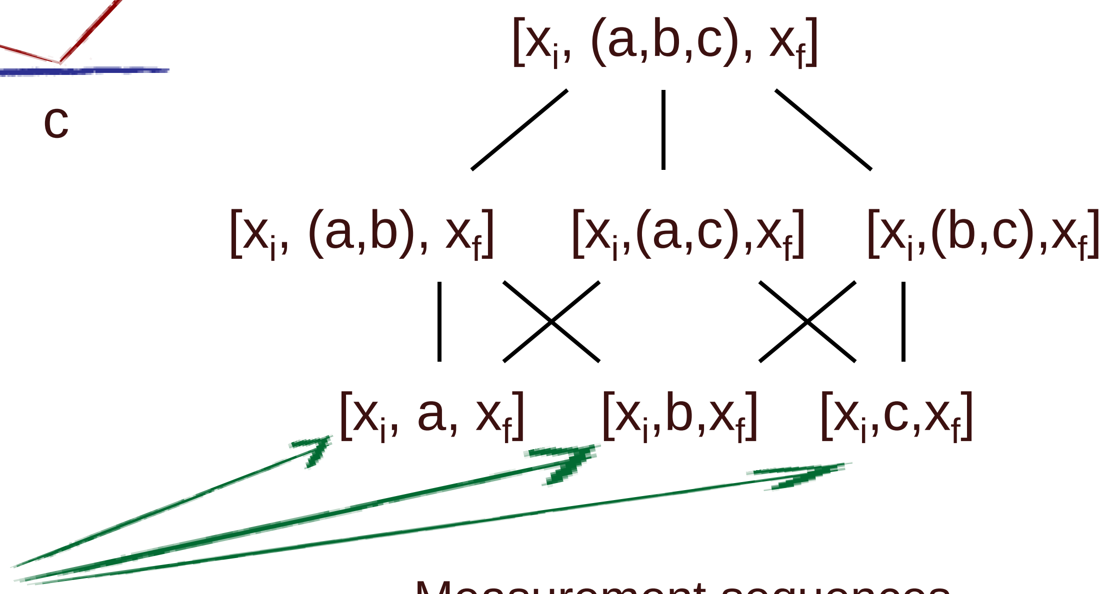


Feynman Mirror Example

Feynman Mirror



states quantified by a pair of real numbers: amplitudes

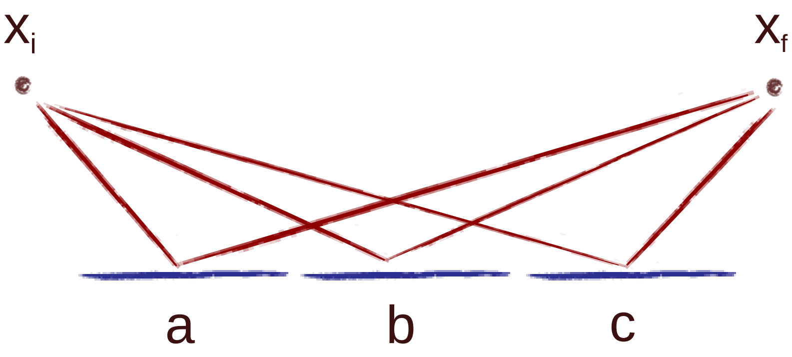


assign amplitudes to join-irreducible elements

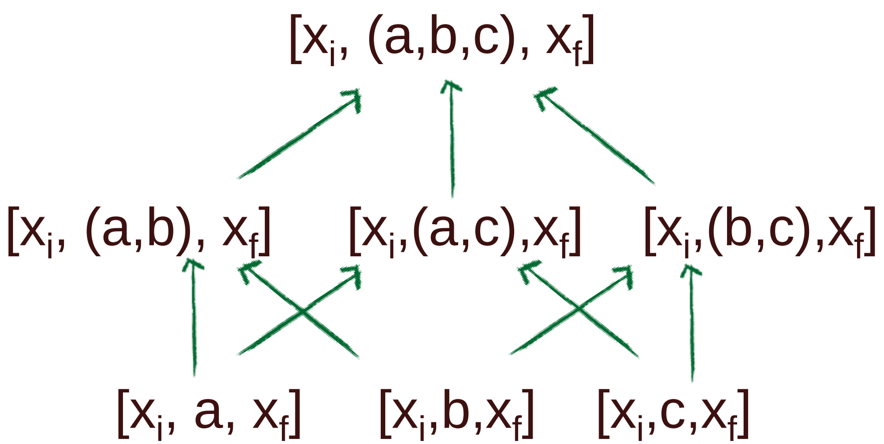
Measurement sequences

Feynman Mirror Example

Feynman Mirror



states quantified by a pair of real numbers: amplitudes



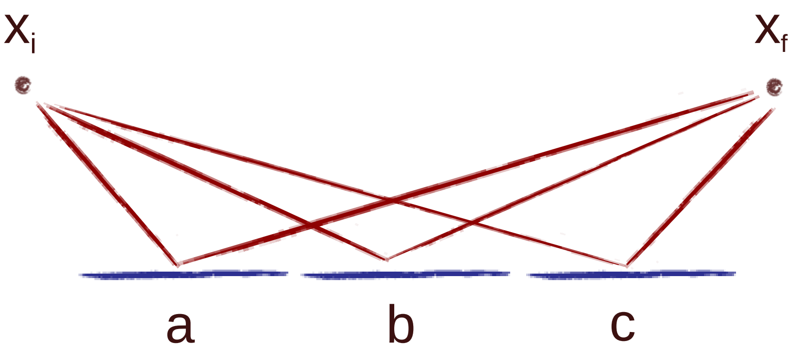
amplitudes combine by rules of complex arithmetic

sum and product rules used to compute amplitudes of interest

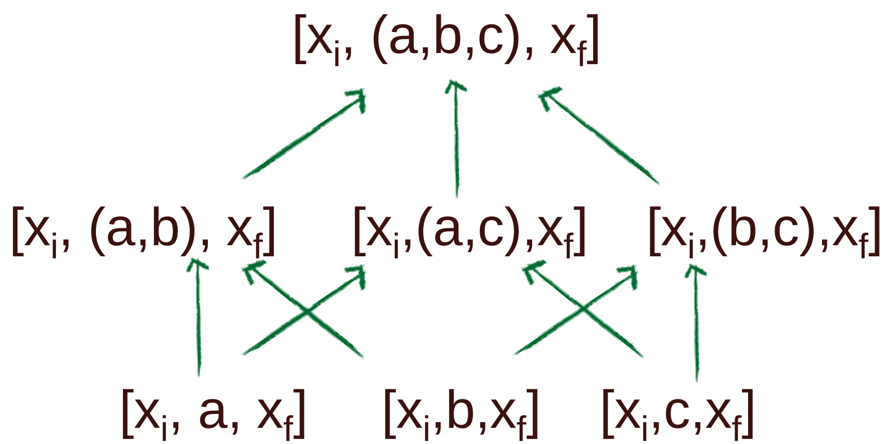
Measurement sequences

Feynman Mirror Example

Feynman Mirror



states quantified by a
pair of real numbers:
amplitudes

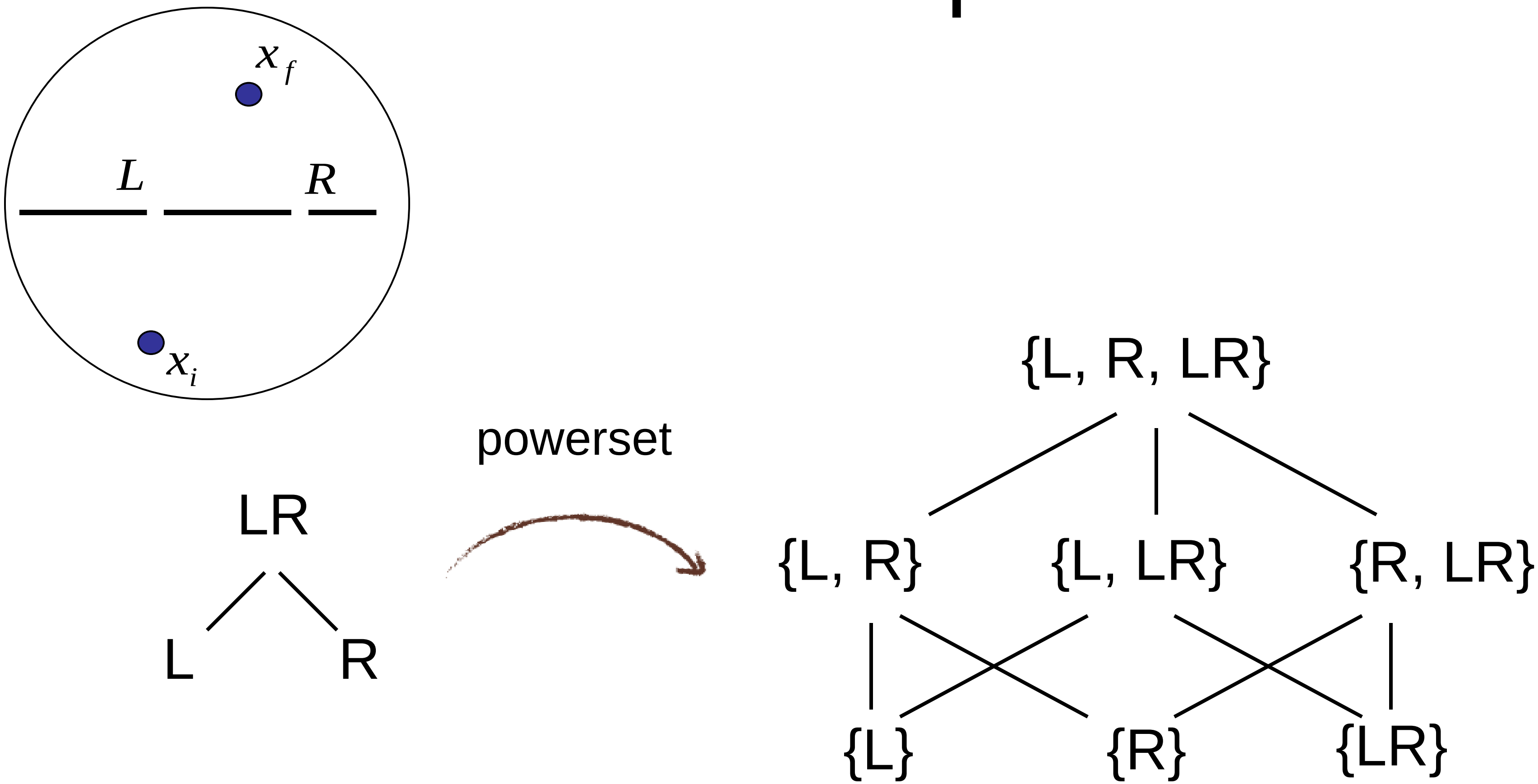


statement lattice
is the powerset of states

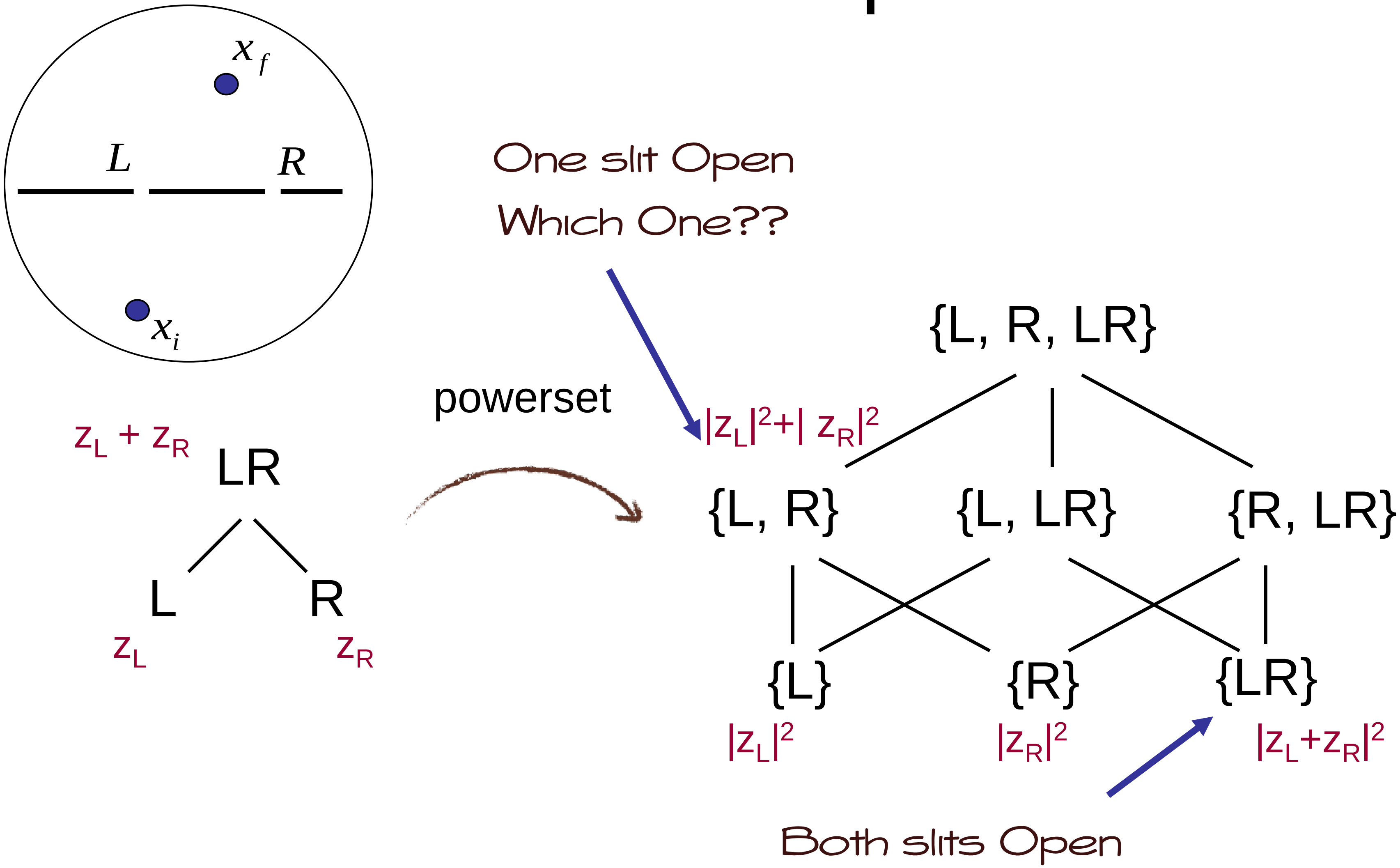
Born Rule gives probabilities
of atomic statements

Measurement sequences

Double-Slit Experiment



Double-Slit Experiment



Conclusions

The complex Feynman formalism is derived from the Pair Postulate, and basic symmetries of combining measurement sequences in series and parallel

Quantum Theory is not distinct from Probability Theory, rather, it relies on it!

Quantum Inference differs from Classical Inference in that the measurement sequences form posets, whereas classically we deal with mutually exclusive, exhaustive states that form an antichain

Quantum Theory and Probability Theory both arise from quantifying fundamental relationships represented by lattices and posets

More Information

kknuth@albany.edu
http://knuthlab.rit.albany.edu

Knuth K.H., Skilling J. 2012. Foundations of Inference. Axioms 1(1), 38-73;
http://www.mdpi.com/2075-1680/1/1/38

Goyal P., Knuth K.H., Skilling J. 2010. Why Quantum Theory is Complex,
Phys. Rev. A 81, 022109.
http://arxiv.org/abs/0907.0909

Goyal P., Knuth K.H. 2011. Quantum theory and probability theory: their relationship and origin in symmetry, Symmetry 3(2):171-206.
http://www.mdpi.com/2073-8994/3/2/171

Knuth K.H., Bahreyni N. 2012. The Physics of Events: A Potential Foundation for Emergent Space-Time.
http://arxiv.org/abs/1209.0881

Knuth K.H. 2012. Inferences about Interactions: Fermions and the Dirac Equation. MaxEnt 2012.
arXiv:1212.2332 [quant-ph]