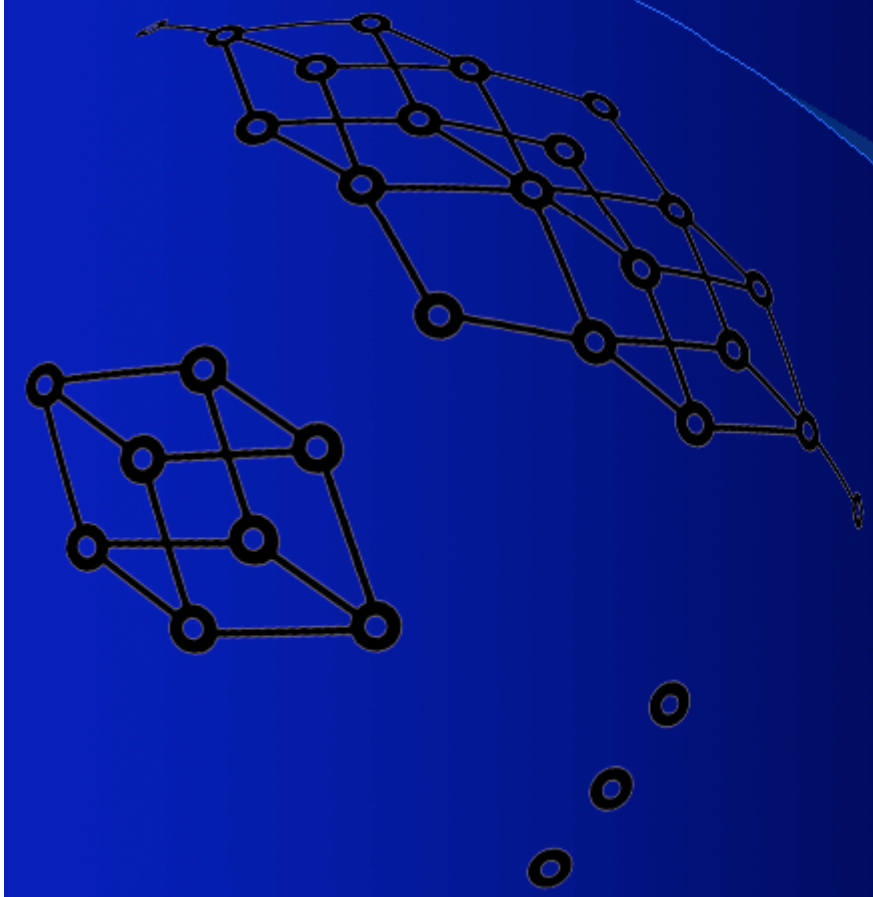
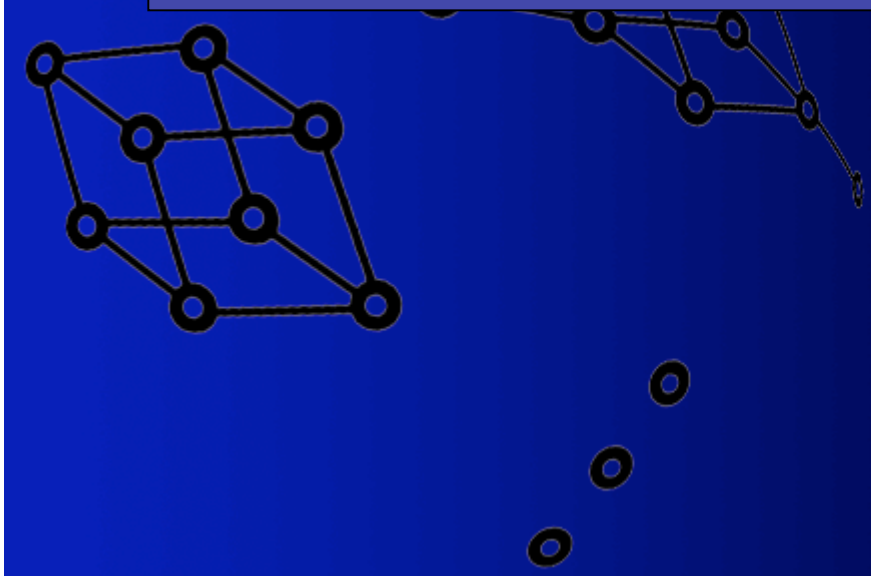
Deriving Laws From Ordering Relations



Kevin H. Knuth, Ph.D.
Computational Sciences Division
NASA Ames Research Center
Moffett Field CA 94035

"But Farmer Hoggett knew that little ideas
that tickled and nagged and refused to go away
should never be ignored
for within them lie the seeds of destiny."

-from the movie Babe based on the novel
of the same title written by Dick King-Smith



Outline

Cox and Probability

Review of Cox's Derivation

Order

The Importance of Order

Posets, Lattices and Algebras

Valuations on a Lattice

Origins

Geometry

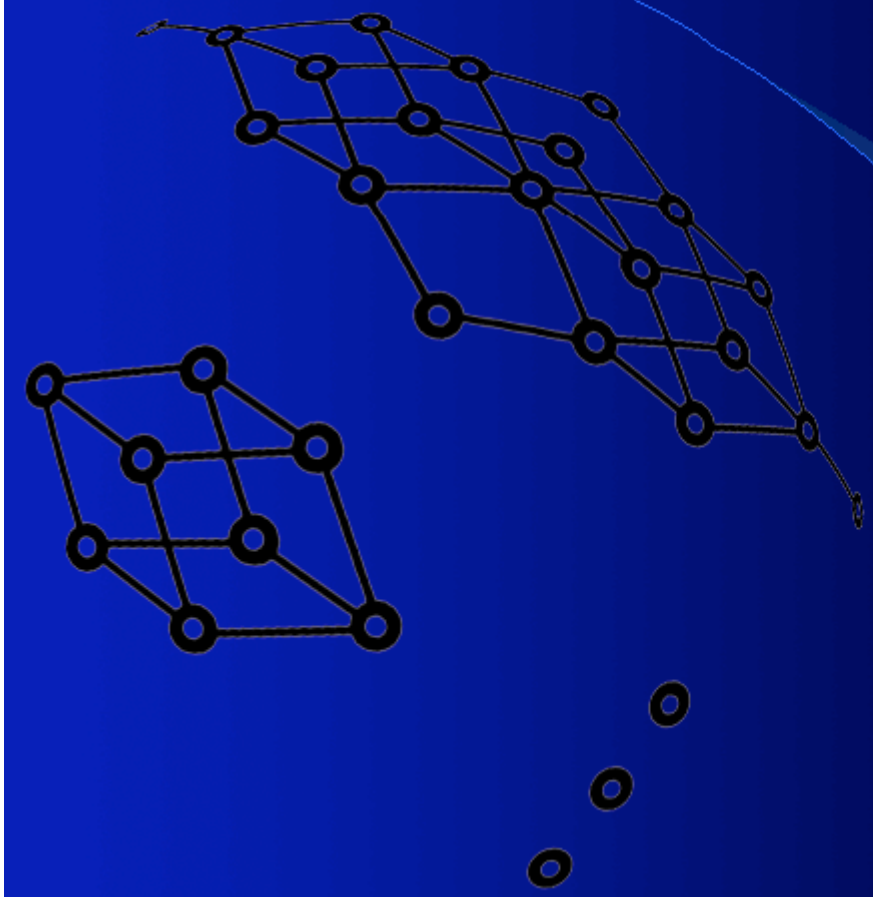
Probability Theory

Quantum Mechanics

A New Methodology

The Role of Order in Science

Cox and Probability



Cox's Contribution

Cox generalized implication among logical statements to degrees of implication represented by real numbers.

Associativity of the Conjunction

Consistency with associativity of the conjunction

$$
\begin{aligned}
p(b \wedge c \wedge d \mid a) &= p((b \wedge c) \wedge d \mid a) \\
&= p(b \wedge (c \wedge d) \mid a)
\end{aligned}
$$

Results in the Product Rule

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

Complementation

Consistency with complementatation

$$p(b \mid a) \quad = \quad p(\sim(\sim b) \mid a)$$

Results in the Sum Rule

$$p(b \mid a) + p(\sim b \mid a) = 1$$

Commutativy of the Conjunction

Consistency with the commutativity of the conjunction

$$p(a \wedge b \mid h) \equiv p(b \wedge a \mid h)$$

Results in Bayes Theorem

$$p(a \mid b \wedge h) = \frac{p(b \mid a \wedge h)\, p(a \mid h)}{p(b \mid h)}$$

Inferential Calculus

The inferential calculus (probability theory) derives directly from consistency with

Associativity

Commutativity

Complementation

These are BASIC mathematical ideas, not restricted to inference.

Other Trails Up the Mountain

Other Trails up the Mountain

Dr. Aczel's major contribution has been his thorough investigation of the functional equations central to this development.

Ray Smith and Gary Erickson investigated all possible forms of the associativity equations.

Anton Garrett derived the sum and product rules using consistency with the NAND operation.

Certainly I am missing other contributions…

A More General Derivation

Ariel Caticha derived the sum rule from consistency with associativity of the disjunction

$$p(a \quad b \mid h) \;=\; p(a \mid h) + p(b \mid h)$$

when a and b are logically independent.

The product rule can then be derived from consistency with distributivity.

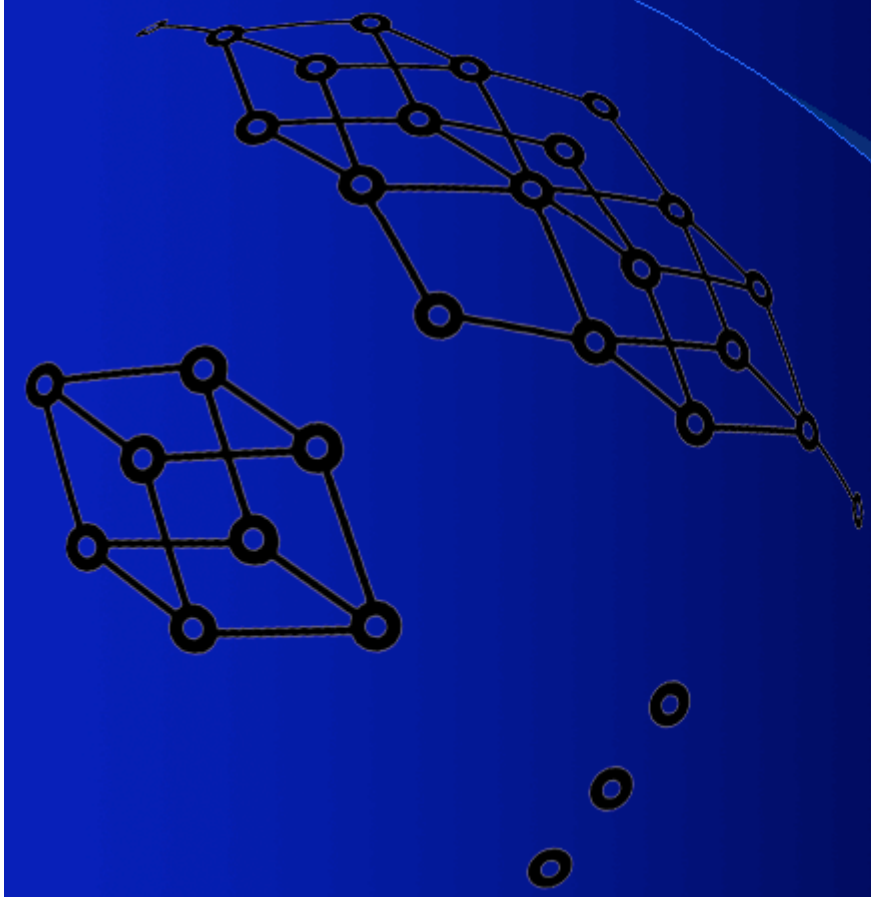
$$p(a \square b \mid h) \;=\; p(a \mid b \square h)\; p(b \mid h)$$

What is the Big Deal?

Because the Sum and Product Rules are not JUST associated with Boolean algebras.

They are associated with Distributive Algebras!

Order



Ordering Relations

To sets of objects, one can often impose additional structure, such as a binary **ordering relation** denoted by $a \sqsubseteq b$, which satisfies for all a, b, c (Birkhoff 1967):

P1.	For all a, $a \sqsubseteq a$.	(Reflexive)
P2.	If $a \sqsubseteq b$ and $b \sqsubseteq a$, then $a = b$	(Antisymmetry)
P3.	If $a \sqsubseteq b$ and $b \sqsubseteq c$, then $a \sqsubseteq c$	(Transitivity)

Now $a \sqsubseteq b$ is read "b contains a" or "b includes *a*"

If $a \sqsubseteq b$ and $a \neq b$ one can write $a < b$ and read "a is less than b" or "a is properly contained in b".

If $a < b$ but $a < x < b$ is not true for any x in the poset *P*, then we say that "b *covers* a", written $a \prec b$.

Posets

Together a set and an ordering relation are called a partially ordered set, or a poset.

The set of Natural numbers {1, 2, 3, 4, 5} with the binary ordering relation "is less than or equal to" $\sqsubseteq$ is a poset.

It is clear that:

$2 \sqsubseteq 2$ (Reflexive)

As $2 \sqsubseteq 2$ and $2 \sqsubseteq 2$, then $2 = 2$ (Antisymmetry)

As $2 \sqsubseteq 3$ and $3 \sqsubseteq 4$, then $2 \sqsubseteq 4$ (Transitivity)

Also $2 < 3$ as $2 \sqsubseteq 3$ but $2 \neq 3$

And $2 \prec 3$ as $2 < 3$ but there is no Natural number x such that $2 < x < 3$

Visualizing the Structure

The covering relation can be used to visualize the structure of a poset.

Whenever $a \lessdot b$ draw b above a:

5

4

3

2

1

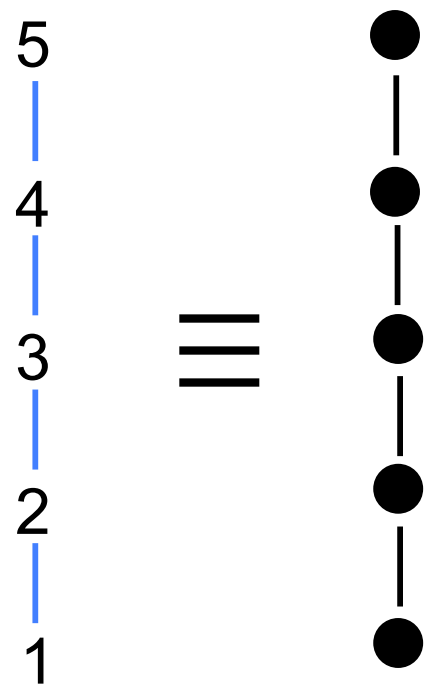
Haase Diagrams

The covering relation can be used to visualize the structure of a poset.

Whenever $a \sqsubset b$ draw b above a.

And whenever $a \prec b$ connect the elements with a line:

This poset forms a **chain**.

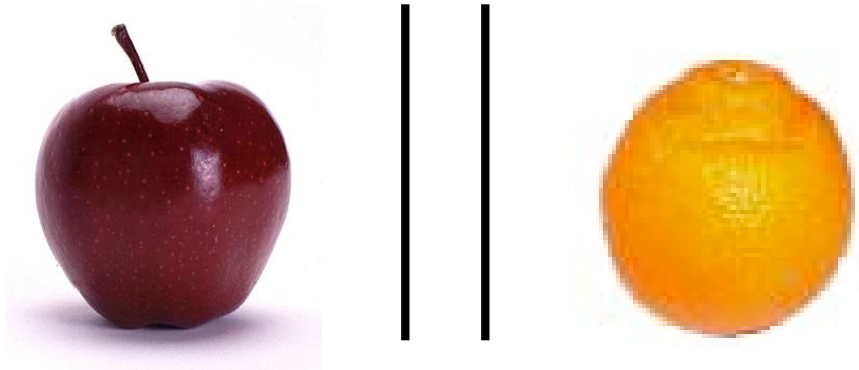


Incomparable Elements

There are times where for a given ordering relation, it is not true that $a \sqsubseteq b$ or $b \sqsubseteq a$.

We then write $a \,||\, b$ read "a is incomparable to b"

Perhaps for the ordering relation "*is healthier than*" we have



Antichains

The diagram corresponding to a poset of three incomparable elements is a picture with the elements placed side-by-side.

This is called an **antichain**.

● ● ●

A More Useful Illustration

Consider the powerset of the set $S = \{ a, b, c \}$

This is the set of all possible subsets of S:

$$P(S) = \left\{ \varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\} \right\}$$

A natural ordering is the relation "*is a subset of*", $\square$

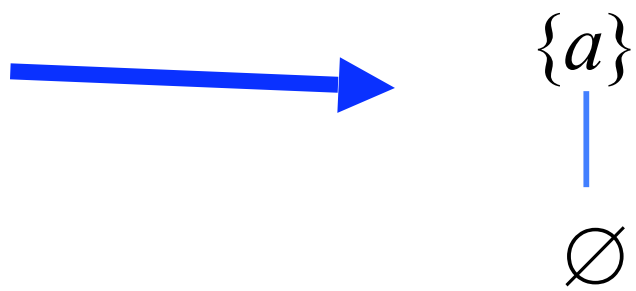
$$P = \left(\left\{ \varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\} \right\}, \square \right)$$

The First Level

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

First we note that $\varnothing \square \{a\}$,
from which we also see that: $\varnothing \prec \{a\}$

So we draw $\{a\}$ above $\varnothing$
and connect them with a line.



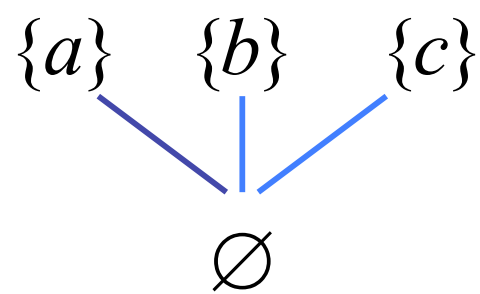
Completing the First Level

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

It is also true that $\varnothing \prec \{b\}$ and $\varnothing \prec \{c\}$ so we draw them above $\varnothing$ as well and connect them with lines.

However, $\{a\} \parallel \{b\}$ as neither one is the subset of the other.

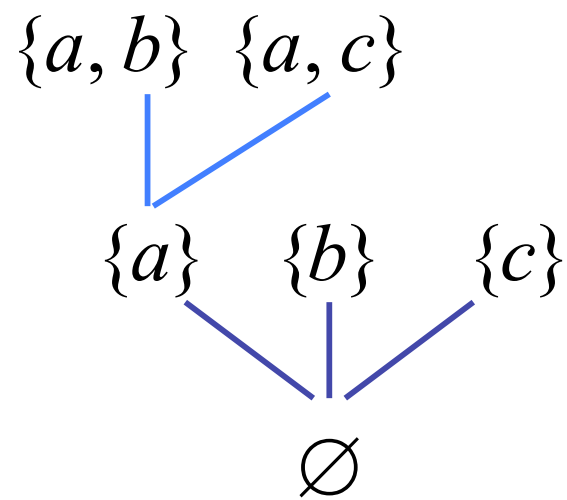
In addition, $\{a\} \parallel \{c\}$ and $\{b\} \parallel \{c\}$. So we draw them on the same level and do not connect them.



The Second Level

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

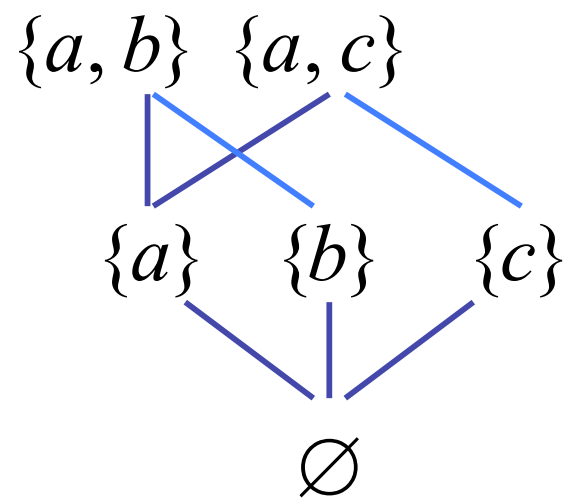
Now we note that $\{a\}$ is covered by two elements $\{a, b\}$ and $\{a, c\}$.



The Second Level

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

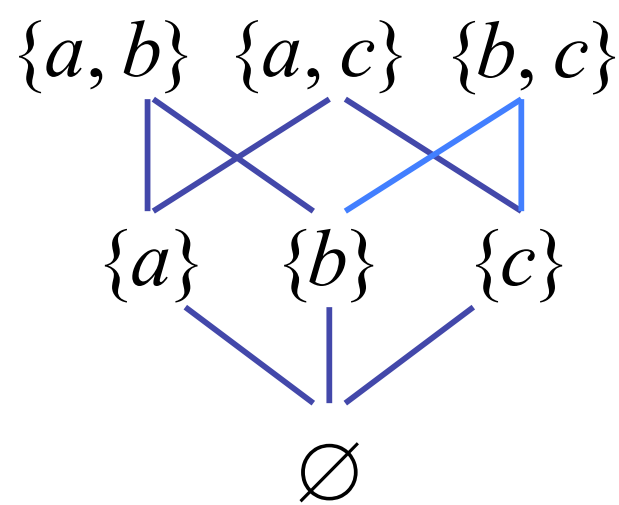
These elements also cover $\{b\}$ and $\{c\}$



Completing the Second Level

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

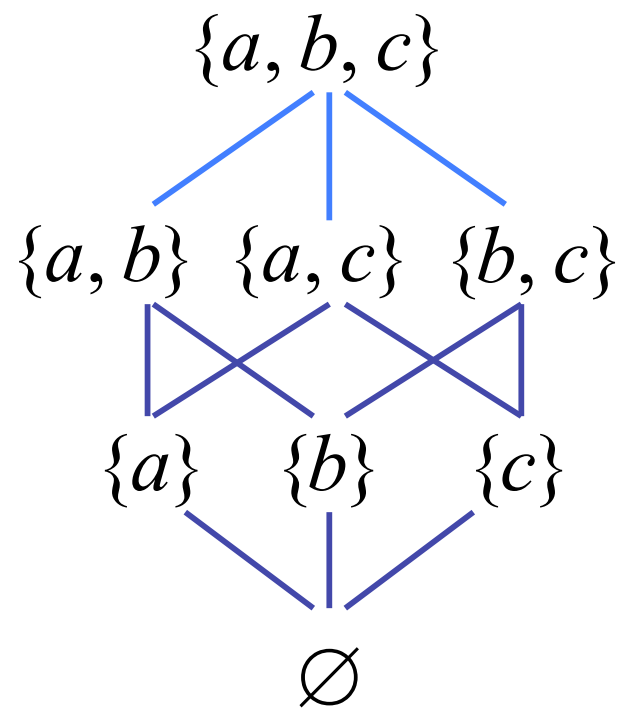
Now $\{b, c\}$ also covers $\{b\}$ and $\{c\}$, but these top elements are also incomparable.



The Third Level

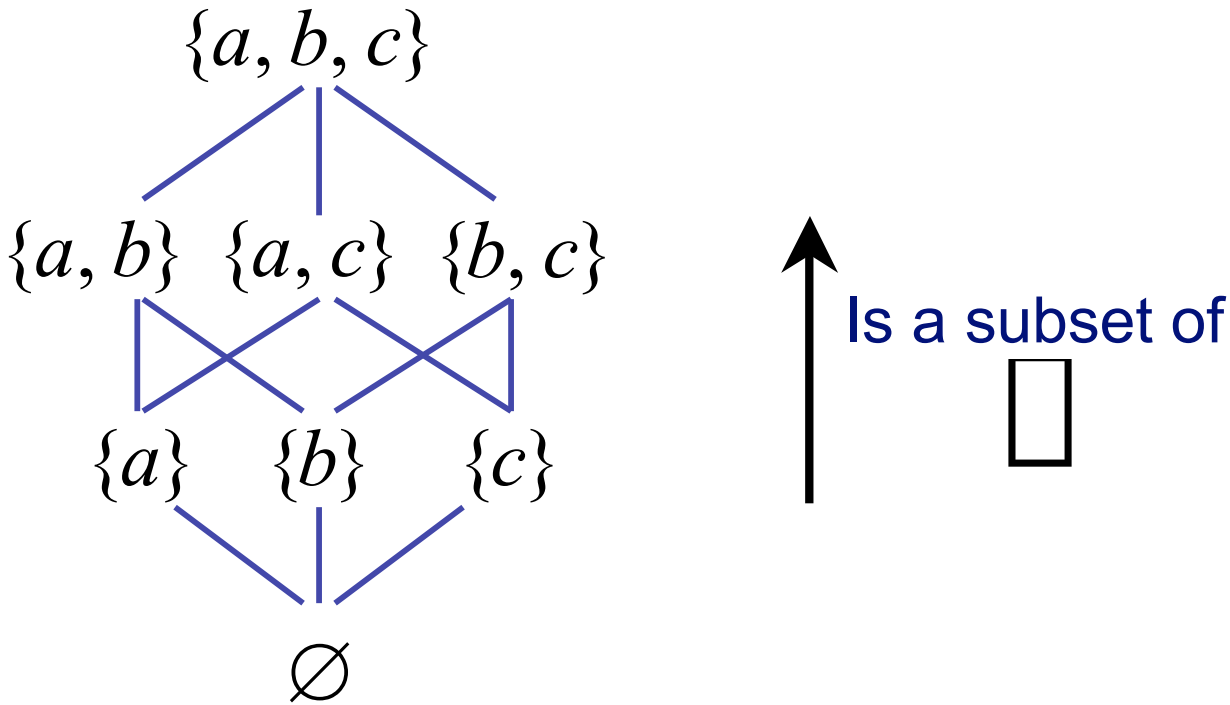
$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$

Finally $\{a, b, c\}$ covers all three two-element subsets.

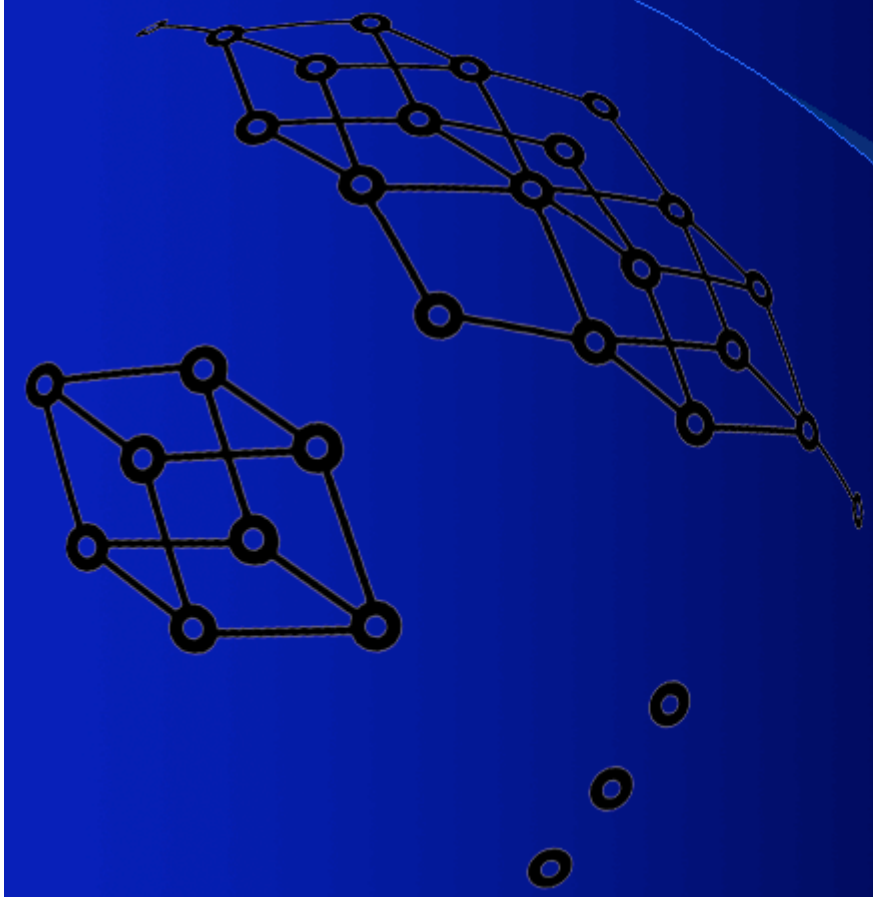


The Powerset of $\{a, b, c\}$

$$P = \left(\left\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\right\}, \square\right)$$



Lattices



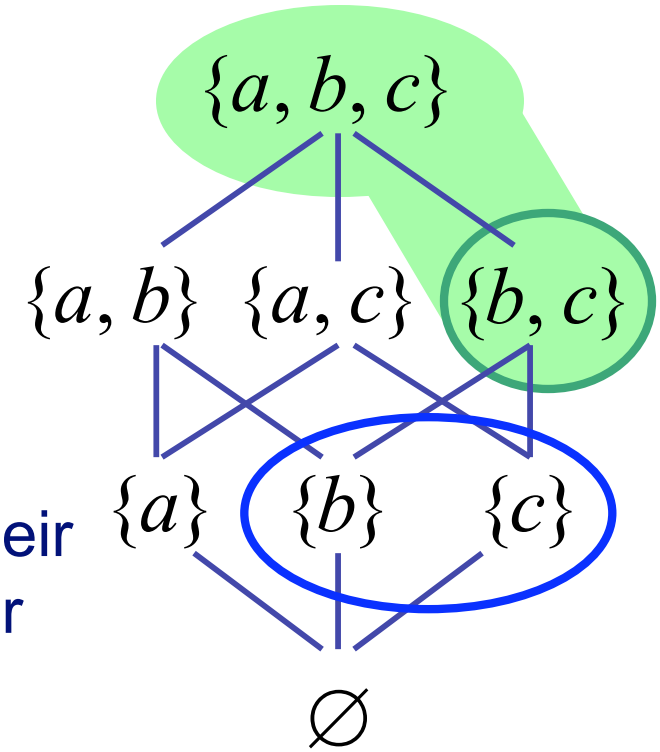
Lattices

A lattice is a poset P where every pair of elements x and y has

a least upper bound called the **join** $x \quad y$

a greatest lower bound called the **meet** $x \square y$

The green elements are upper bounds of the blue circled pair. The green circled element is their least upper bound or their join.



$\{b\} \quad \{c\} = \{b, c\}$

Similarly

$\{a, b\} \square \{b, c\} = \{b\}$

Lattice Identities

The Lattice Identities

L1. $x \sqcap x = x, \quad x \quad x = x$ Idempotent

L2. $x \sqcap y = y \sqcap x, \quad x \quad y = y \quad x$ Commutative

L3. $x \sqcap (y \sqcap z) = (x \sqcap y) \sqcap z, \quad x \quad (y \quad z) = (x \quad y) \quad z$ Associative

L4. $x \sqcap (x \quad y) = x \quad (x \sqcap y) = x$ Absorption

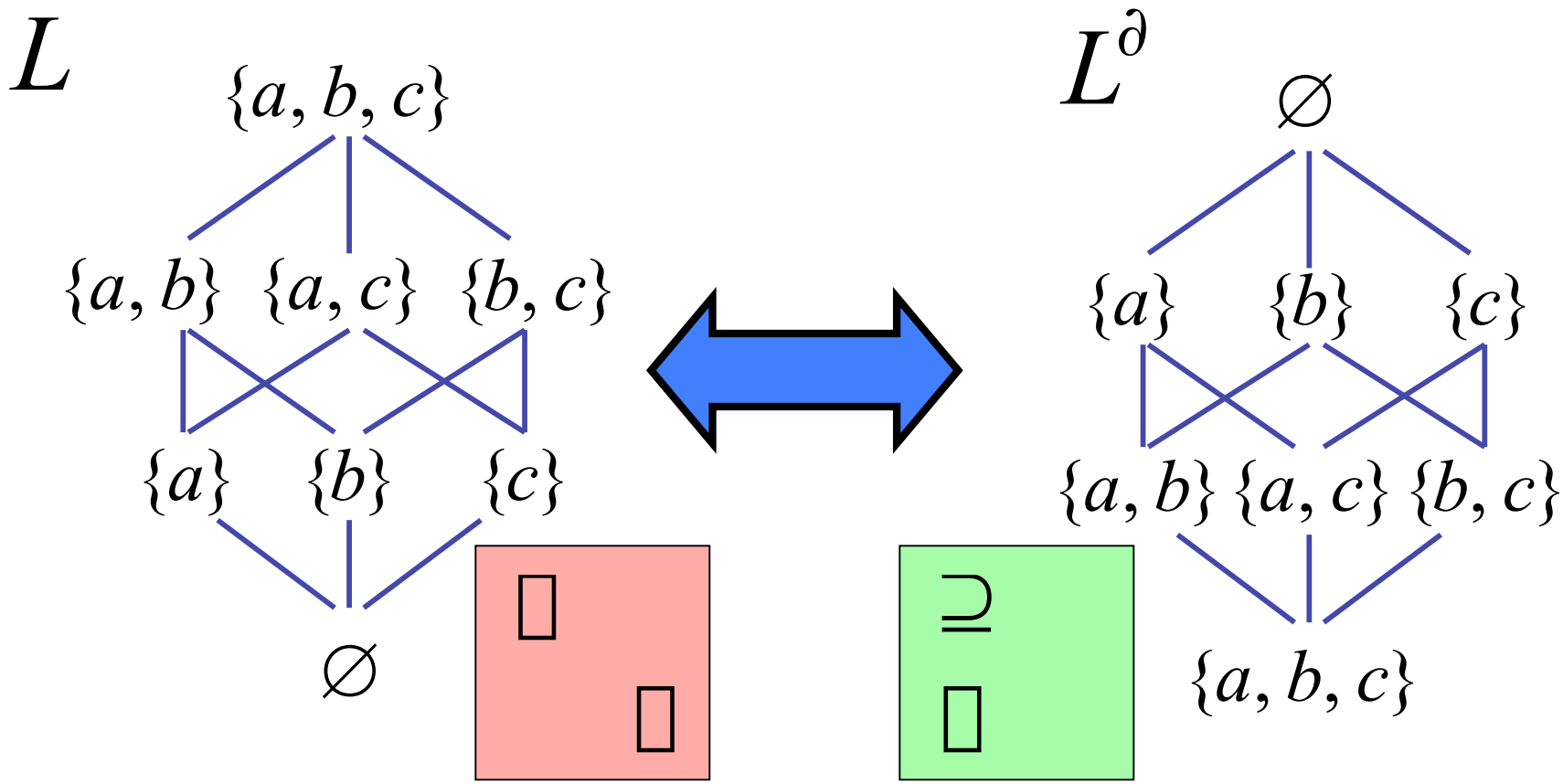
If $x \sqcap y$ the meet and join follow the **Consistency Relations**

C1. $x \sqcap y = x$ (x is the greatest lower bound of x and y)

C2 . $x \quad y = y$ (y is the least upper bound of x and y)

The Dual Lattice

The **dual lattice** can be obtained by reversing the ordering relation

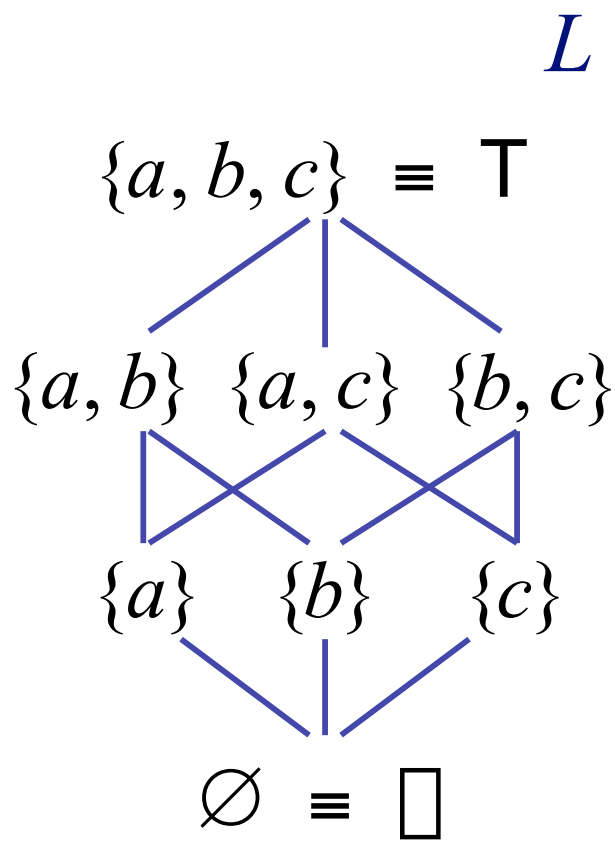


This flips the lattice upside-down and exchanges meets and joins.

Top and Bottom Elements

The greatest element is called the **top** and is symbolized by 1, I, or T.

So that $\mathsf{T} \geq x$ for all x in L.



The least element is called the **bottom** and is symbolized by $\varnothing$ or $\square$.

So that $x \geq \varnothing$ for all x in L.

Distributive Lattices

A **Distributive Lattice** possesses structure additional to L1-4.

It also satisfies the following identity for all elements x, y, z.

D1. $$x \sqcap (y \sqcup z) = (x \sqcap y) \sqcup (x \sqcap z)$$
$$x \sqcup (y \sqcap z) = (x \sqcup y) \sqcap (x \sqcup z)$$ Distributive

Note that these two equations are related by duality as the dual of a distributive lattice is a distributive lattice.

Complemented Distributive Lattices

There is a special case of a distributive lattice that possesses an interesting property where each element is associated with one other element called its **complement**.

The complement has these properties

B1. $x \sqcap \sim x = \sqcap \qquad x \quad \sim x = \mathrm{T}$

B2. $\sim(\sim x) = x$

B3. $\sim(x \sqcap y) = \sim x \quad \sim y \qquad \sim(x \quad y) = \sim x \sqcap \sim y$

Lattices and Algebras

Associated with every lattice is an algebra.

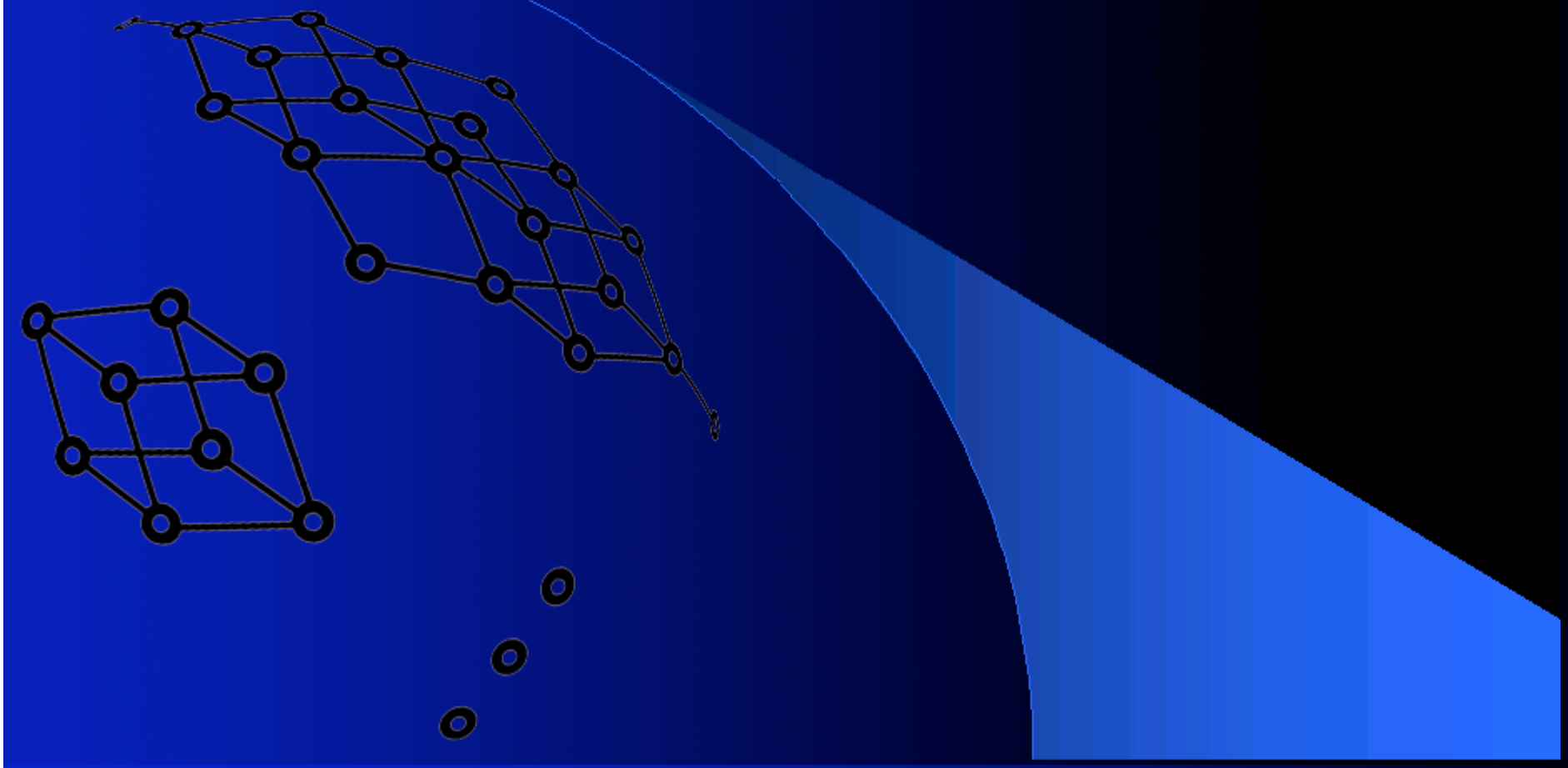
Thus a lattice can be expressed either in terms of its elements and its ordering relation

$$\langle L \,;\, \square \rangle$$

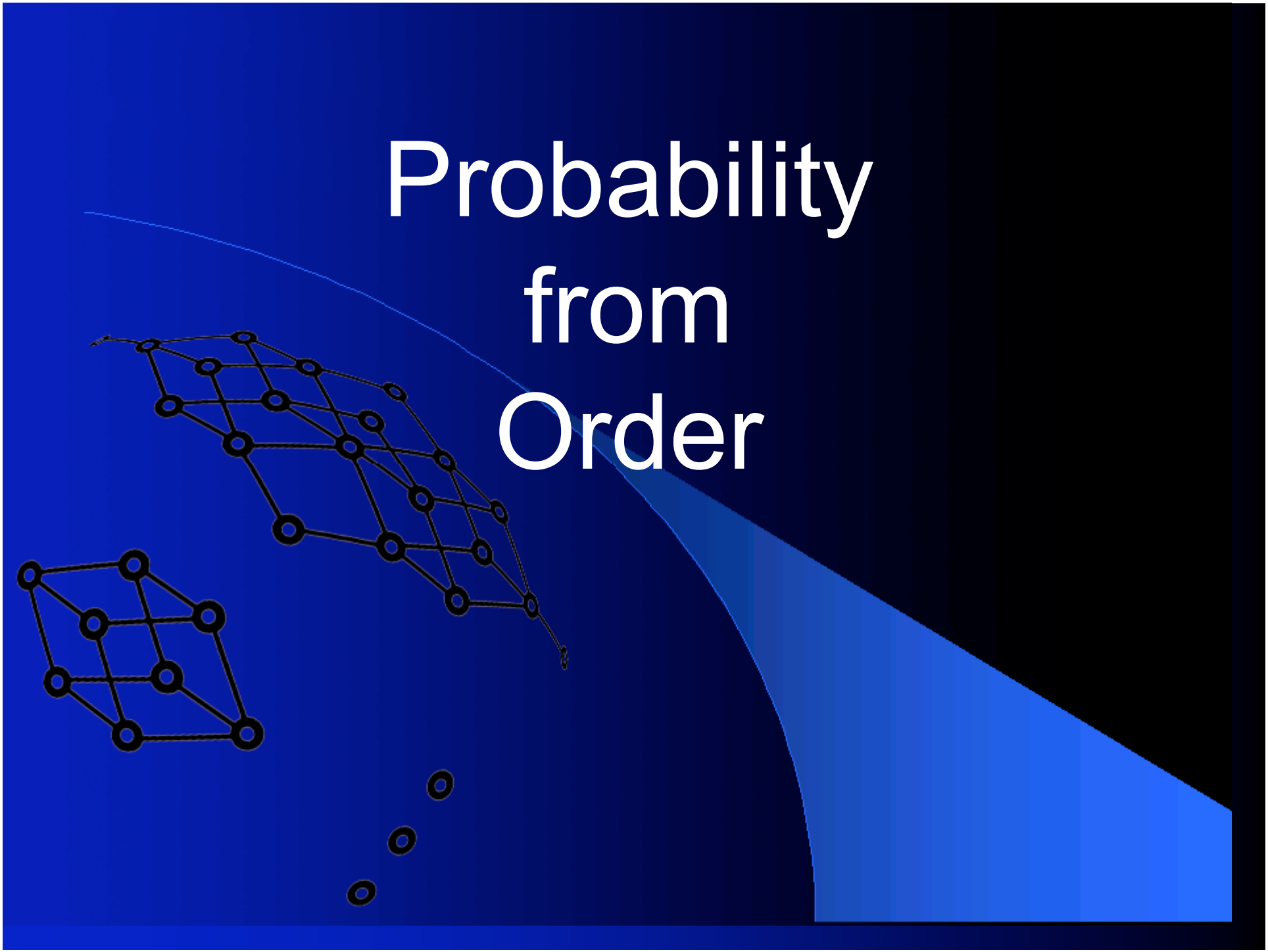
or in terms of its algebra

$$\langle L \,;\, \square, \quad \rangle \quad \text{or perhaps} \quad \langle L \,;\, \square, \quad , \sim \rangle$$

Origins



Probability from Order



Implication as an Ordering Relation

At this point the algebra associated with the complemented distributive lattice should look familiar - **Boolean algebra**.

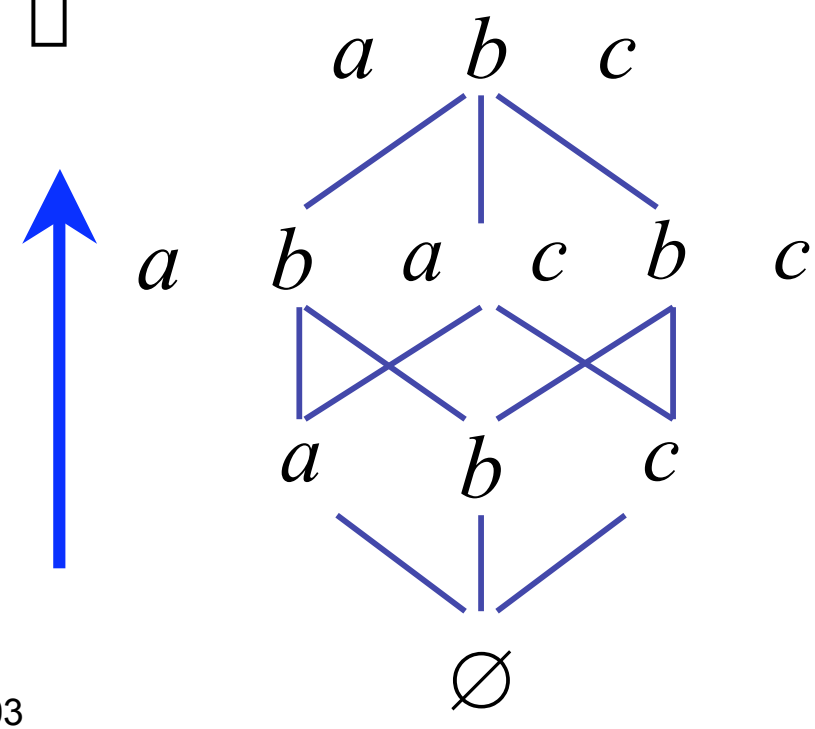
More commonly, the poset is a set of assertions and the ordering relation is "*implies*" □ ≡ □

T is the **Truism**

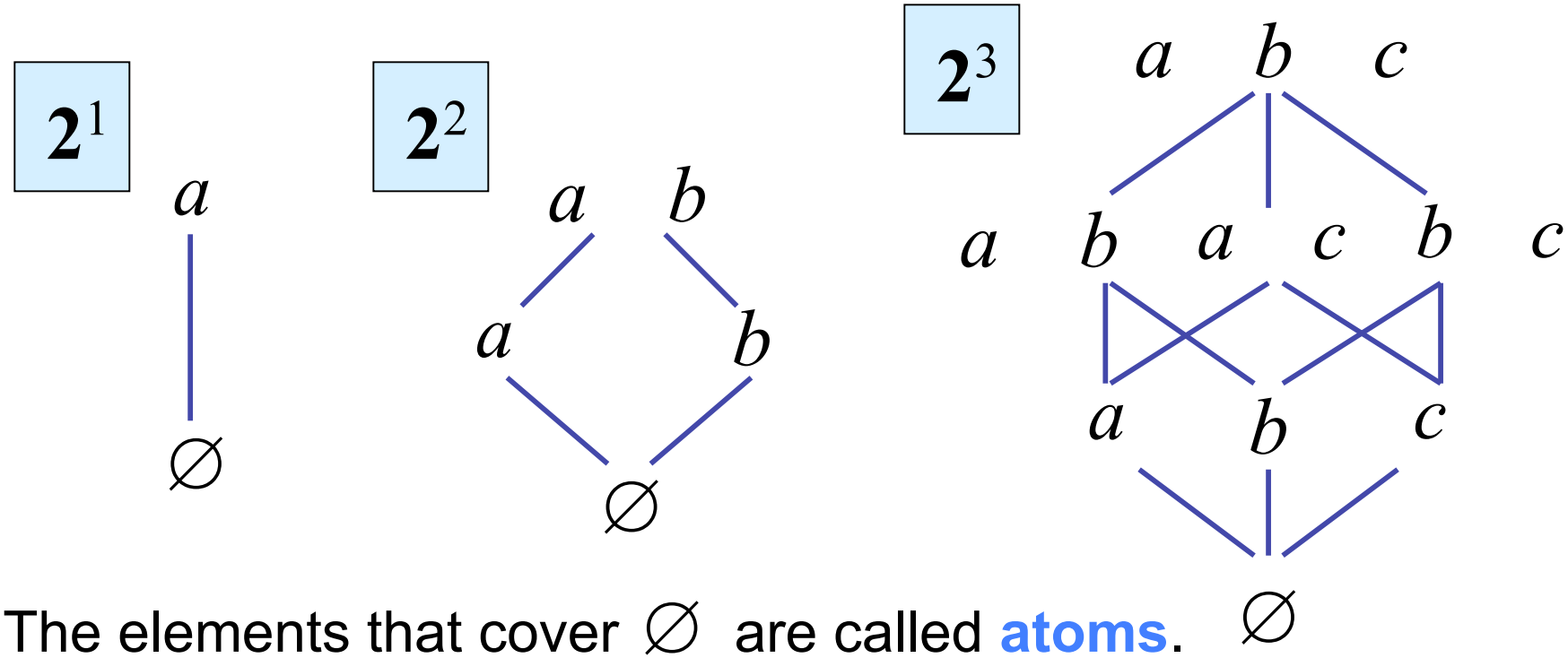
□ is the **Absurdity**

≡ **Logical Disjunction**

□ ≡ **Logical Conjunction**



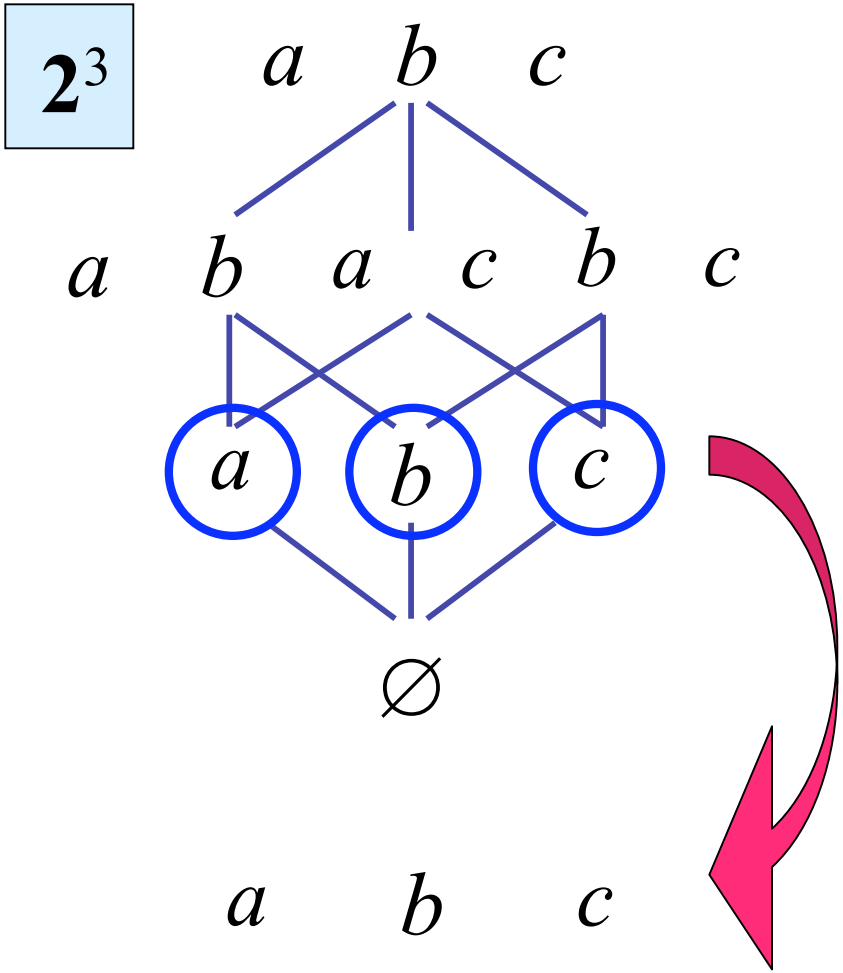
Boolean Lattices



The elements that cover $\varnothing$ are called **atoms**.

In a Boolean lattice the atoms are the **mutually exclusive assertions**. All other elements are joins of the atoms.

Join-Irreducible Elements

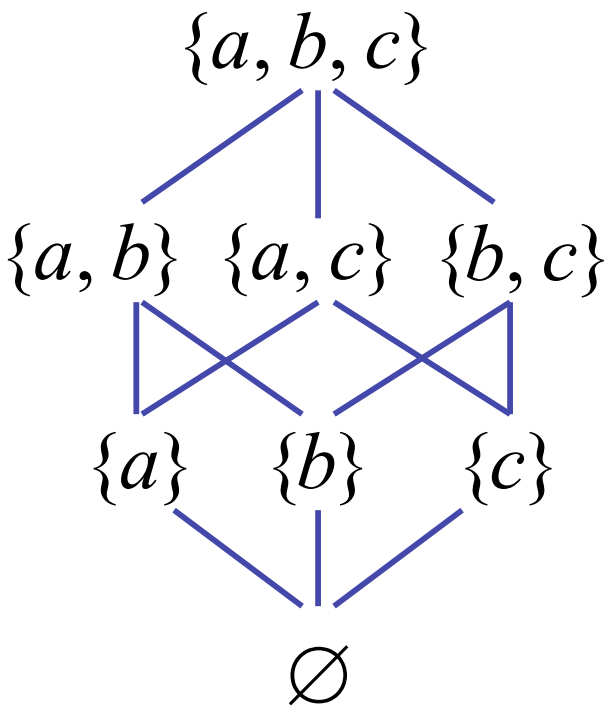


An important subset of a lattice is the set of elements that cannot be written as a join of elements. They are the **join-irreducible elements**.

In a Boolean lattice, these are the atoms, which by themselves form an **antichain**.

One can use this property to identify a Boolean lattice.

Powersets Revisited



The powerset (the set of all subsets) of a set forms a Boolean lattice under the ordering relation $\subseteq$.

Note that the atoms form an antichain.

The complement of a set S is the set T \ S.

Deductive Inference

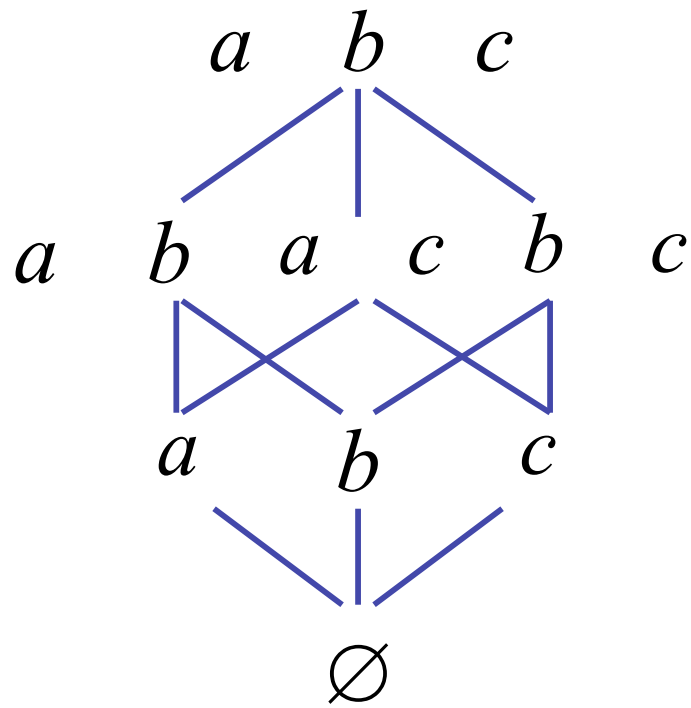
Deductive inference is easy.

For all $x \neq y$

If $x \sqsubseteq y$ then $x \sqsubseteq y$.

If $x \sqsubseteq y$ then $y \not\sqsubseteq x$.

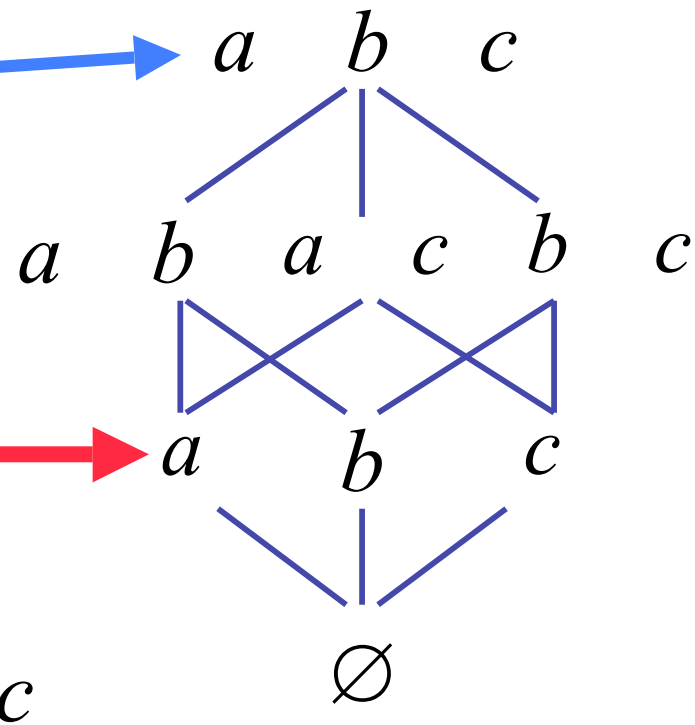
If $x \parallel y$ then $x \not\sqsubseteq y$ and $y \not\sqsubseteq x$.



Note that:
The absurdity $\sqsubseteq$ ($\varnothing$) implies everything.
The truism T is implied by everything.

Inductive Inference

Deduction is nice and all, but sometimes I know that one of a set of possibilities is true



And I want to know to what degree that knowledge implies a more simple hypothesis.

But since $a \quad b \quad c \geq a$

it is only true that $a \quad \square \quad a \quad b \quad c$

Not vice versa!

Generalizing Implication

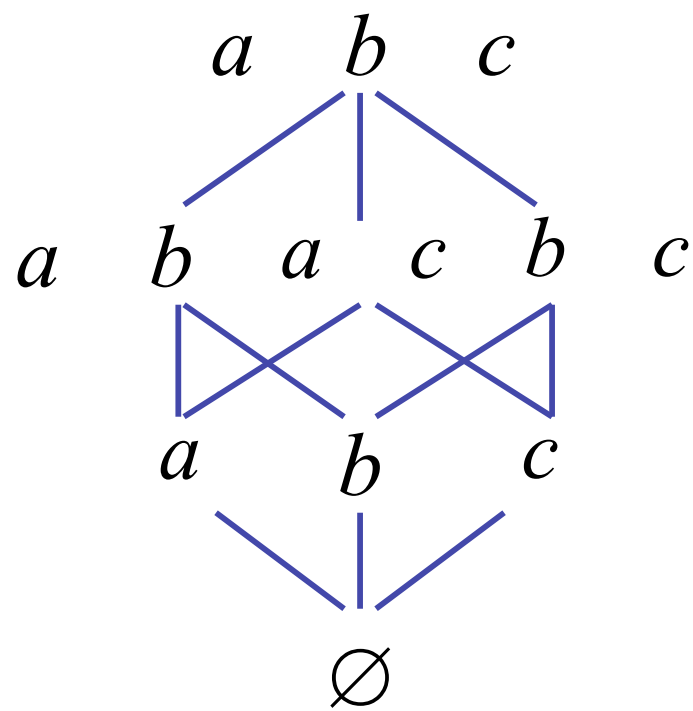
To generalize to degrees of implication, we introduce a real-valued function that takes two lattice elements to a real number:

$$p : L \times L \to \mathbb{R}$$

In particular we write this as:

$$p(a \mid \mathsf{T})$$

Since one of a, b, or c is true, the truism can be considered to be our prior information, in part.



Other Views

This function can be written in two different ways.

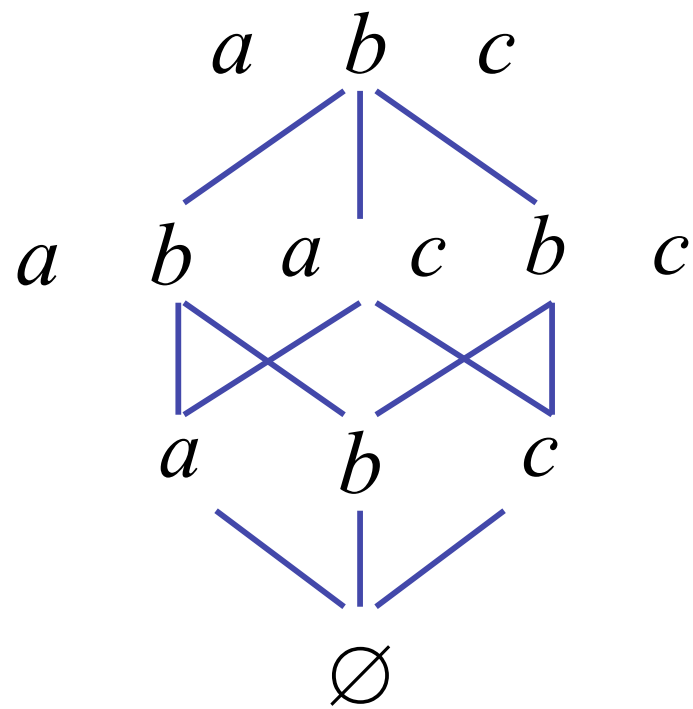
Typically when the premise is the truism, we can write this function as a function that takes a single lattice element to a real number:

$$p(a)$$

And write the function as

$$p(a \mid h)$$

in other cases.



Following the Rules

To be consistent with the Boolean lattice structure, this new measure must follow the rules:

L1-4, D1, and B1-3.

The key ones we use are

L3	Associativity of $\square$	$x \square (y \square z) = (x \square y) \square z$
D1	Distributivity	$x \quad (y \square z) = (x \quad y) \square (y \quad z)$
L2	Commutativity of $\square$	$x \square y = y \square x$

Valuations

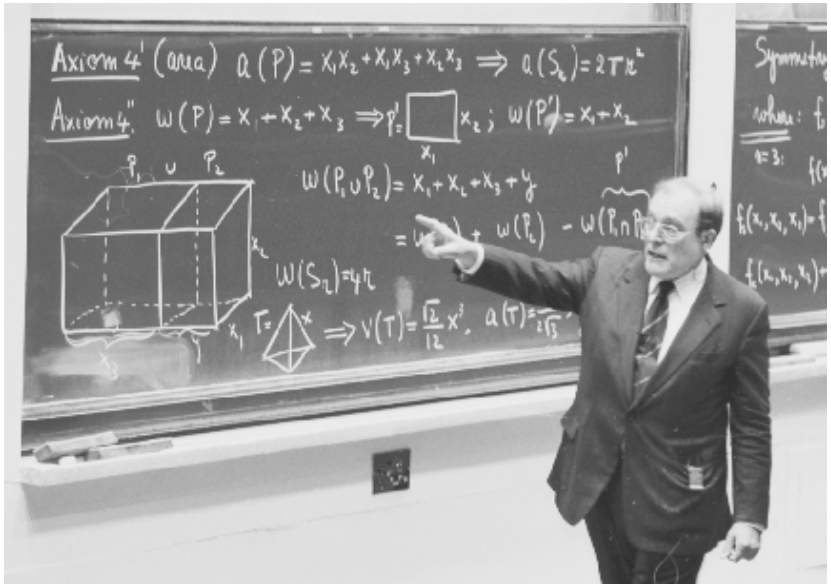
Valuations on Lattices

A valuation on a lattice is defined as a function that takes a lattice element to a commutative ring.

$$v : L \square A$$

where A is an element of a commutative ring.

This has been investigated in great detail by a small group of mathematicians led by Gian-Carlo Rota.



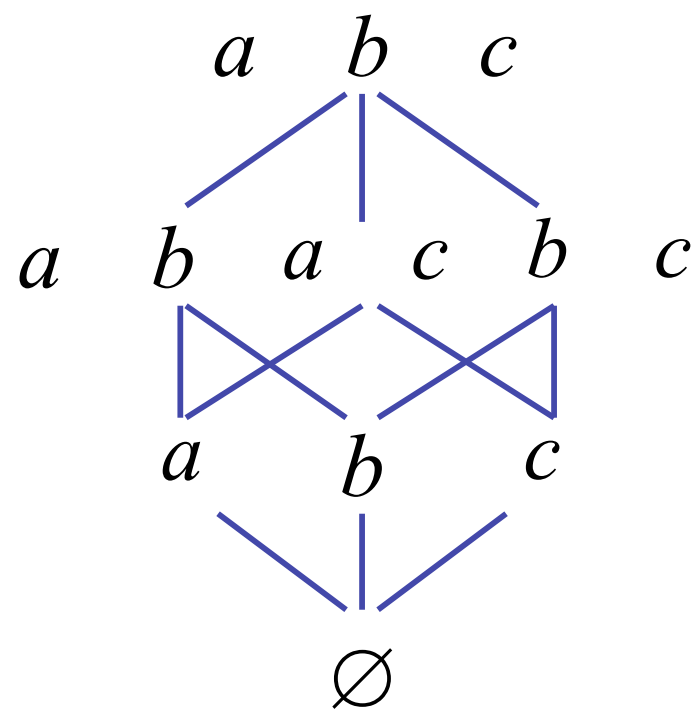
Probability as a Valuation

The function we defined for probability is a valuation.

$$v : L \to A$$

$$p : L \to L$$

$$p(a \mid \top) \equiv p(a)$$



Why is this Useful?

There is a theorem that all valuations can be uniquely determined from the valuations on the join-irreducible elements of the lattice

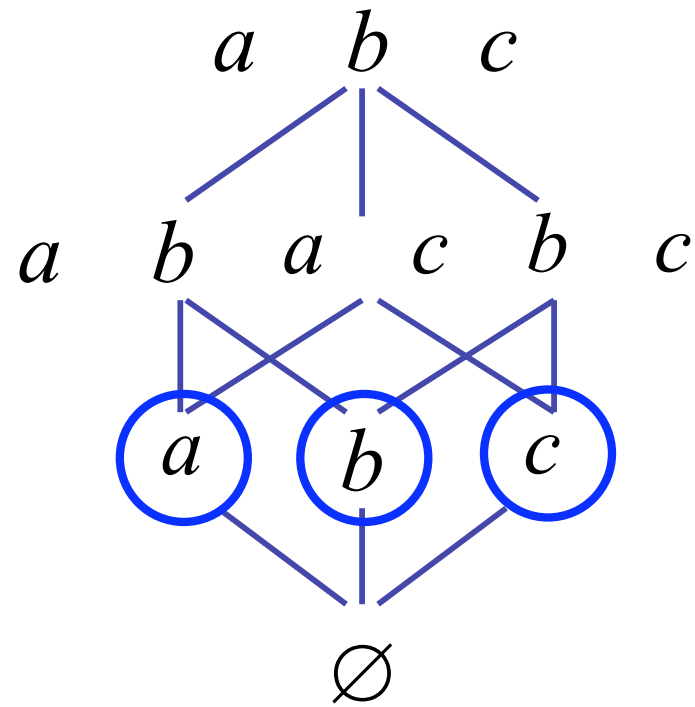
AND

their assignments are arbitrary!

Thus, by assigning the prior probabilities

$$p(a) \quad p(b) \quad p(c)$$

the probabilities of any other pair of elements in the lattice is determined.

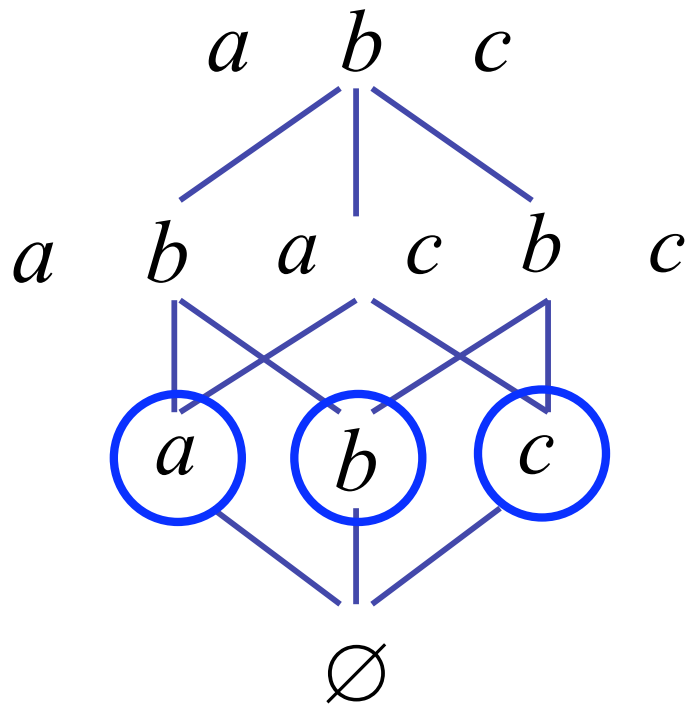


Assigning Priors is Hard

One can now see why assigning priors is difficult.

There is NO structure in the Boolean algebra of assertions that can guide us in these assignments.

We must employ other principles to assign them.



Prior Probabilities

Symmetry, constraints and consistency with other aspects of the problem can be used to assign prior probabilities.

Order-theoretic principles dictate the remaining probabilities.

Probability Theory from Order

Probability Theory and Physics

Thanks to the efforts of Ed Jaynes, Myron Tribus and others, I am able to wave my hands and say that I can derive much of physics from order-theoretic principles.

Understanding Maximum Entropy remains a challenge.

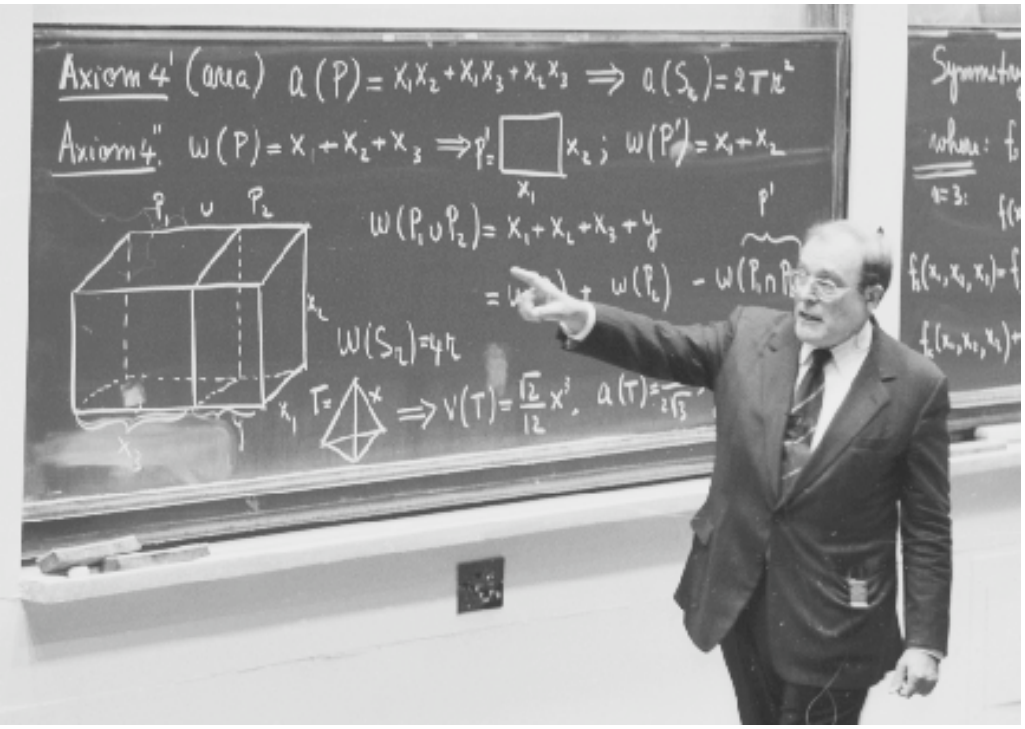
Geometry from Order

Geometric Probability

Many geometric laws can be derived from order-theoretic considerations.

Geometric objects can be ordered, conjoined and disjoined often resulting in a distributive lattice structure.

Valuations are assigned, which are invariant with respect to Euclidean translations and rotations.



These Valuations have a Basis!

For three-dimensional Euclidean geometry all invariant valuations can be written as a linear combination of 4 basis valuations.

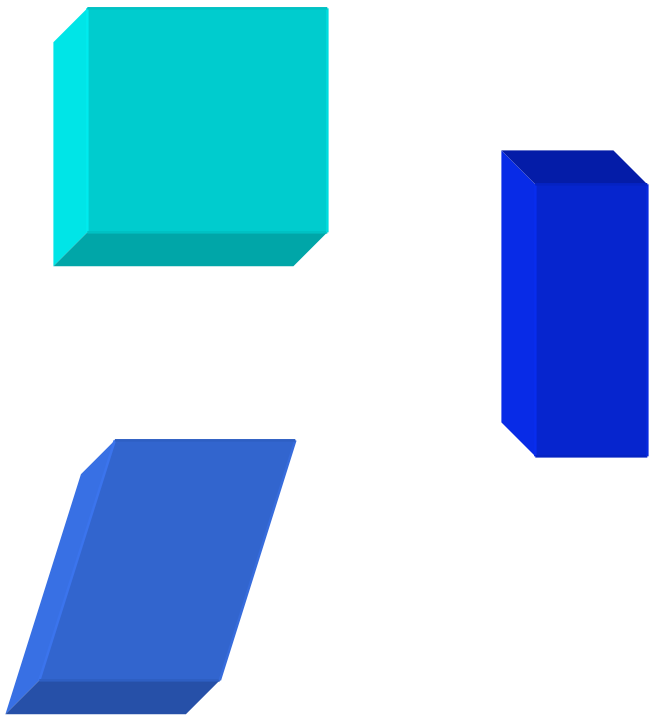
V = volume

A = surface area

W = mean width

χ = Euler characteristic

$\Phi = aV + bA + cW + d\chi$



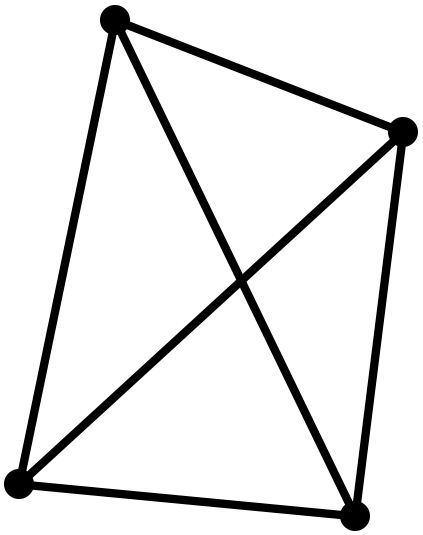
Euler Characteristic

The Euler characteristic is a valuation.

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

For a 3D tetrahedron it is found by

$$\chi(tetra) = 4 - 6 + 4 = 2$$



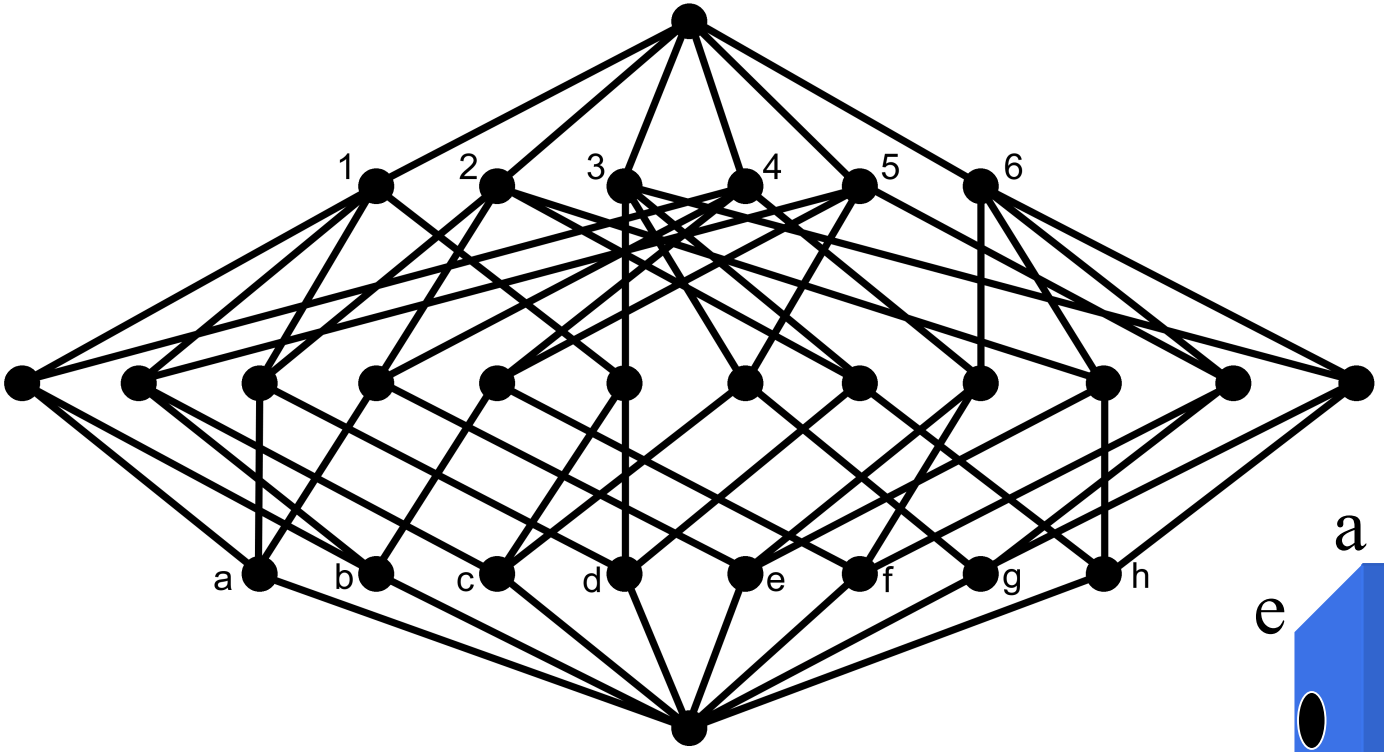
Euler Characteristic

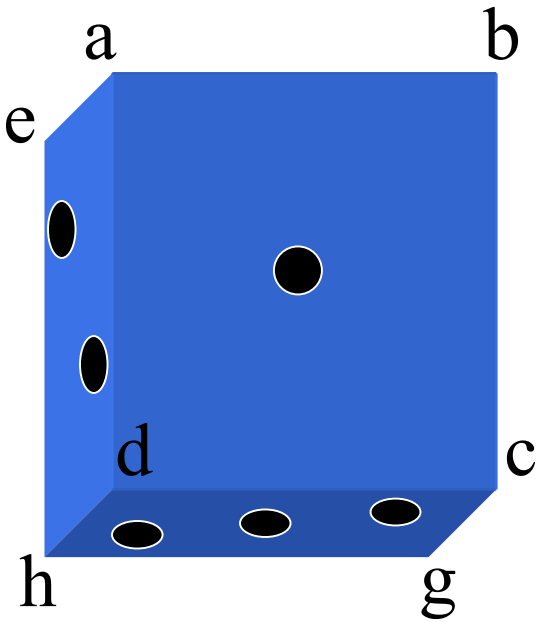
For a cube, it is

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

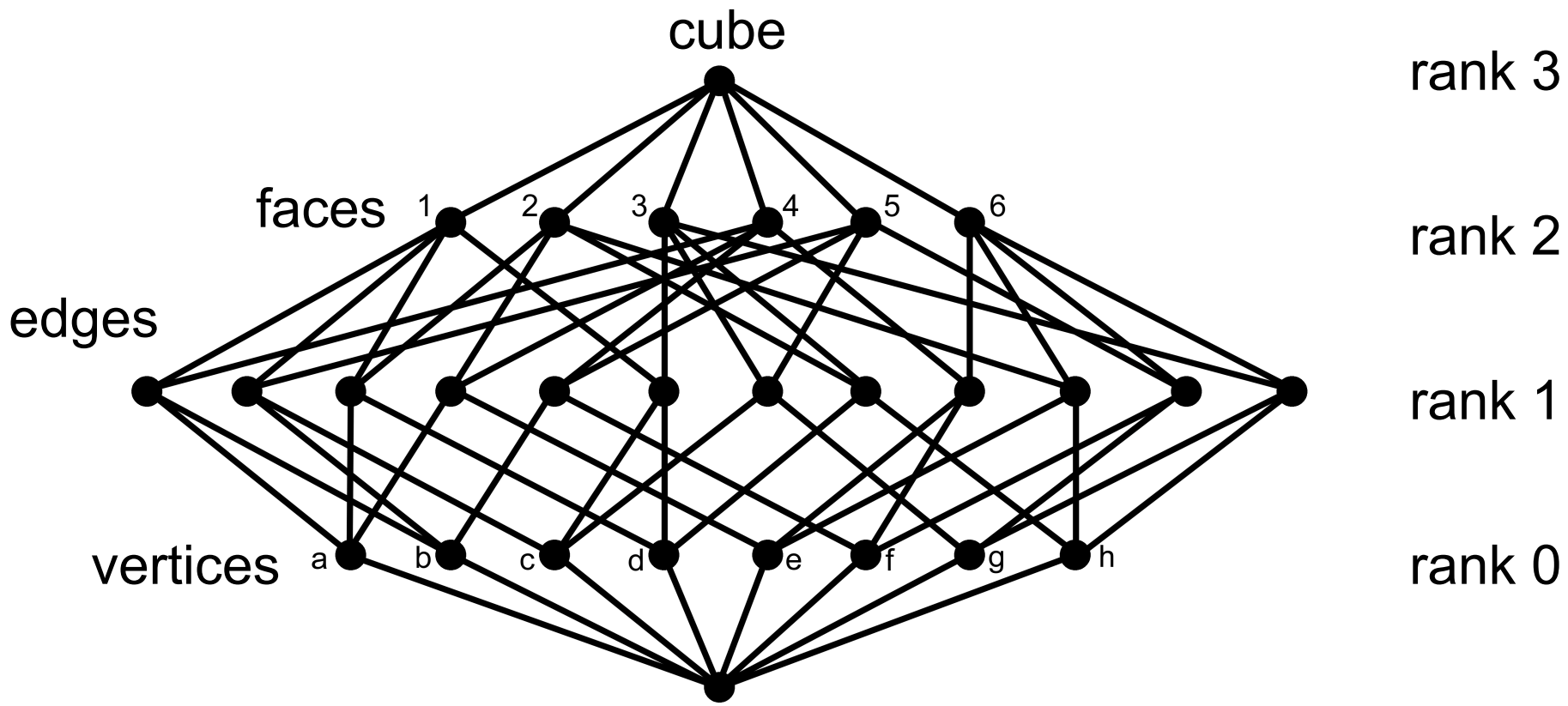
$$\chi(cube) = 6 - 12 + 8 = 2$$

Corresponding Lattice





Lattice Structure of a Cube



Mobius Functions

A Mobius function for a partially ordered set P is a function that satisfies:

$$\mu(x,x) = 1, \qquad x \in P$$

$$\sum_{x \le z \le y} \mu(x,z) = 0 \qquad x < y$$

$$\mu(x,y) = 0 \qquad y > x$$

These functions are important for inverting other functions on the lattice, as we will see later.

This is also related to the Euler characteristic in distributive lattices.

Inclusion-Exclusion Principle

Mobius functions allow us to compute valuations on elements higher in the lattice based on linear combinations of valuations of elements lower in the lattice.

For many familiar structures this leads to
Rota's **inclusion-exclusion principle** where when summing we add at one level and subtract at the next and so on.

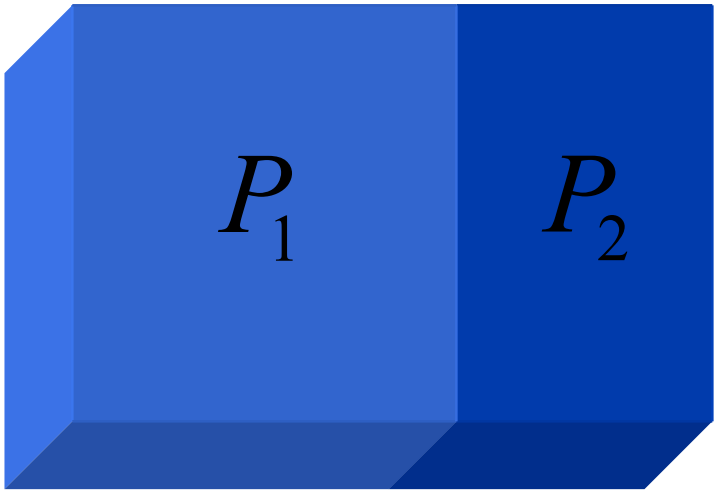
We saw this with the Euler characteristic

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

And we'll see it also on the next slide…

Joining Parallelotopes

$$P_1 \quad P_2$$



$$v(P_1 \quad P_2) = v(P_1) + v(P_2) \square v(P_1 \square P_2)$$

Quantum Mechanics from Order

Ariel Caticha

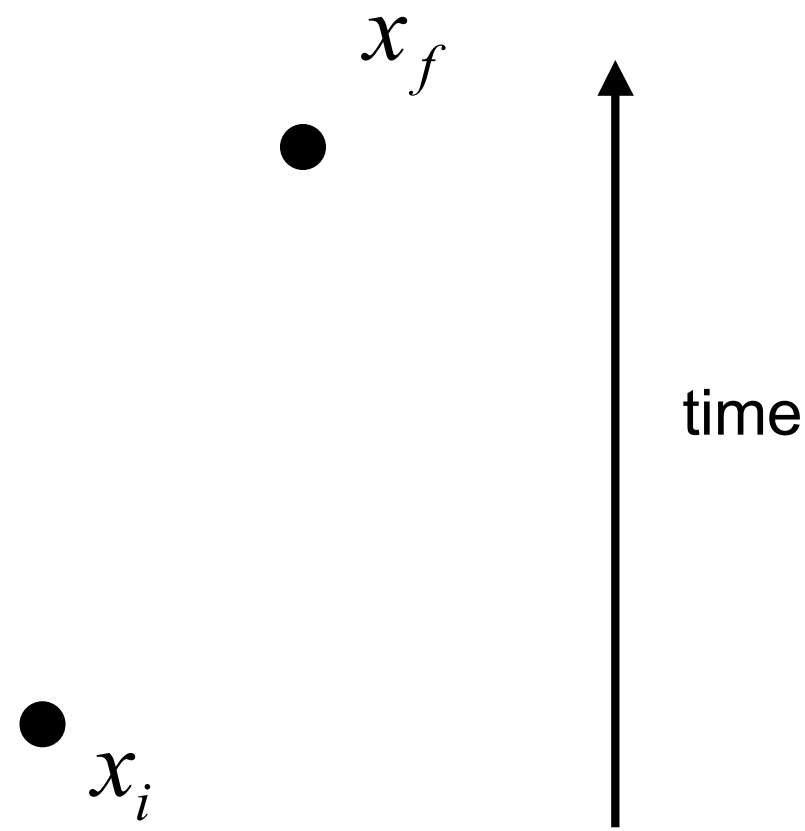
Ariel has also developed a very interesting derivation of quantum mechanics using Cox's method applied to experimental setups rather than logical statements.

The concept of consistency with the order-theoretic structure is central here as well.

Particles and Motion

A particle moves from x_i to x_f

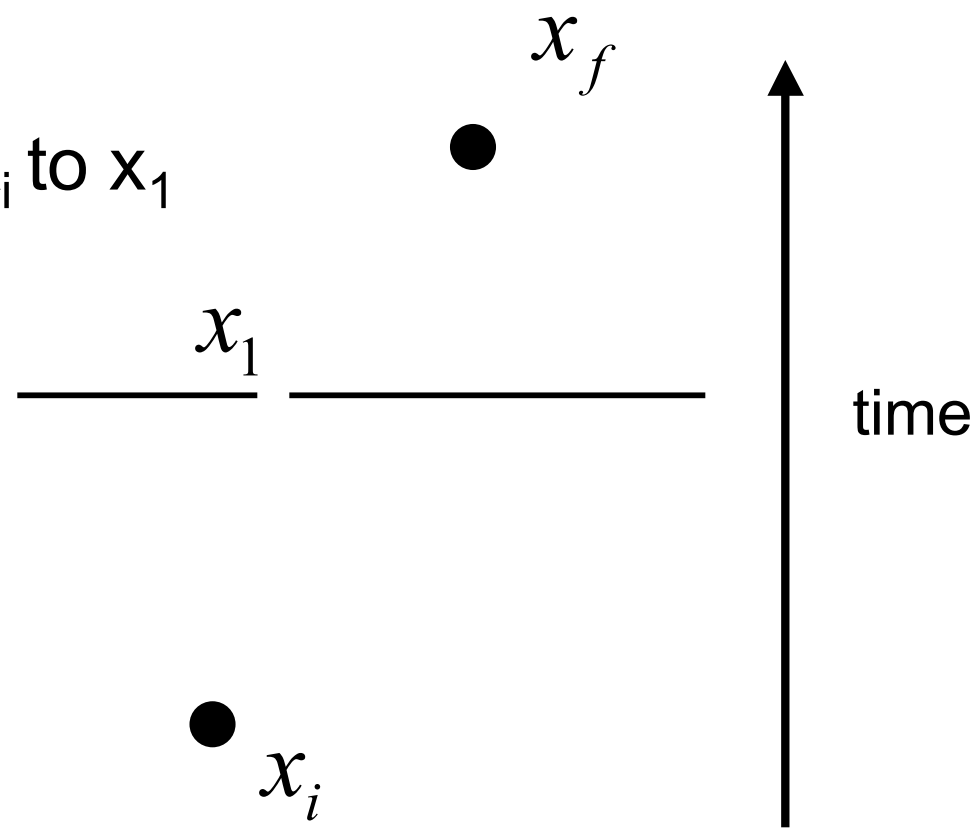
$$[x_f, x_i]$$



A Little More Complex

A particle moves from x_i to x_1 and then from x_1 to x_f

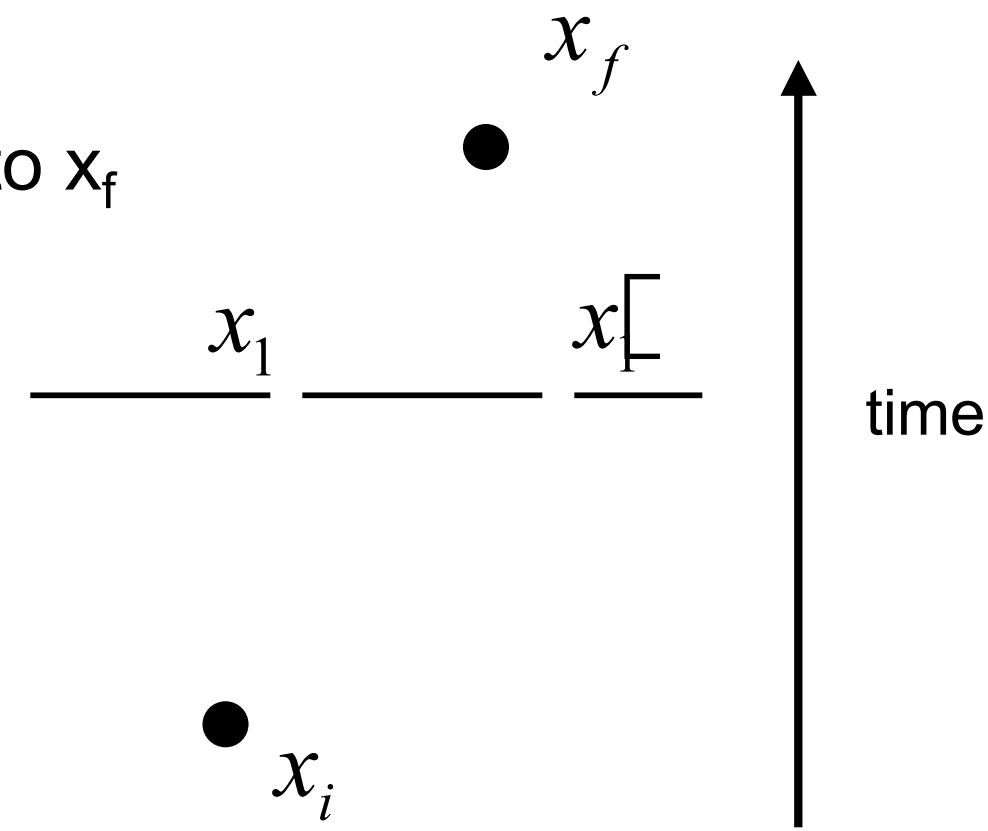
$$[x_f, x_1, x_i]$$



A Little More Complex

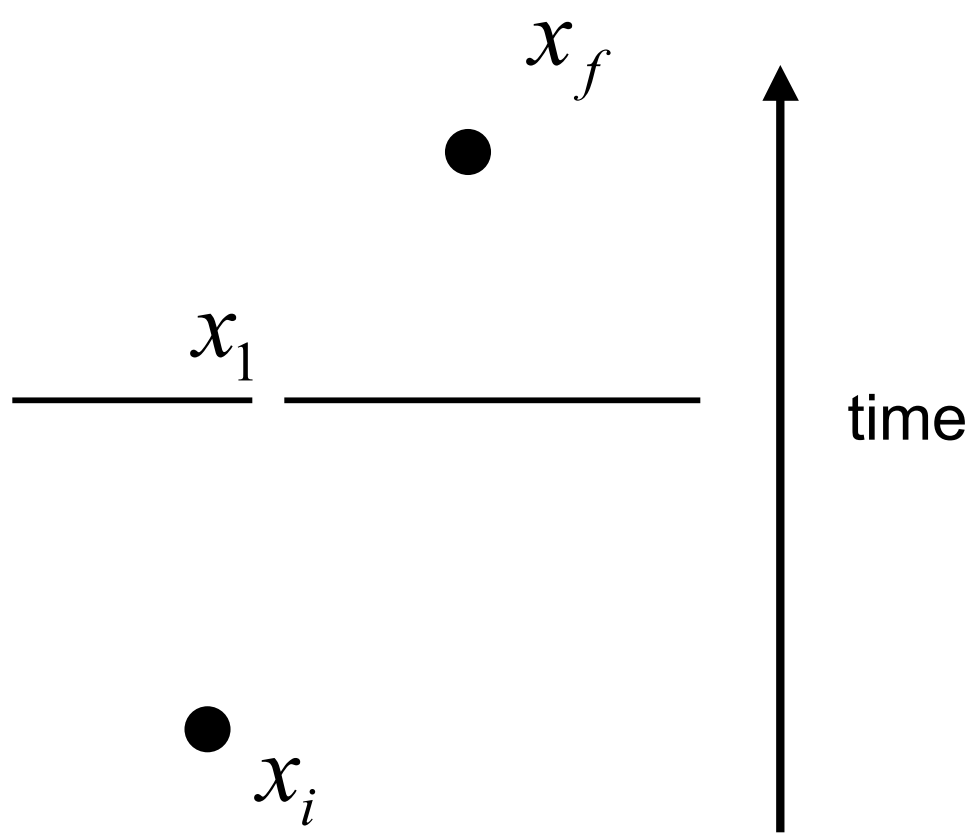
A particle goes from x_i to x_f via x_1 or x'_1

$$[x_f, (x_1, x'_1), x_i]$$



Experimental Setups

We can look at this experimental setup as



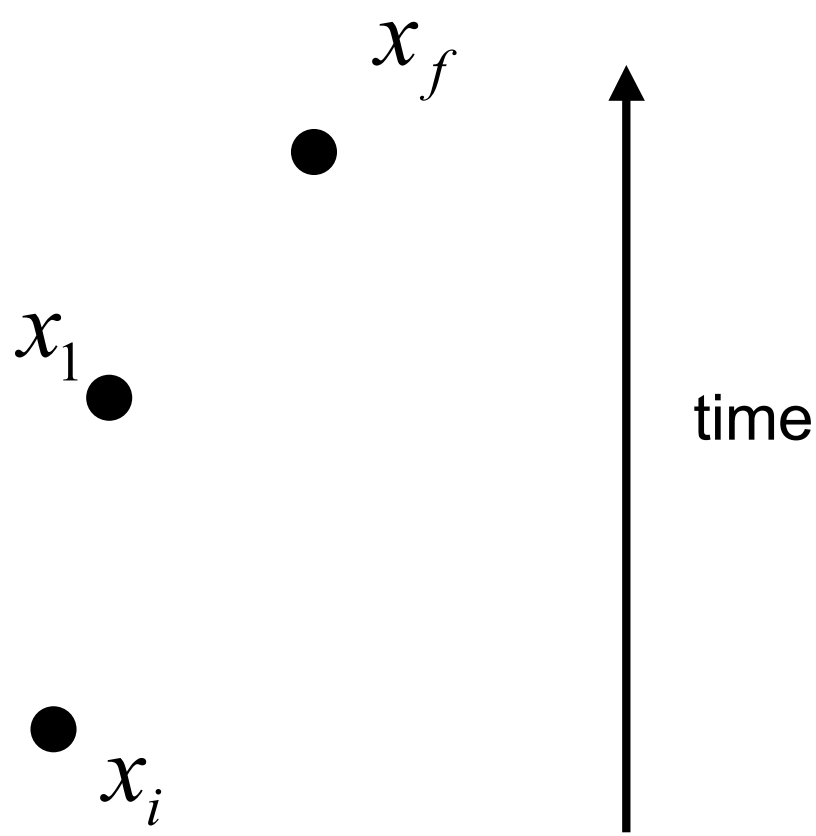
The Meet Operation

We can look at this experimental setup as

being a combination of two setups…

$$[x_f, x_1] \sqcap [x_1, x_i]$$

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

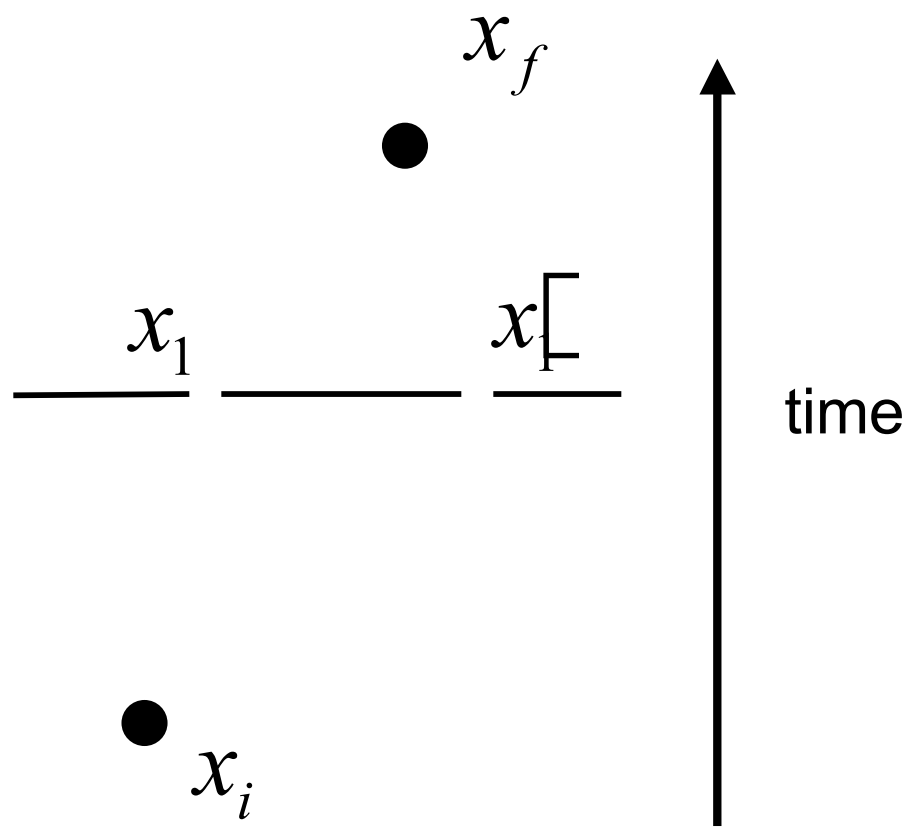


The Join Operation

This is a different way to combine setups

$$[x_f, x_1, x_i] \quad [x_f, x_1, x_i]$$

$$= [x_f, (x_1, x_1), x_i]$$



Associativity of the Meet

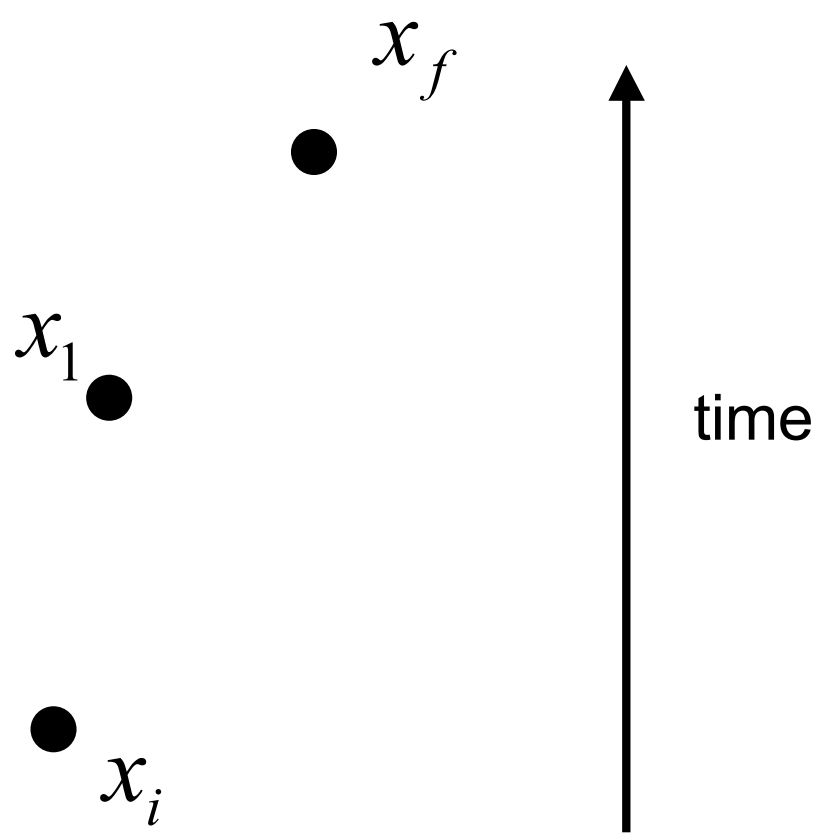
The meet is associative, as

$$
\begin{aligned}
[x_f, x_2] \sqcap [x_2, x_1] \sqcap [x_1, x_i] \\
&= [x_f, x_2] \sqcap \left([x_2, x_1] \sqcap [x_1, x_i]\right) \\
&= \left([x_f, x_2] \sqcap [x_2, x_1]\right) \sqcap [x_1, x_i] \\
&= [x_f, x_2, x_1, x_i]
\end{aligned}
$$

Meet NOT Commutative!

However, experimental setups are not commutative under the meet operation!

This is because it makes no sense for a particle to go from x_1 to x_f and then from x_i to x_1!



Associativity of the Join

The join is also associative, as

$$
\begin{aligned}
[x_f, x_1, x_i] \quad [x_f, x_1, x_i] \quad [x_f, x_1, x_i] \\
&= [x_f, x_1] \quad \left([x_f, x_1] \quad [x_1, x_i]\right) \\
&= \left([x_f, x_1] \quad [x_f, x_1]\right) \quad [x_1, x_i] \\
&= [x_f, (x_1, x_1, x_1), x_i]
\end{aligned}
$$

As long as each of these joins is allowed

Join is Commutative

Joins are commutative

$$[x_f, x_1, \ldots, x_i] \quad [x_f, x_1, \ldots, x_i] = [x_f, x_1, \ldots, x_i] \quad [x_f, x_1, \ldots, x_i]$$

NOT a Lattice Structure

What is interesting about setups is that because not all meets and joins exist, setups do not form a lattice structure.

They do form a poset however.

As the measure we will define is not probability, Ariel represented it with $\square(a)$ rather than $p(a \mid i)$

So lets continue…

Sum and Product Rules Again

Caticha showed that the

Sum Rule is derived from Associativity of the Join.

$$\square(a \quad b) = \square(a) + \square(b)$$

Product Rule from Distributivity.

$$\square(a \square b) = \square(a)\square(b)$$

Amplitudes

If we let the valuations on this poset take on complex values, we have quantum mechanical amplitudes.

Caticha showed that one can then easily derive Schrodinger's Equation.

Feynman Path Integrals are simply analogous to marginalizations.

Probability and QM

Quantum Mechanics is NOT Probability Theory

Probability is a degree of implication defined on a partially ordered set of logical statements.

Amplitudes are an analogous measure defined on the partially ordered set of experimental setups.

This is exciting as it suggests that other analogous measures can be constructed for other partially ordered sets, leading to new laws!

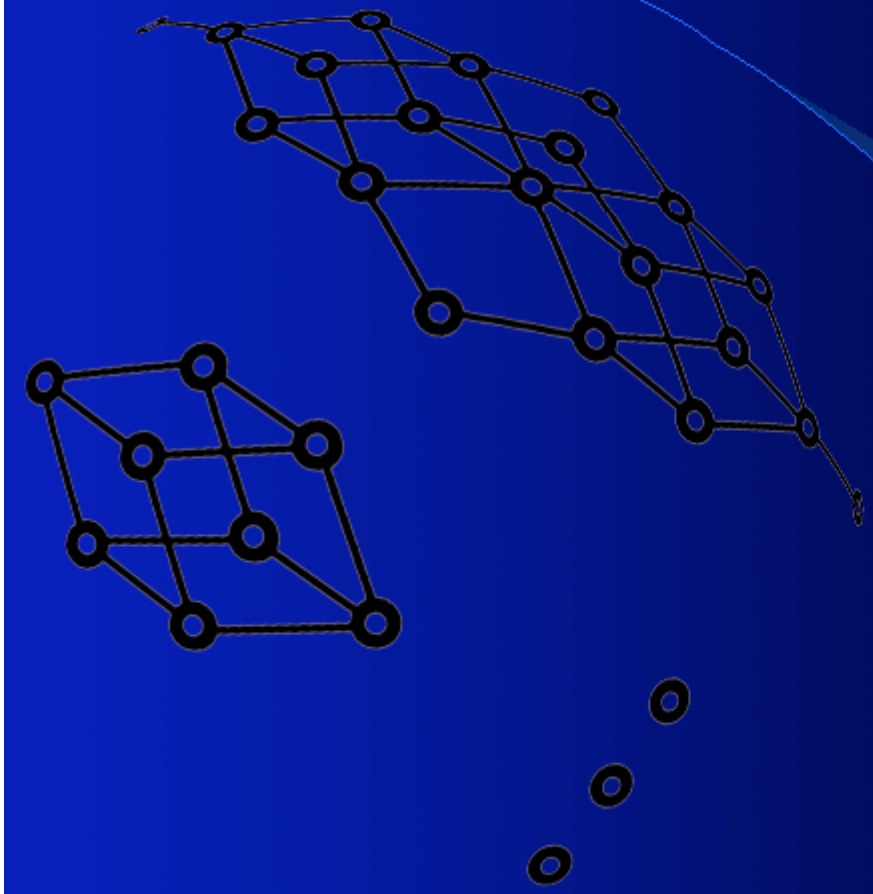
Take Home Message

Symmetry + Order = Laws

I would like to thank Ariel Caticha and Carlos Rodríguez for their discussions, which have enlightened and inspired me.

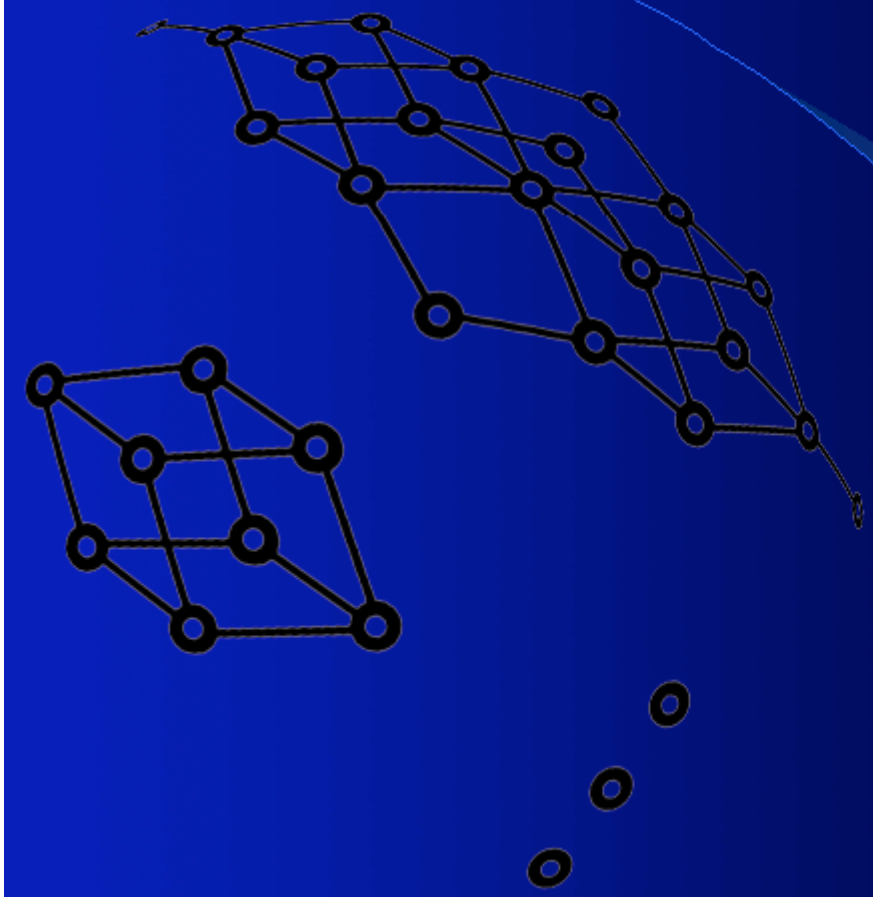
I would like to thank Robert Fry for introducing me to this fascinating area of study.

Deriving Laws From Ordering Relations



Kevin H. Knuth, Ph.D.
Computational Sciences Division
NASA Ames Research Center
Moffett Field CA 94035

Cox Details



Conjunctions

We look at the conjunction of two assertions implied by a premise

$$(a \square\ b \square c)$$

and take as an axiom that this is a function of

$$(a \square\ b) \qquad \text{and} \qquad (a \square b \square\ c)$$

so that

$$(a \square\ b \square c) \quad = \quad F\left[(a \square\ b),(a \square b \square\ c)\right]$$

Conjunctions

We now conjoin an additional assertion

$$(a \square b \square c \square d) = (a \square (b \square c) \square d)$$
$$= F\left[(a \square b \square c), (a \square b \square c \square d)\right]$$

Letting

$$x = (a \square b) \qquad y = (a \square b \square c) \qquad z = (a \square b \square c \square d)$$

We have

$$(a \square b \square c \square d) = F\left[F\left[x, y\right], z\right]$$

Associativity of the Conjunction

We could have grouped the assertions differently

$$
\begin{aligned}
(a \quad b \quad c \quad d) &= (a \quad b \quad (c \quad d)) \\
&= F\left[(a \quad b), (a \quad b \quad c \quad d)\right] \\
&= F\left[(a \quad b), F\left[(a \quad b \quad c), (a \quad b \quad c \quad d)\right]\right] \\
&= F\left[x, F\left[y, z\right]\right]
\end{aligned}
$$

This gives us a functional equation

$$F\left[F\left[x, y\right], z\right] = F\left[x, F\left[y, z\right]\right]$$

The Product Rule

Functional Equation

$$F\left[F\left[x, y\right], z\right] = F\left[x, F\left[y, z\right]\right]$$

As a particular solution we can take $F\left[x, y\right] = xy$
which gives

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

and can be written in a more familiar form by **changing notation**

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

The Product Rule

In general however the solution is

$$F[x, y] = G^{-1}[G[x]G[y]]$$

where G is an **arbitrary** function.

$$G[b \cap c \mid a] = G[b \mid a]G[c \mid a \cap b]$$

We could call G probability!

Logical Complements

The degree to which a premise implies a statement determines the degree to which it implies its contradictory.

$$(a \square \sim b) = f\left[(a \square b)\right]$$

So

$$(a \square \sim(\sim b)) = f\left[(a \square \sim b)\right]$$

$$(a \square \sim(\sim b)) = f\left[f\left[(a \square b)\right]\right]$$

$$(a \square b) = f\left[f\left[(a \square b)\right]\right]$$

$$f\left[f\left[x\right]\right] = x$$

The Sum Rule

Another functional equation

$$f[f[x]] = x$$

A particular solution is which gives $f(x) = 1 - x$

$$(a \cdot b) + (a \cdot \sim b) = 1$$

$$(b \mid a) + (\sim b \mid a) = 1$$

In general

$$g(b \mid a) + g(\sim b \mid a) = C$$

Putting it Together

The solution to the first functional equation puts some constraints on the second

$$G[b \square c \mid a] = G[b \mid a]\, G[c \mid a \square b]$$
$$g[b \mid a] + g[\sim b \mid a] = C$$

The final general solution is and we have

$$G[x] \equiv g[x] = x^r$$

$$(b \square c \mid a)^r = (b \mid a)^r \, (c \mid a \square b)^r$$
$$(b \mid a)^r + (\sim b \mid a)^r = C$$

The Rules of Probability

Setting $r = C = 1$ and writing the function $g(x) = G(x)$ as $p(x)$ we recover the familiar **sum** and **product** rules of probability

$$p(b \wedge c \mid a) = p(b \mid a)\; p(c \mid a \wedge b)$$

$$p(b \mid a) + p(\sim b \mid a) = 1$$

Note that probability is necessarily conditional!
And we never needed the concept of frequencies of events!

The utility of this formalism becomes readily apparent when the implicant is an assertion representing a premise and the implicate is an assertion or proposition representing a hypothesis

$$p(hypothesis \mid premise) \quad \equiv \quad (premise \Rightarrow \; hypothesis)$$

Commutativy of the Conjunction

The symmetry of the conjunction of assertions

$$a \square b \equiv b \square a$$

means that under implication

$$(h \square \; a \square b) \equiv (h \square \; b \square a)$$

also written as

$$p(a \square b \mid h) \equiv p(b \square a \mid h)$$

which means we can write

$$\begin{aligned} p(a \square b \mid h) &= p(a \mid b \square h)\, p(b \mid h) \\ &= p(b \mid a \square h)\, p(a \mid h) \end{aligned}$$

Bayes' Theorem

Likelihood

$$p(model \mid data, I) = p(model \mid I) \frac{p(data \mid model, I)}{p(data \mid I)}$$

Posterior Probability

Prior Probability

Evidence

Bayes' Theorem

Bayes' Theorem is a Learning Rule

Prior Knowledge

$$p(model \mid data, I) = p(model \mid I) \frac{p(data \mid model, I)}{p(data \mid I)}$$

Improved State of Knowledge

Data Dependent Term

Inferential Calculus

In short we have the following calculus:

Product Rule

$$p(x \square y \mid I) = p(y \mid x \square I)\, p(x \mid I)$$ associativity of $\square$

Sum Rule

$$p(x \mid I) + p(\sim x \mid I) = 1$$ complements $x = \sim(\sim x)$

Bayes Theorem

$$p(y \mid x \square I) = p(y \mid I)\frac{p(x \mid y \square I)}{p(x \mid I)}$$ commutativity of $\square$