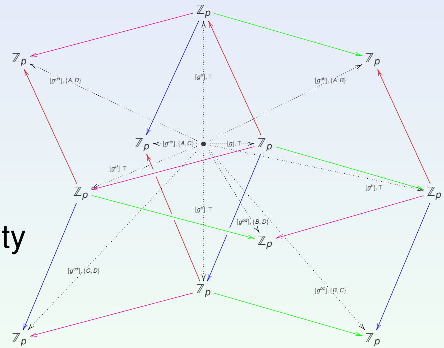
A diagrammatic approach to information flow in encrypted communication

Peter M. Hines
Y.C.C.S.A. , Univ. York



GraMSec
Graphical Models for Security
(online) – June 2020

peter.hines@york.ac.uk www.peterhines.info

An overview ...

This talk is about using tools from category theory to reason about communication:

1. What is category theory?
   - Motivation, definitions, & history.
   - Current theory & applications.
   - Useful tools: diagrammatic & otherwise
2. Why might it be useful for communication?
   - Graphical descriptions of protocols & communication.
   - Reasoning as diagram manipulation.

'Category theory for communication', not vice versa!

Category Theory – the original motivation

A formalism for reasoning about the 'large-scale' properties of mathematical structures.

We might consider the 'category' of all **groups**, or all **rings**, or even all **sets**, etc., and study their properties and relationships with each other.

A category consist of **objects** and **arrows** :

Objects All mathematical structures of a certain kind.

Arrows Structure-preserving mappings between objects.

Composition Arrows may be composed ...

Beyond topology: the spread of category theory

Why should we be interested?

More recently, category theory has been used to model **information flow** in :

- Formal Logic & Deduction
- Quantum algorithms & protocols
- Theoretical & practical computer science,
- Linguistics & natural language processing,
- Cognitive science & psychology.

Why – what is the appeal?

These often use *very simple tools* developed for use within category theory, rather than the actual theory itself.

There's something about category theory ...

Diagrammatic reasoning

Category theory frequently expresses *equations* as *pictures*.

Algebraic manipulations are replaced by *diagram-chasing*.

Our simple aims :

1. Express protocols / communication generally using such graphical tools,
2. Use 'diagram-chasing' to reason about them.

The definition ...

A **category** $\mathcal{C}$ consists of a **class** of objects, $Ob(\mathcal{C})$ and a **set** of arrows $\mathcal{C}(A, B)$ between any two objects.

- Matching arrows can be composed

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "gf"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

- Composition is associative

$$h(gf) = (hg)f$$

- There is an identity 1_A at each object A

These are the tools we are looking for ...

Identities and equations are traditionally expressed graphically.

A **diagram** in the category **Set**

$$\begin{tikzcd}
\mathbb{Z} \arrow[rr, "x \mapsto x^2"] \arrow[d, "x \mapsto abs(x)"'] & & \mathbb{N} \arrow[d, "n \mapsto n \ (mod\ 2)"] \\
\mathbb{N} \arrow[urr, "n \mapsto n^2"] \arrow[rr, "n \mapsto n \ (mod\ 2)"'] & & \{0, 1\}
\end{tikzcd}$$

A diagram **commutes** when all paths with the same source / target describe the same arrow.

peter.hines@york.ac.uk www.peterhines.info

A passing observation!

The **word problem** for groups / monoids is a special case of **deciding commutativity** of diagrams.

Some simple arithmetic bijections ...

$$X(n) = \begin{cases} n & n \ (mod\ 2) = 0 \\ 2n-1 & n \ (mod\ 4) = 1 \\ n+2 & n \ (mod\ 8) = 3 \\ \frac{n-1}{2} & n \ (mod\ 8) = 7 \end{cases} \qquad Y(n) = \begin{cases} 2n & n \ (mod\ 4) = 0 \\ n+2 & n \ (mod\ 8) = 2 \\ \frac{n+1}{2} & n \ (mod\ 8) = 6 \\ n & n \ (mod\ 2) = 1 \end{cases}$$

$$Z(n) = \begin{cases} 4n & n \ (mod\ 2) = 0 \\ n+2 & n \ (mod\ 4) = 1 \\ \frac{n+1}{2} & n \ (mod\ 8) = 3 \\ \frac{n-3}{4} & n \ (mod\ 8) = 7 \end{cases} \qquad T(n) = \begin{cases} 2n & n \ (mod\ 2) = 0 \\ n+1 & n \ (mod\ 4) = 1 \\ \frac{n-1}{2} & n \ (mod\ 4) = 3 \end{cases}$$

We may prove this diagram commutes :

$$\begin{tikzcd}
\mathbb{N} \arrow[r, "Y"] & \mathbb{N} & \mathbb{N} \arrow[l, "T"'] \\
\mathbb{N} \arrow[u, "T"] & \mathbb{N} \arrow[l, "X"] \arrow[u, "Z"'] \arrow[ur, "T"'] &
\end{tikzcd}$$

but how easily can we decide commutativity for *arbitrary* diagrams over $\{X, Y, Z, T\}$?

A simple aim!

We wish to use a single diagram to model

- Underlying algebra
- Knowledge of participants
- Information flow

The aims :

1. Make things clearer by drawing them as pictures!
2. Interpret commutativity / failure of commutativity in terms of communication.
3. Develop tools for (graphical) reasoning about communication.

Illustration by example

Commuting Action Key Exchange (CAKE)

- A general prescription for key exchange protocols.
- Introduced in 2004 by V. Shpilrain & G. Zapata
- Includes many interesting protocols as special cases

We will look at the monoid-theoretic version:

Example 3, Section 3 of *Combinatorial Group Theory and Public Key Cryptography* S.-Z. (2004).

CAKE – sharing protocol

Alice and Bob will come to share a secret element of a monoid $\mathcal{M}$.

1. Alice and Bob both have large **key pools** $A, B \subseteq \mathcal{M}$ that satisfy

$$ab = ba \ \ \forall \ a \in A, \ b \in B.$$

2. A fixed public **root element** $\gamma \in \mathcal{M}$ is chosen.
3. Alice chooses her **private key**, $(\alpha_1, \alpha_2) \in A \times A$, and publicly broadcasts $\alpha_1 \gamma \alpha_2 \in \mathcal{M}$
4. Bob chooses his **private key**,$(\beta_1, \beta_2) \in B \times B$, and publicly broadcasts $\beta_1 \gamma \beta_2 \in \mathcal{M}$.
5. Alice computes $\alpha_1 \beta_1 \gamma \beta_2 \alpha_2$ and Bob computes $\beta_1 \alpha_1 \gamma \alpha_2 \beta_2$.

By the point-wise commutativity of $A, B \subseteq \mathcal{M}$, these are equal, giving Alice and Bob's **shared secret** σ as

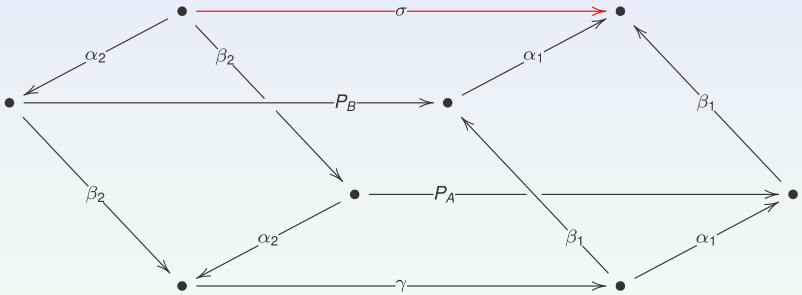
$$\sigma \ = \ \alpha_1 \beta_1 \gamma \beta_2 \alpha_2 \ = \ \beta_1 \alpha_1 \gamma \alpha_2 \beta_2$$

The algebra of CAKE

The required arrows are:

1. The root γ
2. Alice & Bob's private keys, (α_1, α_2) and (β_1, β_2)
3. Alice & Bob's public announcements, P_A and P_B
4. Their shared secret σ

Expressing the required relationships as a commuting diagram :

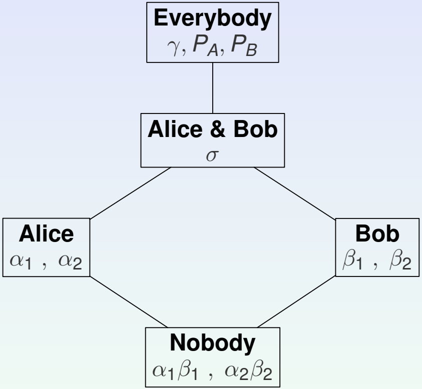


peter.hines@york.ac.uk www.peterhines.info

Knowns and unknowns in semigroup CAKE

In this protocol, who comes to know what?

The epistemic data:



peter.hines@york.ac.uk

Combining algebraic & epistemic data

Introducing epistemic data to diagrams

- Form the subset-lattice of participants.
- Label each edge in the diagram by an element of this lattice:

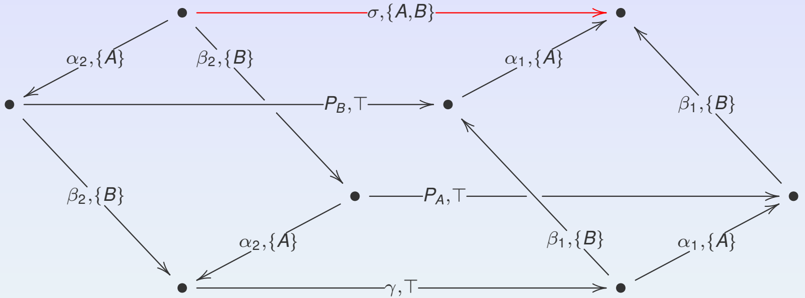
$$\bullet \xrightarrow{f,X} \bullet$$

$X \subseteq \{Alice, Bob, Eve\}$ consists of participants who

- know the value of f, or (more accurately)
- are able to perform the operation f.

CAKE, in summary

The **Algebraic-Epistemic (A-E) diagram** for semigroup-CAKE:



What is and is not shown!

This diagram summarises the ‘final state of affairs’ : who ends up knowing what. We are interested in *deducing* implicit information such as ordering of events, communication between participants, etc.

Commuting diagrams??

Treating $2^{\{A,B,E\}}, \cap$ as a monoid:

Question: Is this diagram for CAKE a commuting diagram over the product category $\mathcal{M} \times 2^{\{A,B,E\}}$?

Answer: No!

Turning a bug into a feature: *The reasons why / points at which it fails to commute are highly significant.*

1. Announcements / information sharing by participants.
2. Different routes to calculating the same value.

Failure of commutativity & public announcements

Diagram 1 commutes, Diagram 2 is from CAKE.

$$\begin{tikzcd} \bullet \arrow[rr, "\beta_1\gamma\beta_2{,}\{B\}"] \arrow[d, "\beta_2{,}\{B\}"'] & \text{Diagram 1} & \bullet \\ \bullet \arrow[rr, "\gamma{,}\top"'] & & \bullet \arrow[u, "\beta_1{,}\{B\}"'] \end{tikzcd} \qquad \begin{tikzcd} \bullet \arrow[rr, "\beta_1\gamma\beta_2{,}\top"] \arrow[d, "\beta_2{,}\{B\}"'] & \text{Diagram 2} & \bullet \\ \bullet \arrow[rr, "\gamma{,}\top"'] & & \bullet \arrow[u, "\beta_1{,}\{B\}"'] \end{tikzcd}$$

1. In **diagram 1**, Bob computes $\beta_2\gamma\beta_1$.
2. In **diagram 2**, Bob computes $\beta_2\gamma\beta_1$, and announces the result.

Public announcements as inequalities

The points at which announcements have been made appear as *inequalities*:

$$
\begin{tikzcd}
\bullet \arrow[rr, "{\beta_1\gamma\beta_2, \top}"] \arrow[d, "{\beta_2, \{B\}}"'] & & \bullet \\
\bullet \arrow[rr, "{\gamma, \{A,B,E\}}"'] \arrow[r, phantom, ""] & \arrow[u, phantom, "\Uparrow \leqslant"] & \bullet \arrow[u, "{\beta_1, \{B\}}"']
\end{tikzcd}
$$

From a category-theory viewpoint ...

Public announcements lead to failure of commutativity.

The other way commutativity fails :

In another sub-diagram of CAKE, we have failure of commutativity without announcements :

$$\begin{tikzcd}
 & \bullet \arrow[rrr, red, "{\sigma,\{A,B\}}"] \arrow[dl, "{\alpha_2,\{A\}}"'] \arrow[ddr, "{\beta_2,\{B\}}"] & & & \bullet & \\
\bullet \arrow[rrr, "{P_B,\top}"] & & & \bullet \arrow[ur, "{\alpha_1,\{A\}}"] & & \\
 & & \bullet \arrow[rrr, "{P_A,\top}"] & & & \bullet \arrow[ul, "{\beta_1,\{B\}}"']
\end{tikzcd}$$

Here, the non-trivial orderings

- $(\alpha_1, \{A\})(P_B, \top)(\alpha_2, \{A\}) \ < \ (\sigma, \{A, B\})$
- $(\beta_1, \{B\}(P_A, \top)(\beta_2, \{B\}) \ < \ (\sigma, \{A, B\})$

arise because Alice and Bob take distinct routes to calculating the shared secret.

A simple definition ...

A diagram $\mathfrak{D}$ over an order-enriched category is the **information flow ordered (IFO)** when:

1. The underlying digraph is acyclic.
2. For any edge e and path $p = p_k \ldots \{V, W\}$ with the same source and target node, the label on p is $\leqslant$ the label on e.

   We draw this diagrammatically as a “2-cell”:

$$\begin{tikzcd} & \cdots \arrow[dr, "g_n"] \arrow[d, phantom, "\Downarrow \leqslant"] & \\ X \arrow[ur, "g_1"] \arrow[rr, "f"'] & & Y \end{tikzcd}$$

(Terminology from 2-category theory ...) Algebraically,

$$g_n g_{n-1} \ldots g_1 \;\leqslant\; f$$

peter.hines@york.ac.uk www.peterhines.info

Interpreting the edge-path condition

We claim this as a generic ‘correctness criterion’ for A-E diagrams.

If it fails, then either:

1. We have failed to account for the results of some announcement,
2. We have missed some route to calculating a secret value,

This is about information flow: nothing at all to do with the difficult of solving problems!

The IFO condition: who knows what?

Consider a fragment of the A-E diagram for some protocol:

$$\begin{tikzcd}
\bullet \arrow[r, "a_2{,}R_2"] & \cdots \arrow[r, "a_{n-1}{,}R_{n-1}"] & \bullet \arrow[d, "a_n{,}R_n"] \\
H \arrow[u, "a_1{,}R_1"] \arrow[rr, "b{,}Q"'] & & K
\end{tikzcd}$$

The IFO condition states that

$$b = a_n \ldots a_1 \quad \text{and} \quad \bigcap_{j=1}^{n} R_j \subseteq Q$$

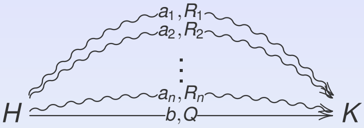
Quite simply:

Every individual $x \in \bigcap_{j=1}^{n} R_j$ knows every operation $\{a_j\}_{j=1..n}$ and therefore also knows their composite $a_n \ldots a_1$.

peter.hines@york.ac.uk www.peterhines.info

No participant left behind

Consider a fragment of an A-E diagram for some protocol with a **single edge** and **multiple paths** from node H to node K.



The IFO condition states that $R_j \subseteq Q$ for all $j = 1..n$.

Again, a simple interpretation:

The members of $R_1, R_2, \ldots, R_n$ are all able to calculate (perform) b, albeit in different ways. Therefore, the set of participants who can perform b must contain each R_j.

Other forms of key-exchange :

Tripartite Diffie-Hellman

A familiar story

Three participants $\{Alice, Bob, Carol\}$ wish to communicate privately, using Diffie-Hellman key exchange.

Using their private keys $a, b, c \in \mathbb{Z}_p$, they may either :

1. produce a single shared secret, $g^{abc} = g^{bca} = g^{cab}$
2. produce a distinct shared secret for each pair:

   Alice - Bob $g^{ab} = g^{ba}$
   Bob - Carol $g^{bc} = g^{cb}$
   Carol - Alice $g^{ca} = g^{ac}$

These give two very distinct A-E diagrams over the same category.

The underlying category

The action takes place in a small subcategory of **Set**:

- **Objects:** $\mathbb{Z}_p$ and $\{*\}$
- **Arrows:**
  1. *modular exponentiation* $(\ )^x : \mathbb{Z}_p \rightarrow \mathbb{Z}_p$, for all $x = 0 \ldots p-1$
  2. *selecting an element* $[x] : \{\star\} \rightarrow \mathbb{Z}_p$, where $[x](\star) = x \in \mathbb{Z}_p$

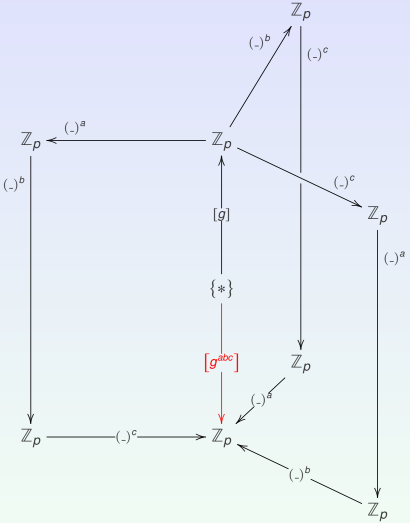
Constructing a single shared secret (I)

The basic identity is $(((\_)^a)^b)^c = (((\_)^b)^c)^a = (((\_)^c)^a)^b$

$$
\begin{tikzcd}
 & & \mathbb{Z}_p \arrow[dd, "(\_)^c"] & \\
\mathbb{Z}_p \arrow[dd, "(\_)^b"'] & \mathbb{Z}_p \arrow[l, "(\_)^a"'] \arrow[ur, "(\_)^b"] \arrow[drr, "(\_)^c"] & & \\
 & & \mathbb{Z}_p \arrow[dl, "(\_)^a"] & \mathbb{Z}_p \arrow[dd, "(\_)^a"] \\
\mathbb{Z}_p \arrow[r, "(\_)^c"] & \mathbb{Z}_p & & \\
 & & & \mathbb{Z}_p \arrow[ull, "(\_)^b"]
\end{tikzcd}
$$

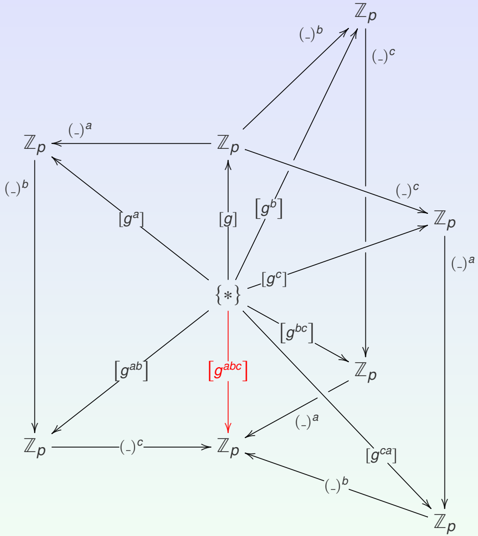
Constructing a single shared secret (II)

We require these equalities *applied to the root* $g \in \mathbb{Z}_p$.



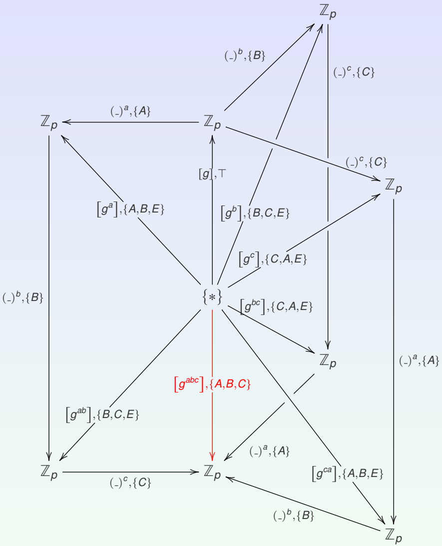
Constructing a single shared secret (III)

The elements $g^a, g^b, g^c, g^{ab}, g^{bc}, g^{ca}$ are all announced:



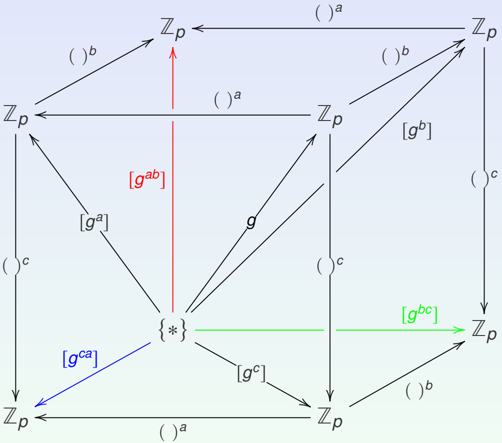
Constructing a single shared secret (IV)

Adding in the ‘who-knows-what’ data, we get the A-E diagram :



Constructing three distinct shared secrets (I)

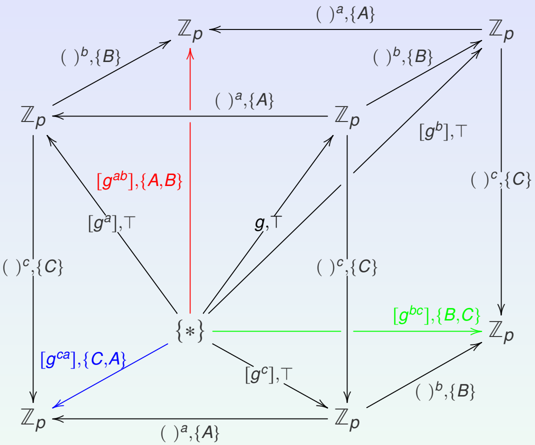
Going through the same procedure for the case of three distinct shared secrets, we get the (commuting) diagram describing the algebra :



peter.hines@york.ac.uk

Constructing three distinct shared secrets (II)

Adding in the epistemic information, we get the A-E diagram



Is there any advantage to this ?

Drawing pictures of protocols may be fun but ... what can we actually do?

Simple diagram-chasing gives us a *systematic* route to answering questions such as :

- Can any additional information be announced without compromising the protocol?
- What happens when Eve discovers (say) Bob's secret key?
- Are these two approaches equivalent?

(All already thoroughly understood – we are *testing the formalism* by asking questions where we already know the answer.)

Can we go further??

Drawing diagrams gives a *visual representation* of algebraic relationships, epistemic knowledge, and information flow.

We can use standard 'diagram-chasing' techniques to answer questions about information flow.

They are also convenient for dealing with *partial information*.

Deductions from partial information

Consider the situation where we have partial information about (for example) which communications have taken place.

Representing as much as we know, diagramatically, we have arrived at:

$$
\begin{tikzcd}
 & \bullet \arrow[dr, "{c,\{X,Y\}}"] & \\
\bullet \arrow[ur, "{b,\{W,X\}}"] & & \bullet \arrow[d, "{d,\{Y,Z\}}"] \\
\bullet \arrow[u, "{a,\{V,W\}}"] \arrow[rr, "{dcba,\top}"'] & & \bullet
\end{tikzcd}
$$

Can we deduce the possible routes by which the composite *dcba* became public knowledge?

— as a starting point, no single individual could have announced this without assistance!

Unambiguous diagrams?

A class of diagrams where announcements are unambiguous :

An A-E diagram is $\mathfrak{D}$ is **triangulated** when every non-identity 2-cell is decomposed into composites of identity two-cells, and non-identity two-cells consisting of three edges.

$$
\begin{tikzcd}
\bullet \arrow[r] \arrow[dr, phantom, "\Searrow"] & \bullet \\
\bullet \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

We wish to consider the possible ways in which that a given diagram is a subdiagram of a triangulated IFO diagram.

peter.hines@york.ac.uk www.peterhines.info

Different options (I)

$\mathfrak{D}_1$

$$\begin{tikzcd}
\bullet \arrow[r, "{b,\{W,X\}}"] & \bullet \arrow[r, "{c,\{X,Y\}}"] \arrow[dr, "{dc,\top}"] & \bullet \arrow[d, "{d,\{Y,Z\}}"] \\
\bullet \arrow[u, "{a,\{V,W\}}"] \arrow[ur, "{ba,\top}"] \arrow[rr, "{dcba,\top}"'] & & \bullet
\end{tikzcd}$$

$\mathfrak{D}_2$

$$\begin{tikzcd}
& \bullet \arrow[dr, "{c,\{X,Y\}}"] & \\
\bullet \arrow[ur, "{b,\{W,X\}}"] \arrow[rr, "{cb,\top}"] & & \bullet \arrow[d, "{d,\{Y,Z\}}"] \\
\bullet \arrow[u, "{a,\{V,W\}}"] \arrow[rr, "{dcba,\top}"'] & & \bullet
\end{tikzcd}$$

- Diagram $\mathfrak{D}_1$ is triangulated. W has publicly announced ba and Z has publicly announced dc; any participant may now compute $dcba$.
- Diagram $\mathfrak{D}_2$ is still not triangulated; there remains ambiguity about how $dcba$ came to be public knowledge.

Different options (II)

Diagram $\mathfrak{D}_2$ may be triangulated in two different ways :

$\mathfrak{D}_3$

$$\begin{tikzcd}
 & \bullet \arrow[dr, "{c,\{X,Y\}}"] & \\
\bullet \arrow[ur, "{b,\{W,X\}}"] \arrow[rr, "{cb,\top}"] & & \bullet \arrow[d, "{d,\{Y,Z\}}"] \\
\bullet \arrow[u, "{a,\{V,W\}}"] \arrow[urr, "{cba,\top}"] \arrow[rr, "{dcba,\top}"'] & & \bullet
\end{tikzcd}$$

$\mathfrak{D}_4$

$$\begin{tikzcd}
 & \bullet \arrow[dr, "{c,\{X,Y\}}"] & \\
\bullet \arrow[ur, "{b,\{W,X\}}"] \arrow[rr, "{cb,\top}"] \arrow[drr, "{dcb,\top}"] & & \bullet \arrow[d, "{d,\{Y,Z\}}"] \\
\bullet \arrow[u, "{a,\{V,W\}}"] \arrow[rr, "{dcba,\top}"'] & & \bullet
\end{tikzcd}$$

- In diagram $\mathfrak{D}_3$, either V or W has announced *cba*, then either Y or Z has announced *dcba*.
- In diagram $\mathfrak{D}_4$, either Y or Z has announced *dcb* followed by either U or V announcing *dcba*.

Combinatorics vs. Common Sense

Elementary combinatorics (& a bit of recursion) will allow us to give all IFO triangulations of a given diagram.

— what can we conclude from these?

Some caution is needed!

We derive *some* potential scenarios for information flow.

Bear in mind our own assumptions.

1. Are we aware of all participants?
2. Is our understanding of their knowledge accurate?
3. Are there other ways to calculate information that we have not accounted for?
4. . . .

peter.hines@york.ac.uk