Two orders for decomposing multivariate information

Conor Finn

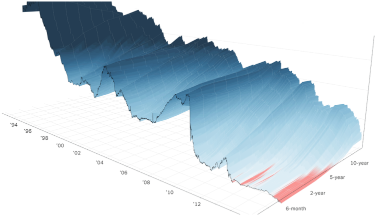
Information Processing in Complex Systems 2022

October 20, 2022

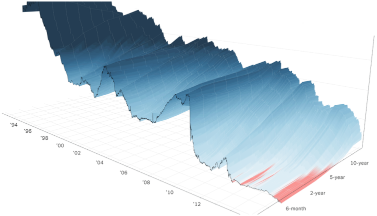
Max Planck Institute for Mathematics in the Sciences

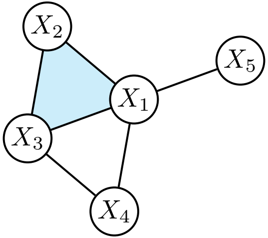
MAX-PLANCK-GESELLSCHAFT

Complex systems



Complex systems





Information theory

- Represent each component using a random variable X_i
- The **entropy** then quantifies our uncertainty about each component

$$H(X_i) = -\sum_{x_i \in X_i} p(x_i) \log p(x_i) \geq 0$$

- The **mutual information** quantifies the dependence between components

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

- Advantages for complex systems:

Information theory

- Represent each component using a random variable X_i
- The **entropy** then quantifies our uncertainty about each component

$$H(X_i) = -\sum_{x_i \in X_i} p(x_i) \log p(x_i) \geq 0$$

- The **mutual information** quantifies the dependence between components

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

- Advantages for complex systems:
  - Captures both linear and non-linear dependencies

Information theory

- Represent each component using a random variable X_i
- The **entropy** then quantifies our uncertainty about each component

$$H(X_i) = -\sum_{x_i \in X_i} p(x_i) \log p(x_i) \geq 0$$

- The **mutual information** quantifies the dependence between components

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

- Advantages for complex systems:
  - Captures both linear and non-linear dependencies
  - Variables can represent very different quantities

Information theory

- Represent each component using a random variable X_i
- The **entropy** then quantifies our uncertainty about each component

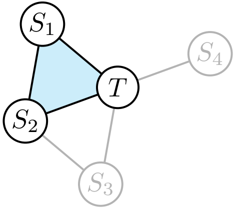
$$H(X_i) = -\sum_{x_i \in X_i} p(x_i) \log p(x_i) \geq 0$$

- The **mutual information** quantifies the dependence between components

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

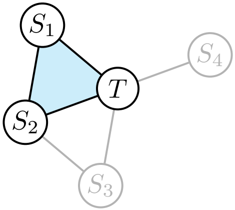
- Advantages for complex systems:
  - Captures both linear and non-linear dependencies
  - Variables can represent very different quantities
  - Model free

Information theory and higher-order interactions



- There are four distinct ways in which T can depend on S_1 and S_2

Information theory and higher-order interactions

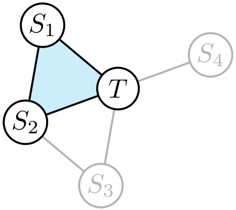


- There are four distinct ways in which T can depend on S_1 and S_2

Uniquely on S_1

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Information theory and higher-order interactions



- There are four distinct ways in which T can depend on S_1 and S_2

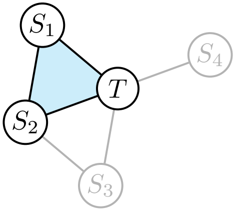
Uniquely on S_1

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Uniquely on S_2

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	0
1/4	1	1	1

Information theory and higher-order interactions



- There are four distinct ways in which T can depend on S_1 and S_2

Uniquely on S_1

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

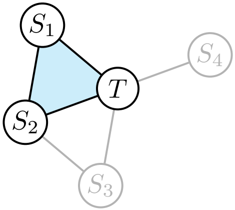
Uniquely on S_2

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	0
1/4	1	1	1

Redundantly

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/2	0	0	0
1/2	1	1	1

Information theory and higher-order interactions



- There are four distinct ways in which T can depend on S_1 and S_2

Uniquely on S_1

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Uniquely on S_2

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	0
1/4	1	1	1

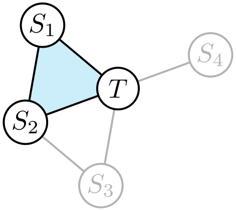
Redundantly

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/2	0	0	0
1/2	1	1	1

Higher-order

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	1
1/4	1	1	0

Information theory and higher-order interactions



- There are four distinct ways in which T can depend on S_1 and S_2

Uniquely on S_1

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Uniquely on S_2

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	0
1/4	1	1	1

Redundantly

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/2	0	0	0
1/2	1	1	1

Higher-order

$\boldsymbol{p}$	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{t}$
1/4	0	0	0
1/4	0	1	1
1/4	1	0	1
1/4	1	1	0

- Three ways to use the mutual information: $I(S_1;T)$, $I(S_2;T)$ and $I\big((S_1,S_2);T\big)$

The problem

Shannon information theory is ill-equipped for analysing complex systems

The problem

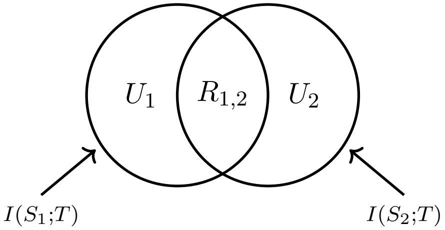
Shannon information theory is ill-equipped for analysing complex systems

Solution: information decomposition

Information decomposition

- Mutual information captures

$$I(S_1;T) = U_1 + R_{1,2}$$
$$I(S_2;T) = U_2 + R_{1,2}$$



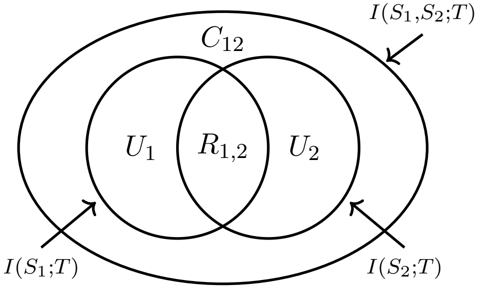
Information decomposition

- Mutual information captures

$$I(S_1;T) = U_1 + R_{1,2}$$
$$I(S_2;T) = U_2 + R_{1,2}$$

- Joint mutual information captures

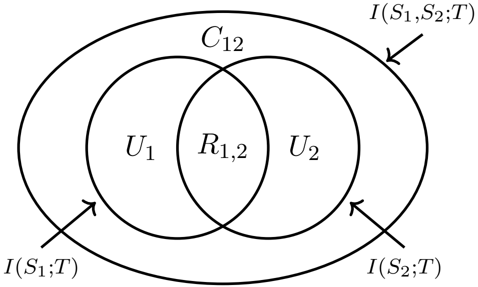
$$I\big((S_1, S_2);T\big) = U_1 + U_2 + R_{1,2} + C_{12}$$



Information decomposition

- Mutual information captures

$$I(S_1;T) = U_1 + R_{1,2}$$
$$I(S_2;T) = U_2 + R_{1,2}$$



- Joint mutual information captures

$$I\big((S_1, S_2);T\big) = U_1 + U_2 + R_{1,2} + C_{12}$$

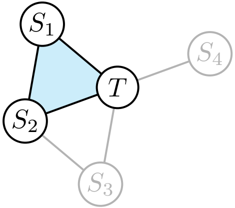
- Can we define one of the quantities to solve the system?
- How do we generalise this idea to consider more variables?

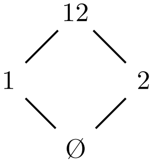
Partial information decomposition

- Axiomatic framework for information decomposition (Williams and Beer, 2010)

Partial information decomposition

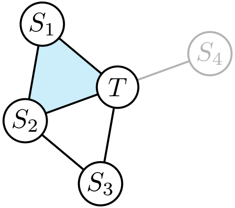
- Axiomatic framework for information decomposition (Williams and Beer, 2010)
- Consider each way n source variables can provide information about T

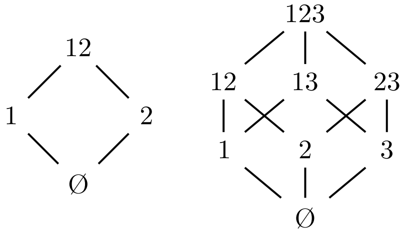




Partial information decomposition

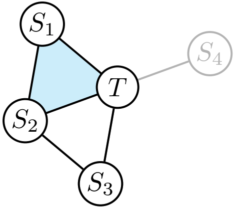
- Axiomatic framework for information decomposition (Williams and Beer, 2010)
- Consider each way n source variables can provide information about T

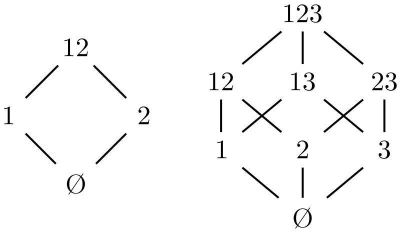




Partial information decomposition

- Axiomatic framework for information decomposition (Williams and Beer, 2010)
- Consider each way n source variables can provide information about T

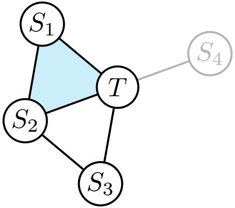


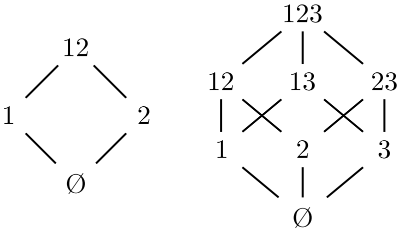


- Let $\boldsymbol{A}_i$ represent a distinct way sources can provide information T

Partial information decomposition

- Axiomatic framework for information decomposition (Williams and Beer, 2010)
- Consider each way n source variables can provide information about T

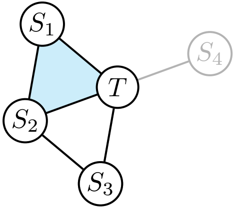


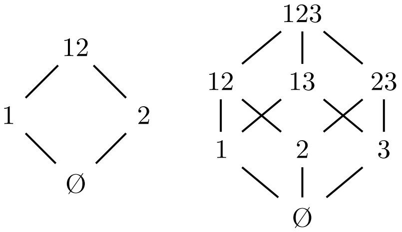


- Let $\boldsymbol{A}_i$ represent a distinct way sources can provide information T
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$

Partial information decomposition

- Axiomatic framework for information decomposition (Williams and Beer, 2010)
- Consider each way n source variables can provide information about T





- Let $\boldsymbol{A}_i$ represent a distinct way sources can provide information T
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
- Define a function $I_\cap$ that quantifies the redundant information provided by the $\boldsymbol{A}_i$'s

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- $I_\cap$ can be applied to any combination of $\boldsymbol{A}_i$'s, but many are equivalent
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - But by Axiom 3, $I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_3; T) = I_\cap(\boldsymbol{A}_1; T)$ since $\boldsymbol{A}_1 \subseteq \boldsymbol{A}_3$

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- $I_\cap$ can be applied to any combination of $\boldsymbol{A}_i$'s, but many are equivalent
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - But by Axiom 3, $I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_3; T) = I_\cap(\boldsymbol{A}_1; T)$ since $\boldsymbol{A}_1 \subseteq \boldsymbol{A}_3$
- Left with all combinations of sources $\boldsymbol{A}_i$ s.t. no source is a subset of any other
  - For $n = 2$, we have

$$I_\cap(\boldsymbol{A}_1; T), \quad I_\cap(\boldsymbol{A}_2; T), \quad I_\cap(\boldsymbol{A}_3; T) \quad \text{and} \quad I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_2; T)$$

- The remaining combinations of sources are the antichains of the inclusion lattice

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over the remaining combinations

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over the remaining combinations
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - By Axiom 3, $I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_2; T) \leq I_\cap(\boldsymbol{A}_1; T)$

Redundant information

Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over the remaining combinations
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - By Axiom 3, $I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_2; T) \leq I_\cap(\boldsymbol{A}_1; T)$
  - Similarly, $I_\cap(\boldsymbol{A}_1; T) = I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_3; T) \leq I_\cap(\boldsymbol{A}_3; T)$ since $\boldsymbol{A}_1 \subseteq \boldsymbol{A}_3$

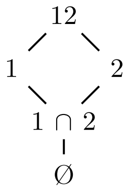
Redundant information

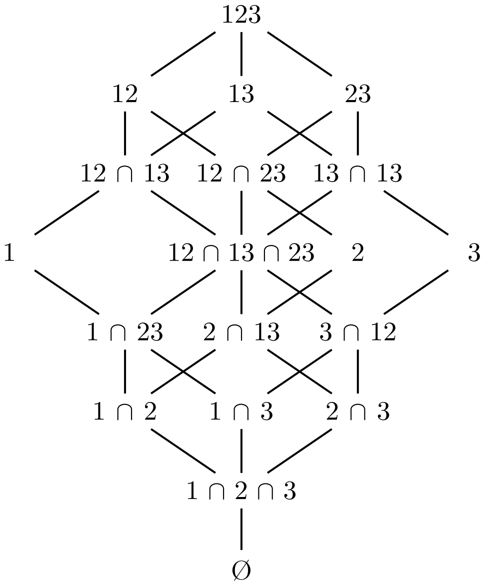
Williams and Beer axioms

1. *Symmetry*: $I_\cap(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \leq I_\cap(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

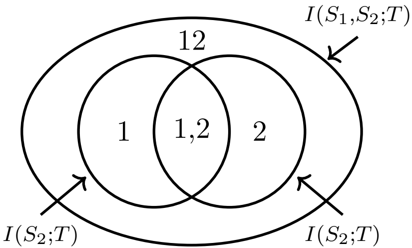
- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over the remaining combinations
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - By Axiom 3, $I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_2; T) \leq I_\cap(\boldsymbol{A}_1; T)$
  - Similarly, $I_\cap(\boldsymbol{A}_1; T) = I_\cap(\boldsymbol{A}_1, \boldsymbol{A}_3; T) \leq I_\cap(\boldsymbol{A}_3; T)$ since $\boldsymbol{A}_1 \subseteq \boldsymbol{A}_3$
- Applying $\preccurlyeq_\cap$ to all of the combinations of sources yields a lattice structure

Redundancy lattice

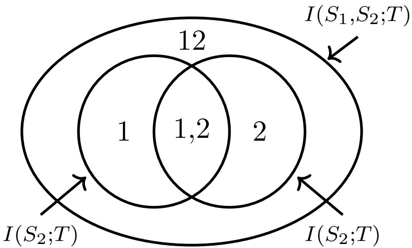


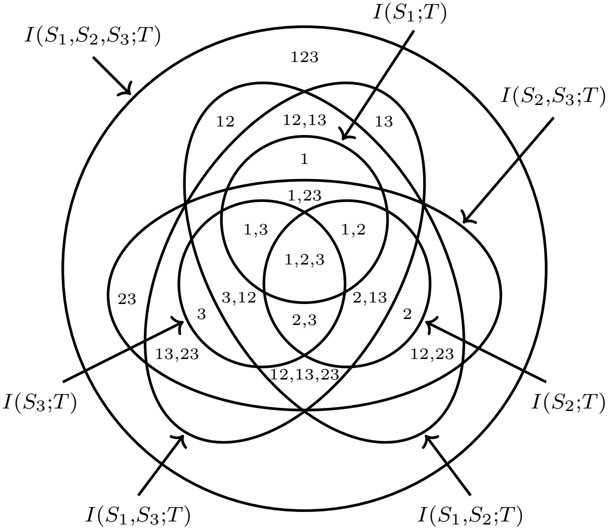


Partial information decomposition (PID)



Partial information decomposition (PID)





Proposed measures

- Still need to actually define a measure of redundant information

Proposed measures

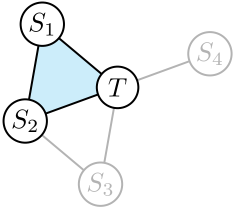
- Still need to actually define a measure of redundant information
- Providing this definition has been a contentious area of research
  - Williams and Beer (2010) $I_{\min}$.
  - Harder et al. (2013) I_{red}.
  - Bertschinger et al. (2014) $\widetilde{UI}$, or equivalently Griffith and Koch (2014) S_{VK}.
  - Barrett (2015) I_{MMI}.
  - Finn and Lizier (2018a) $r^{\pm}$.
  - ...

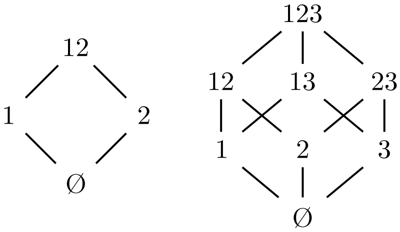
Union information decomposition

- Williams and Beer focused on defining the redundancy $I_\cap$ between sources
- Can we instead quantify the union information $I_\cup$ provided by sources?

Union information decomposition

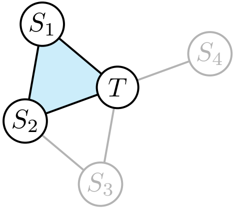
- Williams and Beer focused on defining the redundancy $I_\cap$ between sources
- Can we instead quantify the union information $I_\cup$ provided by sources?

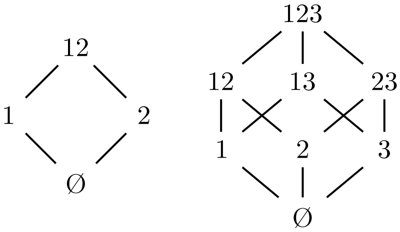




Union information decomposition

- Williams and Beer focused on defining the redundancy $I_\cap$ between sources
- Can we instead quantify the union information $I_\cup$ provided by sources?





- Let $\boldsymbol{A}_i$ represent a distinct way sources can provide information T
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$

Union information

Union information axioms

1. *Symmetry*: $I_\cup(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-information*: $I_\cup(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \geq I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

Union information

Union information axioms

1. *Symmetry*: $I_\cup(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-information*: $I_\cup(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \geq I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- $I_\cup$ can be applied to any combination of $\boldsymbol{A}_i$'s, but many are equivalent
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - But by Axiom 3, $I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_3; T) = I_\cup(\boldsymbol{A}_3; T)$ since $\boldsymbol{A}_3 \supseteq \boldsymbol{A}_1$

Union information

Union information axioms

1. *Symmetry*: $I_\cup(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-information*: $I_\cup(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \geq I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- $I_\cup$ can be applied to any combination of $\boldsymbol{A}_i$'s, but many are equivalent
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - But by Axiom 3, $I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_3; T) = I_\cup(\boldsymbol{A}_3; T)$ since $\boldsymbol{A}_3 \supseteq \boldsymbol{A}_1$
- Left with all combinations of sources $\boldsymbol{A}_i$ s.t. no source is a *superset* of any other
  - For $n = 2$, we have

$$I_\cup(\boldsymbol{A}_1; T), \quad I_\cup(\boldsymbol{A}_2; T), \quad I_\cup(\boldsymbol{A}_3; T) \quad \text{and} \quad I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2; T)$$

- The remaining combinations of sources are the same as for the redundancy

Union information

Union information axioms

1. *Symmetry*: $I_\cup(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-information*: $I_\cup(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_k; T) \geq I_\cup(\boldsymbol{A}_1; \ldots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- Axiom 3 also defines a partial order $\preccurlyeq_\cup$ over the remaining combinations
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - By Axiom 3, $I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2; T) \geq I_\cup(\boldsymbol{A}_1; T)$
  - Similarly, $I_\cup(\boldsymbol{A}_3; T) = I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3; T) \geq I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2; T)$

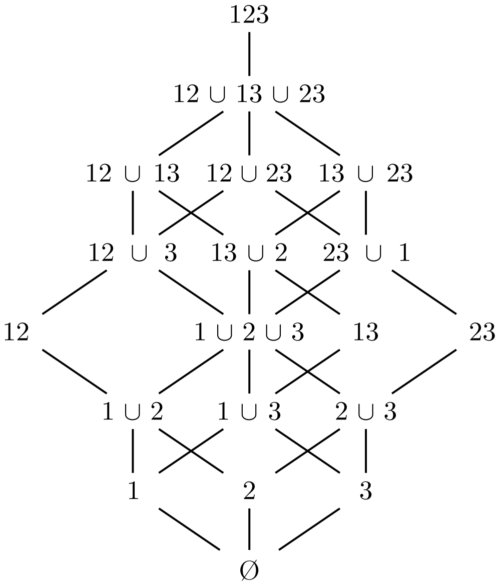
Union information

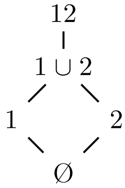
Union information axioms

1. *Symmetry*: $I_\cup(\boldsymbol{A}_1, \dots, \boldsymbol{A}_k; T)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-information*: $I_\cup(\boldsymbol{A}_i : T) = I(\boldsymbol{A}_i; T)$
3. *Monotonicity*: $I_\cup(\boldsymbol{A}_1; \dots; \boldsymbol{A}_k; T) \geq I_\cup(\boldsymbol{A}_1; \dots; \boldsymbol{A}_{k-1}; T)$ with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- Axiom 3 also defines a partial order $\preccurlyeq_\cup$ over the remaining combinations
  - For $n = 2$, we have $\boldsymbol{A}_1 = S_1$, $\boldsymbol{A}_2 = S_2$ and $\boldsymbol{A}_3 = (S_1, S_2)$
  - By Axiom 3, $I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2; T) \geq I_\cup(\boldsymbol{A}_1; T)$
  - Similarly, $I_\cup(\boldsymbol{A}_3; T) = I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3; T) \geq I_\cup(\boldsymbol{A}_1, \boldsymbol{A}_2; T)$
- We have the same combinations of sources, but a different partial order
- Applying $\preccurlyeq_\cup$ to all of the combinations of sources yields a lattice structure

Union information lattice

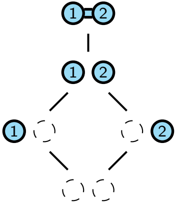


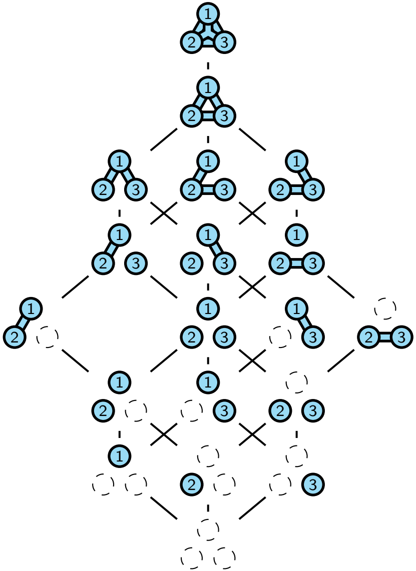


Redundancy and union information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ aligns well with higher-order interactions

Interaction hierarchy





Redundancy and union information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ aligns well with higher-order interactions
- Seems natural to demand a consistency between the approaches

$$I_\cup(S_1, S_2; T) = I(S_1; T) + I(S_2; T) - I_\cap(S_1, S_2; T)$$

- Kolchinsky (2022) argues that we should not make this demand

Inclusion-exclusion principle

- Many approaches to PID already have an implicit measure of union information

Inclusion-exclusion principle

- Many approaches to PID already have an implicit measure of union information
  - Williams and Beer (2010) $I_{\min}$, Barrett (2015) I_{MMI}, Finn and Lizier (2018a) $r^{\pm}$

$$
\begin{aligned}
I_{\cap} &= I_{\mathsf{MMI}}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = \min\big(I(\boldsymbol{A}_1; T), \ldots, I(\boldsymbol{A}_k; T)\big) \\
\implies I_{\cup} &= I_{\mathsf{MaxMI}} = \max\big(I(\boldsymbol{A}_1; T), \ldots, I(\boldsymbol{A}_k; T)\big)
\end{aligned}
$$

Inclusion-exclusion principle

- Many approaches to PID already have an implicit measure of union information
  - Williams and Beer (2010) $I_{\min}$, Barrett (2015) I_{MMI}, Finn and Lizier (2018a) $r^{\pm}$

$$
\begin{aligned}
I_\cap &= I_{\mathsf{MMI}}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_k) = \min\big(I(\boldsymbol{A}_1;T), \dots, I(\boldsymbol{A}_k;T)\big) \\
\implies I_\cup &= I_{\mathsf{MaxMI}} = \max\big(I(\boldsymbol{A}_1;T), \dots, I(\boldsymbol{A}_k;T)\big)
\end{aligned}
$$

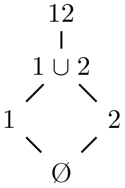
  - Bertschinger et al. (2014) $\widetilde{UI}$, Griffith and Koch (2014) S_{VK}

$$
\begin{aligned}
I_\cap &= \widetilde{SI}(S_1, S_2; T) = \max_{Q \in \Delta P} I_Q(S_1; S_2; T) \\
\implies I_\cup &= \widetilde{UnI}(S_1, S_2; T) = \min_{Q \in \Delta P} I_Q\big((S_1; S_2); T\big)
\end{aligned}
$$

Bounds on the bivariate union information

- Similar to $I_\cap$, the union information $I_\cup$ increases monotonically on the lattice
- In the bivariate case, we have that

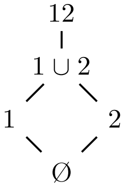
$$I\big((S_1,S_2);T\big) \geq I_\cup(S_1,S_2;T) \geq I(S_1;T), I(S_2;T) \geq 0$$



Bounds on the bivariate union information

- Similar to $I_\cap$, the union information $I_\cup$ increases monotonically on the lattice
- In the bivariate case, we have that

$$I((S_1, S_2); T) \geq I_\cup(S_1, S_2; T) \geq I(S_1; T), I(S_2; T) \geq 0$$



- Assuming that $I_\cup(S_1, S_2; T)$ depends only on $P(S_1, T)$ and $P(S_2, T)$

$$\widetilde{UnI}(S_1, S_2; T) \geq I_\cup(S_1, S_2; T) \geq I_{\mathsf{MaxMI}}$$

Conclusion

Redundant information $I_\cap$ is only one side of the information decomposition problem

We also need to consider the union information $I_\cup$

References

Adam B Barrett. Exploration of synergistic and redundant information sharing in static and dynamical gaussian systems. *Physical Review E*, 91(5):052802, 2015.

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

Conor Finn and Joseph T. Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy*, 20(4):297, 2018a.

Conor Finn and Joseph T. Lizier. Probability mass exclusions and the directed components of mutual information. *Entropy*, 20(11):826, 2018b.

Conor Finn and Joseph T Lizier. Generalised measures of multivariate information content. *Entropy*, 22(2):216, 2020.

Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception*, pages 159–190. Springer, 2014.

Malte Harder, Christoph Salge, and Daniel Polani. Bivariate measure of redundant information. *Physical Review E*, 87(1):012130, 2013.

Artemy Kolchinsky. A novel approach to the partial information decomposition. *Entropy*, 24(3):403, 2022.

Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

Announcement

DeMICS — Decomposing Multivariate Information in Complex Systems

June 5th-9th, 2023

Max Planck Institute for the Physics of Complex systems,
Dresden, Germany

Multivariate mutual information

McGill generalised the MI by defining the **multivariate mutual information**,

$$I(X;Y;Z) = I(X;Y) + I(X;Z) - I\big(X;(Y,Z)\big)$$

Multivariate mutual information

McGill generalised the MI by defining the **multivariate mutual information**,

$$I(X;Y;Z) = I(X;Y) + I(X;Z) - I\big(X;(Y,Z)\big)$$

- MMI is not non-negative.
- Historically: "no intuitive meaning".

Multivariate mutual information

McGill generalised the MI by defining the **multivariate mutual information**,

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

- MMI is not non-negative.
- Historically: “no intuitive meaning”.