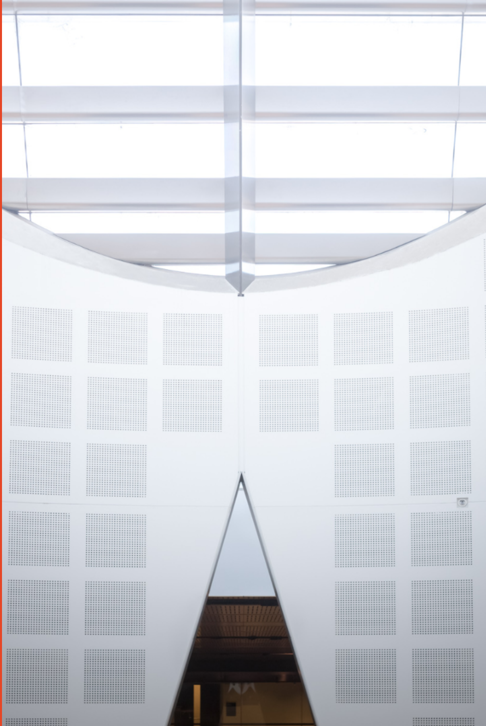
Local interpretations of PID

PID workshop, FIAS

Conor Finn

December 15, 2016

THE UNIVERSITY OF SYDNEY



What do we mean by local information?

- The entropy and mutual information are the expected values over the local or pointwise values:

$$H(X) = \langle h(x) \rangle, \qquad \text{where,} \qquad h(x) = -\log p(x);$$

$$I(X;Y) = \langle i(x;y) \rangle, \qquad \text{where,} \qquad i(x;y) = \log \frac{p(x|y)}{p(x)}.$$

What do we mean by local information?

- The entropy and mutual information are the expected values over the local or pointwise values:

$$H(X) = \langle h(x) \rangle, \qquad \text{where,} \qquad h(x) = -\log p(x);$$

$$I(X;Y) = \langle i(x;y) \rangle, \qquad \text{where,} \qquad i(x;y) = \log \frac{p(x|y)}{p(x)}.$$

- Local mutual information can be negative! 

Why a local interpretation of information?

- Why not?

Why a local interpretation of information?

- Why not?
- Fano — local mutual information can be derived as the primary citizen from four postulates:
  1. once differentiability;
  2. same form for conditionals;
  3. addativity; and
  4. separation for independent ensembles.

Why a local interpretation of information?

- Why not?
- Fano — local mutual information can be derived as the primary citizen from four postulates:
  1. once differentiability;
  2. same form for conditionals;
  3. addativity; and
  4. separation for independent ensembles.
- If X and Y are time series, local values measure dynamics over time which would be useful in applications of PID.

What do we mean by local PID?

- Formally, there should be some kind of axiom:

What do we mean by local PID?

- Formally, there should be some kind of axiom:

Axiom: *Localizability*

There exists a local measure $i_\cap(t; \mathbf{s}_1, \ldots, \mathbf{s}_k)$ for the redundancy of a specific observation $\{t, \mathbf{s}_1, \ldots, \mathbf{s}_k\}$ of $\{T, \mathbf{S}_1, \ldots, \mathbf{S}_k\}$, such that:

1. $i_\cap(t; \mathbf{s}_1, \ldots, \mathbf{s}_k)$ satisfies the corresponding symmetry and self-redundancy axioms as per $I_\cap(T; \mathbf{S}_1, \ldots, \mathbf{S}_k)$;
2. $I_\cap(T; \mathbf{S}_1, \ldots, \mathbf{S}_k) = \langle i_\cap(t; \mathbf{s}_1, \ldots, \mathbf{s}_k) \rangle$;
3. $i_\cap(t; \mathbf{s}_1, \ldots, \mathbf{s}_k)$ is *once-differentiable* with respect to changes in $p(t, \mathbf{s}_1, \ldots, \mathbf{s}_k)$; and
4. $i_\cap(t; \mathbf{s}_1, \ldots, \mathbf{s}_k)$ is *uniquely defined* for the given candidate redundancy measure.

What do we mean by local PID?

- In practice, define a PI type locally, solve for the others and calculate the expectation value for each type.

What do we mean by local PID?

- In practice, define a PI type locally, solve for the others and calculate the expectation value for each type.

Global PID for ensemble

$$\begin{aligned} I(T;S_1) &= U_1 + R \\ I(T;S_2) &= U_2 + R \\ I(T;S_1,S_2) &= U_1 + U_2 + R + S \end{aligned}$$

What do we mean by local PID?

- In practice, define a PI type locally, solve for the others and calculate the expectation value for each type.

Global PID for ensemble

$$I(T;S_1) = U_1 + R$$
$$I(T;S_2) = U_2 + R$$
$$I(T;S_1,S_2) = U_1 + U_2 + R + S$$

Local PID for each configuration

$$i(t;s_1) = u_1^i + r^i$$
$$i(t;s_2) = u_2^i + r^i$$
$$i(t;s_1,s_2) = u_1^i + u_2^i + r^i + s^i$$

What do we mean by local PID?

- In practice, define a PI type locally, solve for the others and calculate the expectation value for each type.

Global PID for ensemble

$$I(T;S_1) = U_1 + R$$
$$I(T;S_2) = U_2 + R$$
$$I(T;S_1,S_2) = U_1 + U_2 + R + S$$

Local PID for each configuration

$$i(t;s_1) = u_1^i + r^i$$
$$i(t;s_2) = u_2^i + r^i$$
$$i(t;s_1,s_2) = u_1^i + u_2^i + r^i + s^i$$

$$U_1 = \langle u_1^i \rangle \qquad U_2 = \langle u_2^i \rangle \qquad R = \langle r^i \rangle \qquad S = \langle s^i \rangle$$

What do we mean by local PID?

- In practice, define a PI type locally, solve for the others and calculate the expectation value for each type.

Global PID for ensemble

$$I(T;S_1) = U_1 + R$$
$$I(T;S_2) = U_2 + R$$
$$I(T;S_1,S_2) = U_1 + U_2 + R + S$$

Local PID for each configuration

$$i(t;s_1) = u_1^i + r^i$$
$$i(t;s_2) = u_2^i + r^i$$
$$i(t;s_1,s_2) = u_1^i + u_2^i + r^i + s^i$$

$$U_1 = \langle u_1^i \rangle \qquad U_2 = \langle u_2^i \rangle \qquad R = \langle r^i \rangle \qquad S = \langle s^i \rangle$$

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
$1/4$	0	0	0	$\log 4/3$	$\log 4/3$	$\log 4/3$	u_1^1	u_2^1	r^1	s^1
$1/4$	0	0	1	$\log 4/3$	$\log 2/3$	$\log 4/3$	u_1^2	u_2^2	r^2	s^2
$1/4$	0	1	0	$\log 2/3$	$\log 4/3$	$\log 4/3$	u_1^3	u_2^3	r^3	s^3
$1/4$	1	1	1	1	1	2	u_1^4	u_2^4	r^4	s^4
Expected values				0.311	0.31	0.811	U_1	U_2	R	S

Advantages of the local approach to PID

"*However, [I_{min}] ignore[s] that even though X and Y give the same amount of information about an outcome z, they tell something different about the change of the distribution $p(z)$ after an observation in X or Y has been made.*"

— Harder, Salge, and Polani 2013

Advantages of the local approach to PID

"*However, [I_{min}] ignore[s] that even though X and Y give the same amount of information about an outcome z, they tell something different about the change of the distribution $p(z)$ after an observation in X or Y has been made.*"

— Harder, Salge, and Polani 2013

"*The problem is that I min does not distinguish whether different random variables carry the same information or just the same amount of information. This phenomenon has also been observed by others.*"

— Bertschinger, Rauh, Olbrich, and Jost 2013

Advantages of the local approach to PID

"*However, [I_{min}] ignore[s] that even though X and Y give the same amount of information about an outcome z, they tell something different about the change of the distribution $p(z)$ after an observation in X or Y has been made.*"

— Harder, Salge, and Polani 2013

"*The problem is that I min does not distinguish whether different random variables carry the same information or just the same amount of information. This phenomenon has also been observed by others.*"

— Bertschinger, Rauh, Olbrich, and Jost 2013

"*Altogether, we conclude that S_{max} overestimates the intuitive synergy by miscategorizing merely unique information as synergistic whenever two or more predictors have unique information about the target.*"

— Griffith and Koch 2014

Advantages of the local approach to PID

"*Even though* S_1 *and* S_2 *give the same amount of information about an outcome* t*, they tell you something different.*"

Advantages of the local approach to PID

"*Even though S_1 and S_2 give the same amount of information about an outcome t, they tell you something different.*"

- Reduced OR (`ROR`):

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$
$1/2$	0	0	0	0.415	0.415	1
$1/4$	1	0	1	-0.585	1	1
$1/4$	1	1	0	1	-0.585	1
Expected values				0.311	0.311	1

Advantages of the local approach to PID

"*Even though S_1 and S_2 give the same amount of information about an outcome t, they tell you something different.*"

- Reduced OR (`ROR`):

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$
$1/2$	0	0	0	0.415	0.415	1
$1/4$	1	0	1	-0.585	1	1
$1/4$	1	1	0	1	-0.585	1
Expected values				0.311	0.311	1

- On average, S_1 and S_2 each give you the same amount of information (0.311 bits) yet locally they appear to tell you something different.

Advantages of the local approach to PID

"*Even though S_1 and S_2 give the same amount of information about an outcome t, they tell you something different.*"

- Reduced OR (ROR):

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$
$1/2$	0	0	0	0.415	0.415	1
$1/4$	1	0	1	-0.585	1	1
$1/4$	1	1	0	1	-0.585	1
Expected values				0.311	0.311	1

- On average, S_1 and S_2 each give you the same amount of information (0.311 bits) yet locally they appear to tell you something different.

- Unique information may depend on your perspective.

Existing measures and localisation: I_{red}

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

Existing measures and localisation: I_{red}

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

- Harder, Salge, and Polani 2013

$$I_{\text{red}}(T : S_1; S_2) = \min\left\{I_T^{\pi}(S_1 \searrow S_2), I_T^{\pi}(S_2 \searrow S_1)\right\},$$

where

$$I_T^{\pi}(S_1 \searrow S_2) = \sum_{t, s_1} p(t, s_1) \log \frac{p_{(s_1 \searrow S_2)}(t)}{p(t)}$$

Existing measures and localisation: I_{red}

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

- Harder, Salge, and Polani 2013
$$I_{\mathsf{red}}(T : S_1; S_2) = \min\left\{ I_T^{\pi}(S_1 \searrow S_2), I_T^{\pi}(S_2 \searrow S_1) \right\},$$
where
$$I_T^{\pi}(S_1 \searrow S_2) = \sum_{t,\, s_1} p(t, s_1) \log \frac{p_{(s_1 \searrow S_2)}(t)}{p(t)}$$
- Defining,
$$r(t : s_1; s_2) = \log \frac{p_{(s_1 \searrow S_2)}(t)}{p(t)},$$
does not work when $I_T^{\pi}(S_1 \searrow S_2) = I_T^{\pi}(S_2 \searrow S_1)$ because
$$\log \frac{p_{(s_1 \searrow S_2)}(t)}{p(t)} = \log \frac{p_{(s_2 \searrow S_1)}(t)}{p(t)},$$
is not generally true—i.e. the choice of $r(t : s_1; s_2)$ is not unique.

Existing measures and localisation: $\widetilde{UI}(T{:}S_1 \backslash S_2)$

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

Existing measures and localisation: $\widetilde{UI}(T{:}S_1 \backslash S_2)$

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

- Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

$$\widetilde{UI}(T : S_1 \setminus S_2) = \min_{Q \in \Delta_P} MI_Q(T : S_1 | S_2)$$

Existing measures and localisation: $\widetilde{UI}(T{:}S_1\backslash S_2)$

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

- Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

$$\widetilde{UI}(T : S_1 \setminus S_2) = \min_{Q \in \Delta_P} MI_Q(T : S_1 | S_2)$$

- Defining

$$u_1(t : s_1 \setminus s_2) = mi_Q(t : s_1 | s_2)$$

  does not work as Q does not necessarily have the same support as P.

Existing measures and localisation: $\widetilde{UI}(T{:}S_1 \backslash S_2)$

Defining a PID for the average joint mutual information does not yield unique local PID for each of the configurations.

- Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

$$\widetilde{UI}(T : S_1 \setminus S_2) = \min_{Q \in \Delta_P} MI_Q(T : S_1 | S_2)$$

- Defining

$$u_1(t : s_1 \setminus s_2) = mi_Q(t : s_1 | s_2)$$

  does not work as Q does not necessarily have the same support as P. For example, ROR

p	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	1
$1/4$	1	1	0

$$\xrightarrow[\text{to } Q]{P \text{ minimises}}$$

q	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	0
$1/4$	1	1	1

More on ROR

The unique information in ROR depends on the chosen perspective

More on ROR

The unique information in ROR depends on the chosen perspective

1. "*Shared information and unique information should depend only on the marginal distributions of the pairs* (X, Y) *and* (X, Z)*.*"

— Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

More on ROR

The unique information in ROR depends on the chosen perspective

1. "*Shared information and unique information should depend only on the marginal distributions of the pairs* (X, Y) *and* (X, Z)*.*"

— Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

p	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	1
$1/4$	1	1	0

$\xrightarrow[\text{to } Q]{P \text{ minimises}}$

q	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	0
$1/4$	1	1	1

More on ROR

The unique information in ROR depends on the chosen perspective

1. "*Shared information and unique information should depend only on the marginal distributions of the pairs* (X, Y) *and* (X, Z)*.*"

— Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

p	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	1
$1/4$	1	1	0

$\xrightarrow[\text{to } Q]{P \text{ minimises}}$

q	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	0
$1/4$	1	1	1

2. Say an observer viewing S_1 knows the full joint distribution. If they observe $S_1 = 1$, then they uniquely know the value of the target $T = 1$ in that local case.

More on RoR

The unique information in RoR depends on the chosen perspective

1. *"Shared information and unique information should depend only on the marginal distributions of the pairs (X, Y) and (X, Z)."*

— Bertschinger, Rauh, Olbrich, Jost, and Ay 2014

p	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	1
$1/4$	1	1	0

$\xrightarrow[\text{to } Q]{P \text{ minimises}}$

q	t	s_1	s_2
$1/2$	0	0	0
$1/4$	1	0	0
$1/4$	1	1	1

2. Say an observer viewing S_1 knows the full joint distribution. If they observe $S_1 = 1$, then they uniquely know the value of the target $T = 1$ in that local case.

- The first perspective is common to all of I_{min}, I_{red} and $\widetilde{UI}$. They say there is no unique information in RoR seems to be incompatible with the second perspective.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1				
0.1	1	1	0	1	$\log(1/3)$	1				
0.4	1	1	1	1	1	1				
Expected values				1	0.610	1				

Assumptions:

Another interesting local example

ImperfectRdn — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1				
0.1	1	1	0	1	$\log(1/3)$	1				
0.4	1	1	1	1	1	1				
Expected values				1	0.610	1		0		

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1				
0.1	1	1	0	1	$\log(1/3)$	1				
0.4	1	1	1	1	1	1				
Expected values				1	0.610	1	0.390	0	0.610	0

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1		0		
0.1	1	1	0	1	$\log(1/3)$	1				
0.4	1	1	1	1	1	1		0		
Expected values				1	0.610	1	0.390	0	0.610	0

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1	$\log(6/5)$	0	$\log(5/3)$	0
0.1	1	1	0	1	$\log(1/3)$	1				
0.4	1	1	1	1	1	1	0	0	1	0
Expected values				1	0.610	1	0.390	0	0.610	0

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1	$\log(6/5)$	0	$\log(5/3)$	0
0.1	1	1	0	1	$\log(1/3)$	1			$\log(1/3)$	
0.4	1	1	1	1	1	1	0	0	1	0
Expected values				1	0.610	1	0.390	0	0.610	0

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

3a. There is only redundant information in S_2 and it has to be the expected value over the local redundancies for s_2, hence $r = \log(1/3)$ in the second row.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1	$\log(6/5)$	0	$\log(5/3)$	0
0.1	1	1	0	1	$\log(1/3)$	1			$\log(1/3)$	
0.4	1	1	1	1	1	1	0	0	1	0
Expected values				1	0.610	1	0.390	0	0.610	0

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

3a. There is only redundant information in S_2 and it has to be the expected value over the local redundancies for s_2, hence $r = \log(1/3)$ in the second row.
   – Strange! What is misinformative about S_1 in the second row?

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1	$\log(6/5)$	0	$\log(5/3)$	0
0.1	1	1	0	1	$\log(1/3)$	1		$\log(1/3)$		
0.4	1	1	1	1	1	1	0	0	1	0
Expected values				1	0.610	1	0.390	$-$	0.610	$+$

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

3a. There is only redundant information in S_2 and it has to be the expected value over the local redundancies for s_2, hence $r = \log(1/3)$ in the second row.
   - Strange! What is misinformative about S_1 in the second row?

3b. Maybe this misinformation is unique to u_2 instead? But then $U_1 < 0$!
   - That might make sense with the local interpretation—the only thing unique about S_1 is when it is misinformative.

Another interesting local example

IMPERFECTRDN — Griffith et al. 2014; Ince 2016

p	t	s_1	s_2	$i(t;s_1)$	$i(t;s_2)$	$i(t;s_1,s_2)$	u_1	u_2	r	s
0.5	0	0	0	1	$\log(5/3)$	1	$\log(6/5)$	0	$\log(5/3)$	0
0.1	1	1	0	1	$\log(1/3)$	1	1	$\log(1/3)$	0	$+$
0.4	1	1	1	1	1	1	0	0	1	0
Expected values				1	0.610	1	0.390	$-$	0.610	$+$

Assumptions:

1. T is an exact copy of S_1 (i.e. fully informative) so $U_2 = 0$.
2. The same seems to be the case for the first and third rows as $i(t;s_2) = i(t;s_1)$, hence $u_2 = 0$ in those rows.

3a. There is only redundant information in S_2 and it has to be the expected value over the local redundancies for s_2, hence $r = \log(1/3)$ in the second row.
   - Strange! What is misinformative about S_1 in the second row?

3b. Maybe this misinformation is unique to u_2 instead? But then $U_1 < 0$!
   - That might make sense with the local interpretation—the only thing unique about S_1 is when it is misinformative.
   - But even this is troublesome as this introduces synergy!!

Final aside: semantics and identity axiom

The 2 bit copy problem

p	s_1	s_2	t
$1/4$	00	0	0
$1/4$	01	0	1
$1/4$	10	1	0
$1/4$	11	1	1

Final aside: semantics and identity axiom

The 2 bit copy problem

p	s_1	s_2	t
$1/4$	00	0	0
$1/4$	01	0	1
$1/4$	10	1	0
$1/4$	11	1	1

- The identity axiom is an axiom about a semantic situation

$$R\big(T = (S_1, S_2) : S_1; S_2\big) = MI(S_1, S_2).$$

Final aside: semantics and identity axiom

The 2 bit copy problem

p	s_1	s_2	t	t'
$1/4$	00	0	0	a
$1/4$	01	0	1	b
$1/4$	10	1	0	c
$1/4$	11	1	1	d

- The identity axiom is an axiom about a semantic situation

$$R\big(T = (S_1, S_2) : S_1; S_2\big) = MI(S_1, S_2).$$

- This is potentially problematic as the same probability distribution may describe two different semantic set-up.

Final aside: semantics and identity axiom

The 2 bit copy problem

p	s_1	s_2	t	t'	t''
$1/4$	00	0	0	a	00
$1/4$	01	0	1	b	11
$1/4$	10	1	0	c	10
$1/4$	11	1	1	d	01

- The identity axiom is an axiom about a semantic situation

$$R\big(T = (S_1, S_2) : S_1; S_2\big) = MI(S_1, S_2).$$

- This is potentially problematic as the same probability distribution may describe two different semantic set-up.

References

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, and Jürgen Jost. Shared information—new insights and problems in decomposing information in complex systems. In *Proceedings of the European Conference on Complex Systems 2012*, pages 251–269. Springer, 2013.

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

Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In Mikhail Prokopenko, editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 159–190. 2014.

Virgil Griffith, Edwin KP Chong, Ryan G James, Christopher J Ellison, and James P Crutchfield. Intersection information based on common randomness. *Entropy*, 16(4):1985–2000, 2014.

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

Robin AA Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *arXiv preprint arXiv:1602.05063*, 2016.