Caltech

Causal Macro Variables

Frederick Eberhardt

(joint work with Krzysztof Chalupka and Pietro Perona)

Causal Discovery

truth
(unknown)

$$x \rightarrow y \rightarrow z$$

samples

x	y	z
1	1	1
0	1	1
1	0	0
0	0	0
...	...	...

Causal Discovery

truth
(unknown)

$x \rightarrow y \rightarrow z$

samples

x	y	z
1	1	1
0	1	1
1	0	0
0	0	0
...	...	...

inference-algorithm

Causal Discovery

truth

(unknown)

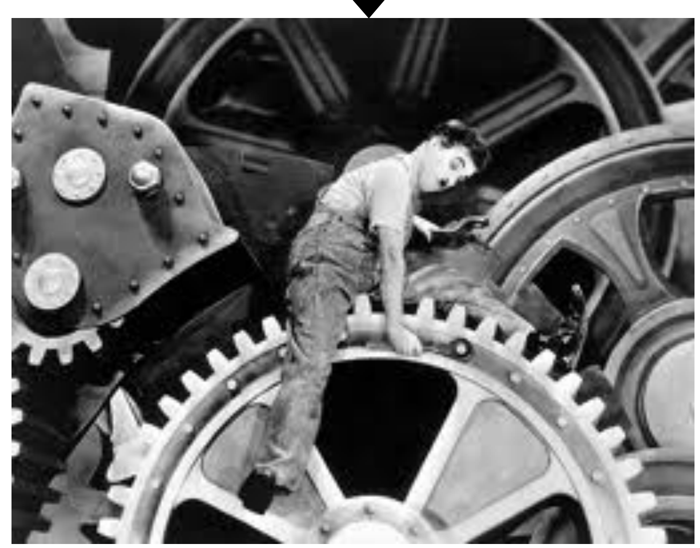
$x \to y \to z$

samples

x	y	z
1	1	1
0	1	1
1	0	0
0	0	0
...	...	...

assumptions, e.g.

- acyclicity
- no latent variables
- parametric form
- time order
- etc.



inference-algorithm

Causal Discovery

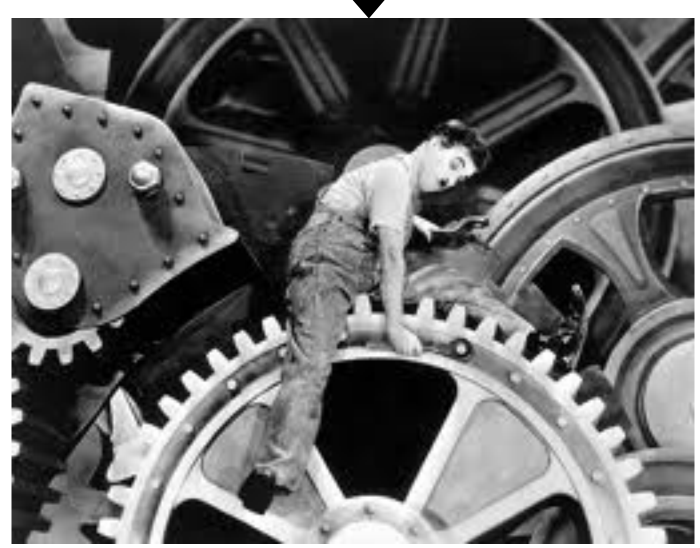
truth
(unknown)

$x \rightarrow y \rightarrow z$

samples

x	y	z
1	1	1
0	1	1
1	0	0
0	0	0
...	...	...

assumptions, e.g.
- acyclicity
- no latent variables
- parametric form
- time order
- etc.



inference-algorithm

equivalence classes of causal structures

$x \rightarrow y \rightarrow z$

$x \leftarrow y \leftarrow z$

$x \leftarrow y \rightarrow z$

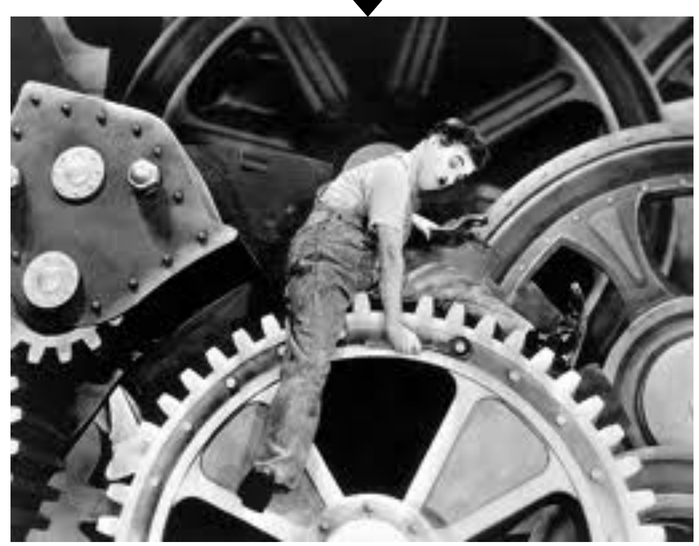
Causal Discovery

truth
(unknown)

$x \rightarrow y \rightarrow z$

data

assumptions, e.g.
- acyclicity
- no latent variables
- parametric form
- time order
- etc.



inference-algorithm

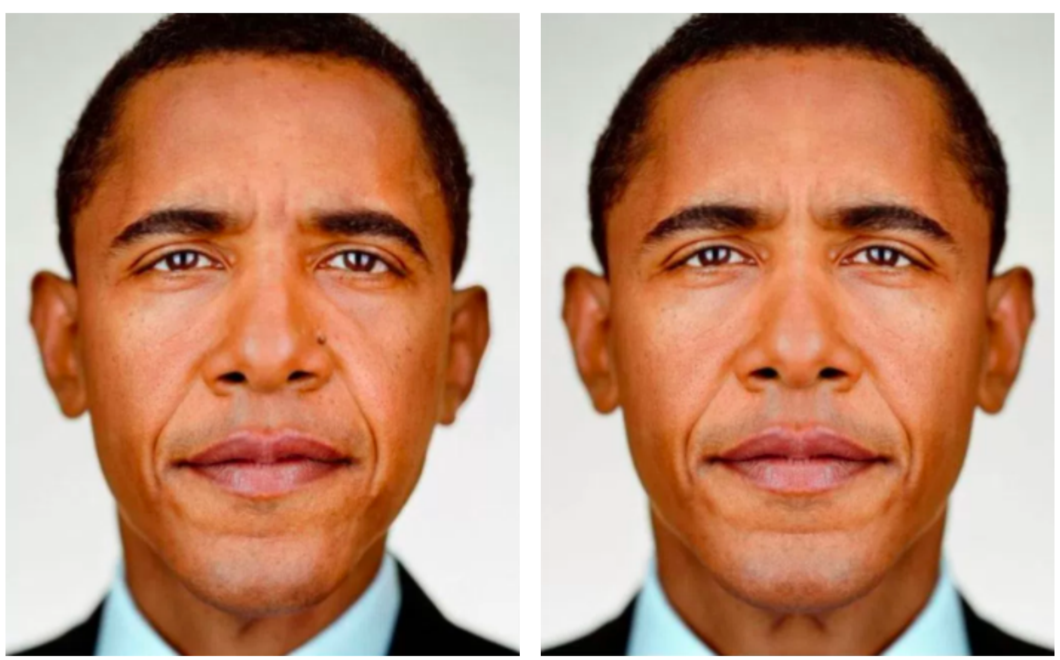
equivalence classes
of causal structures

$x \rightarrow y \rightarrow z$

$x \leftarrow y \leftarrow z$

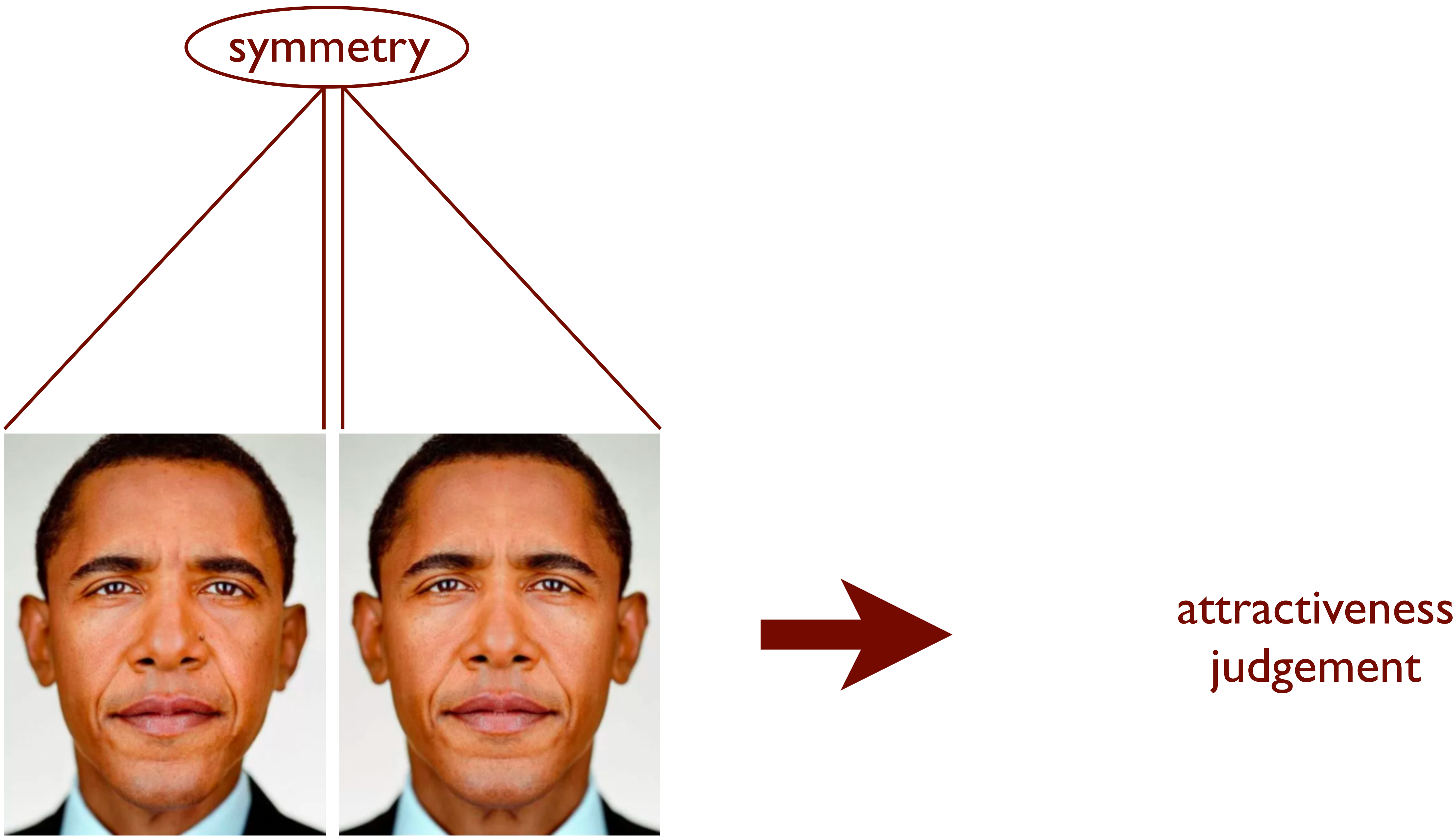
$x \leftarrow y \rightarrow z$

Causal Macro Variables

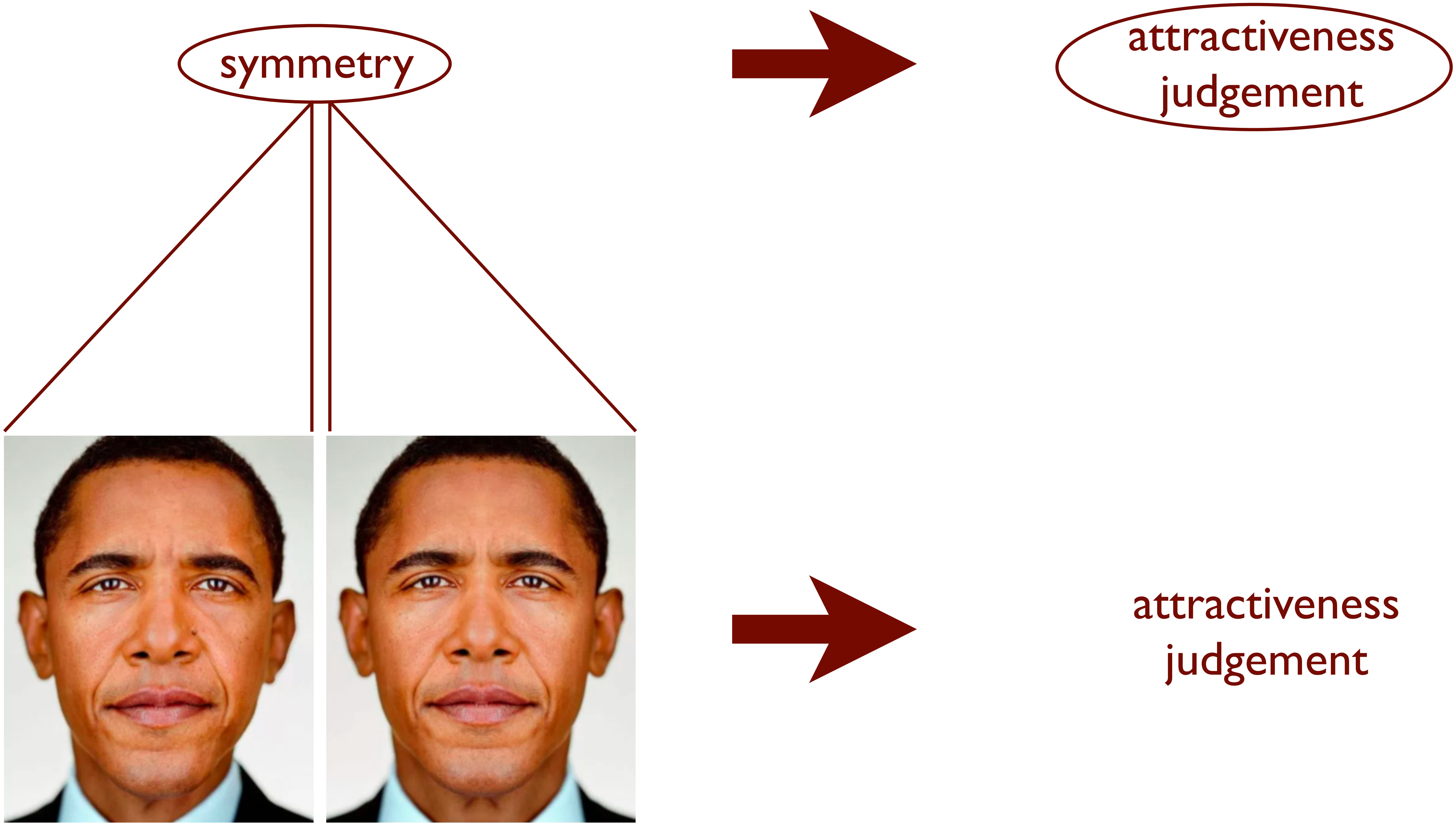


attractiveness judgement

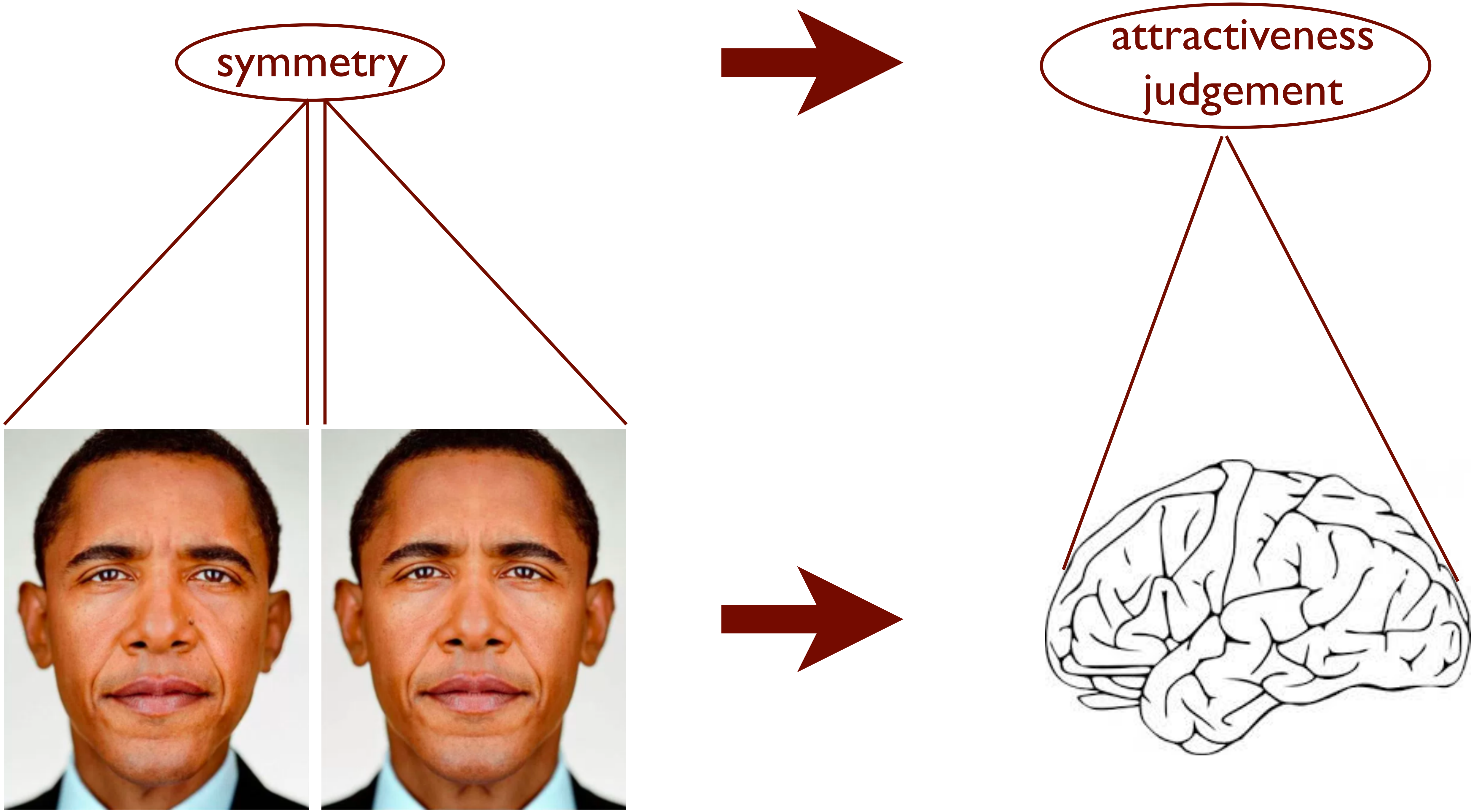
Causal Macro Variables



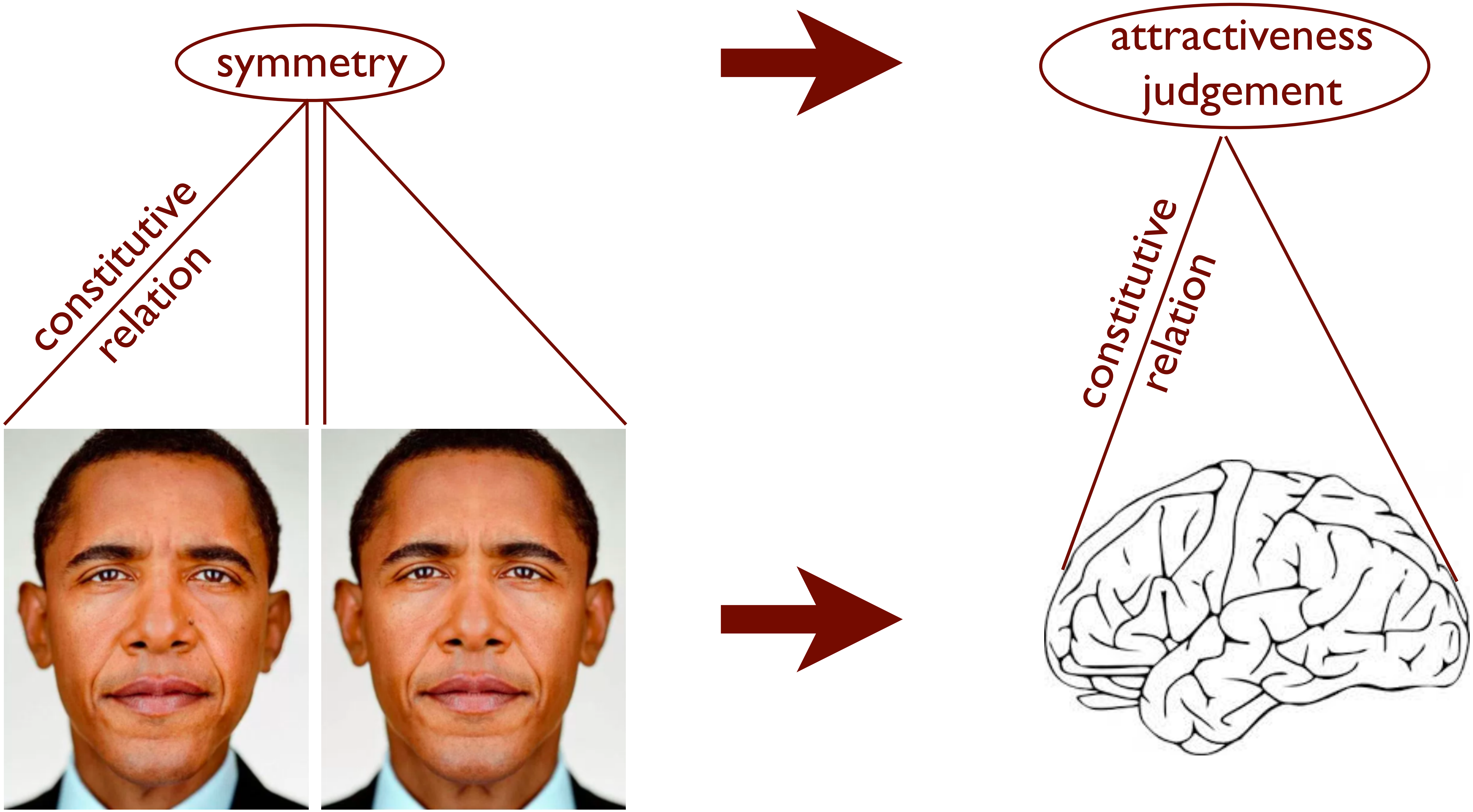
Causal Macro Variables



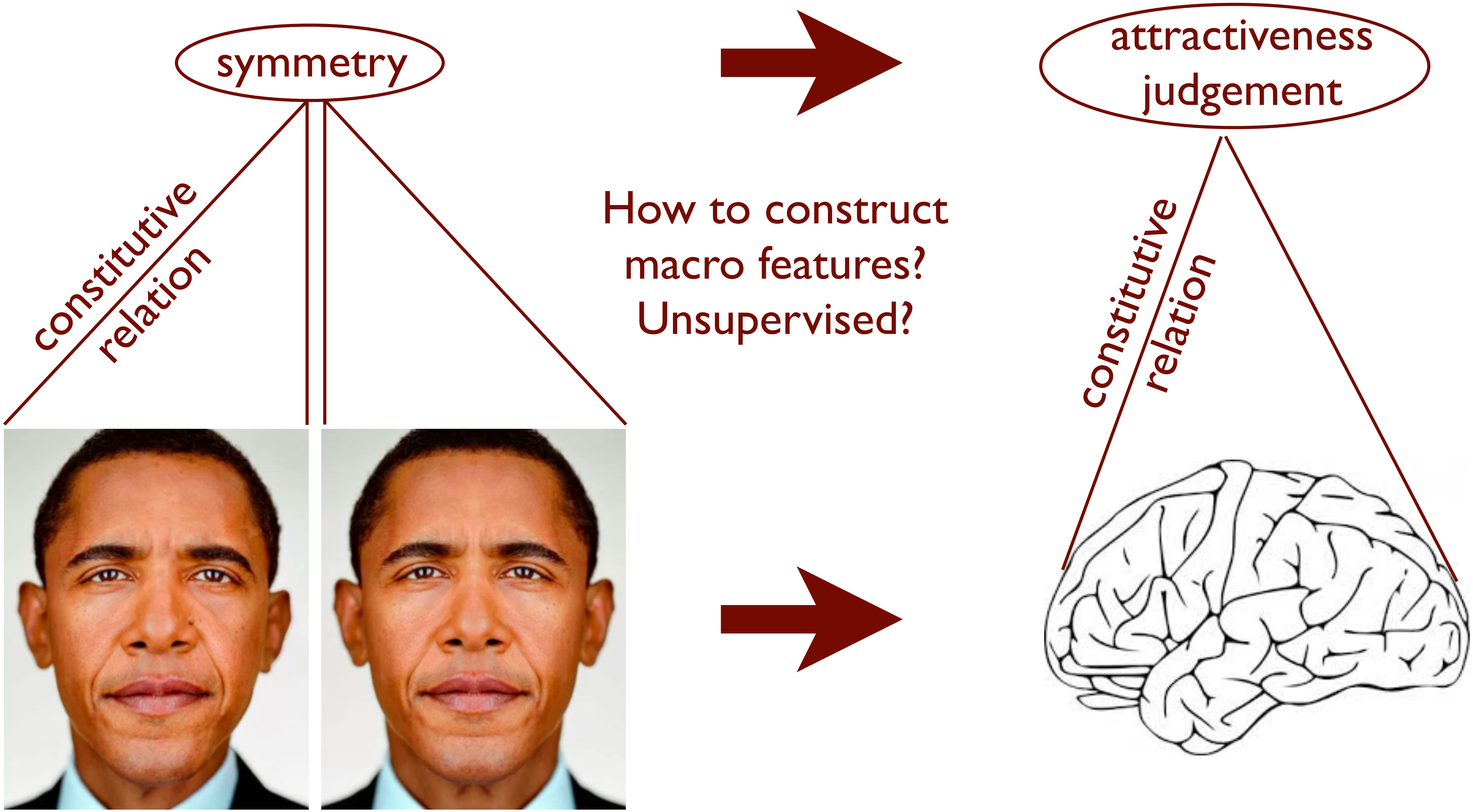
Causal Macro Variables



Causal Macro Variables



Causal Macro Variables



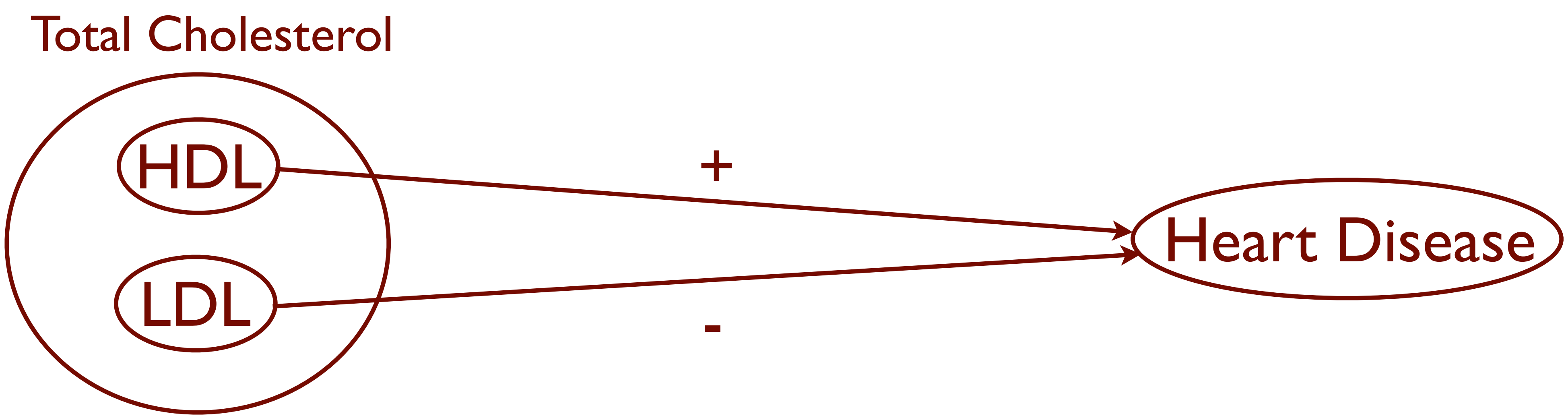
The Aim

- ➡ **account of the construction of causal variables**
- ➡ **applicable to complex macro-level causes**
- ➡ **domain general**
- ➡ **supports an interpretation of causation as invariance under intervention**

Ambiguous Manipulation

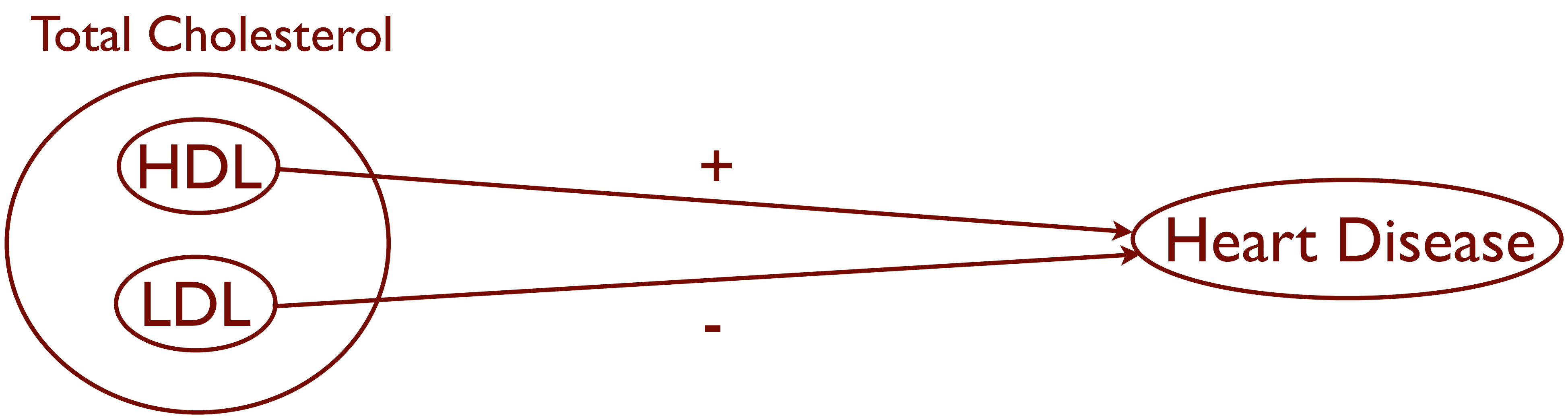
Spirtes & Scheines (2002)

Ambiguous Manipulation



Spirtes & Scheines (2002)

Ambiguous Manipulation

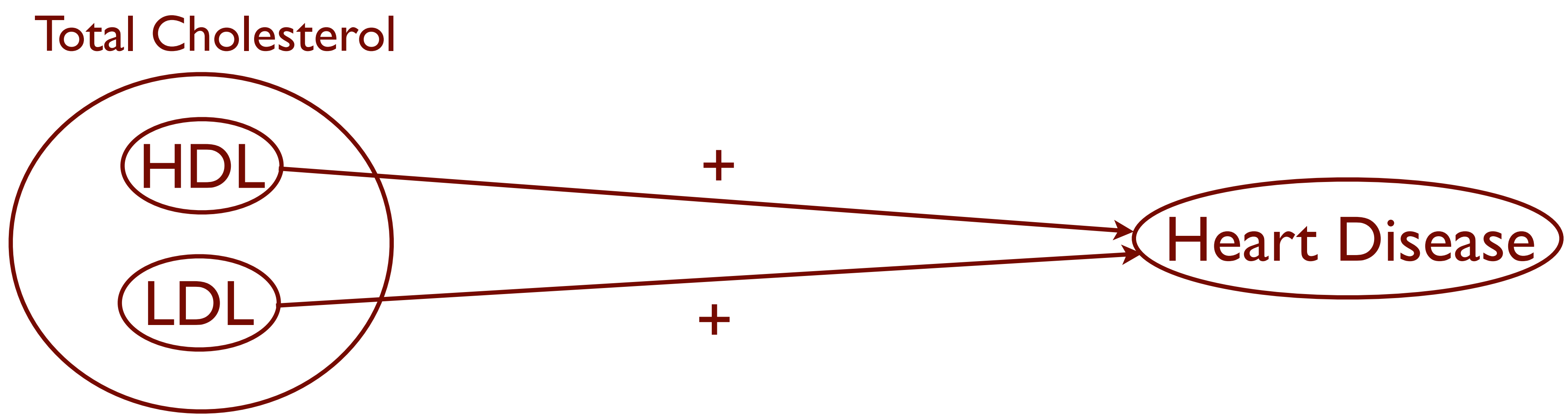


- the causal effect of *Total Cholesterol* on *Heart Disease* is **ambiguous**

➡ *Total Cholesterol* is over-aggregated, it cannot be described as a cause of *Heart Disease*

Spirtes & Scheines (2002)

Ambiguous Manipulation



- if *HDL* and *LDL* have the **same** causal effect on *Heart Disease* then the causal effect of *Total Cholesterol* on *Heart Disease* is NOT ambiguous

➡ we can aggregate *HDL* and *LDL* into *Total Cholesterol,* which is a cause of *Heart Disease*

Spirtes & Scheines (2002)

Ambiguous Manipulation

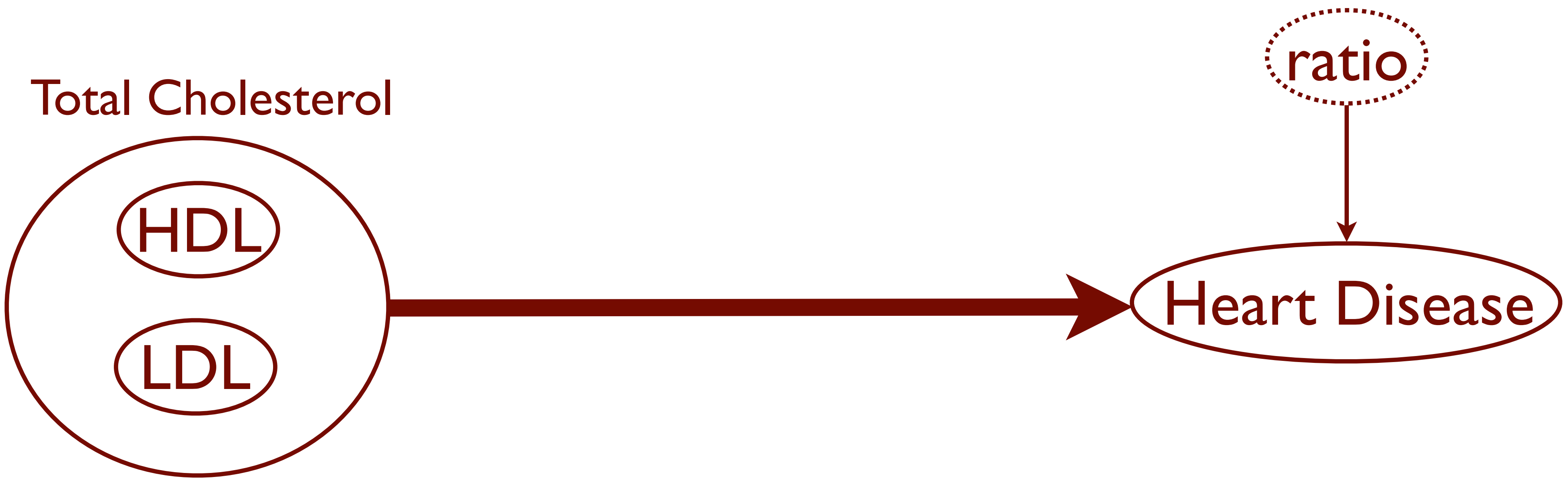
- if *HDL* and *LDL* have the **same** causal effect on *Heart Disease* then the causal effect of *Total Cholesterol* on *Heart Disease* is NOT ambiguous

➡ we can aggregate *HDL* and *LDL* into *Total Cholesterol,* which is a cause of *Heart Disease*

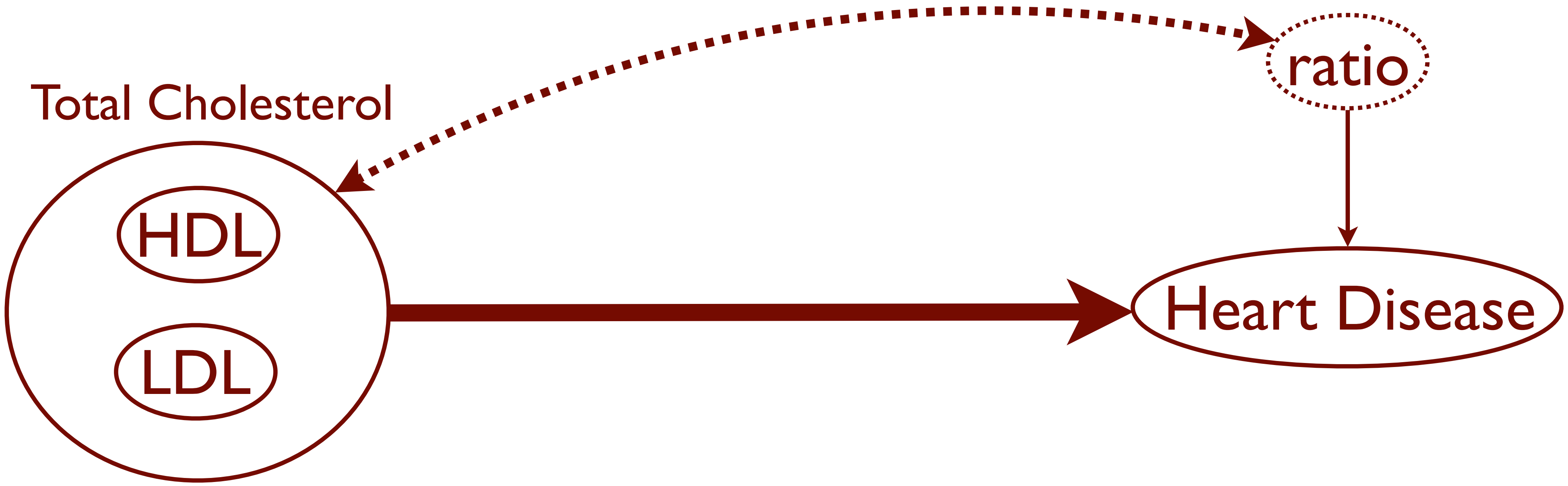
Spirtes & Scheines (2002)

Ambiguous Manipulation

Ambiguous Manipulation



Ambiguous Manipulation



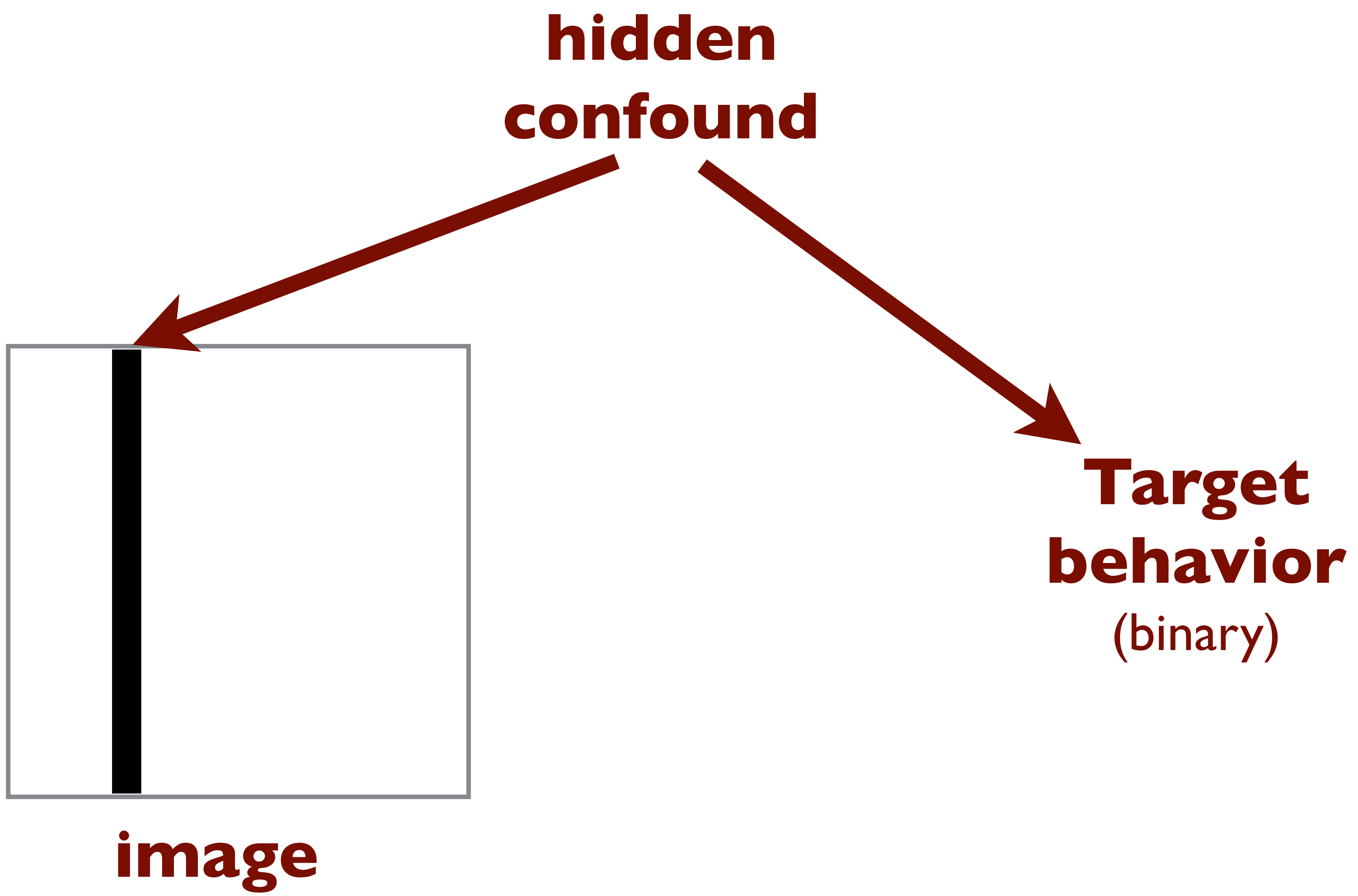
- arbitrary choices of variables imply correlated errors
- interventions would be interventions on the variable *and* the error term

Constructing / Identifying Macro Variable

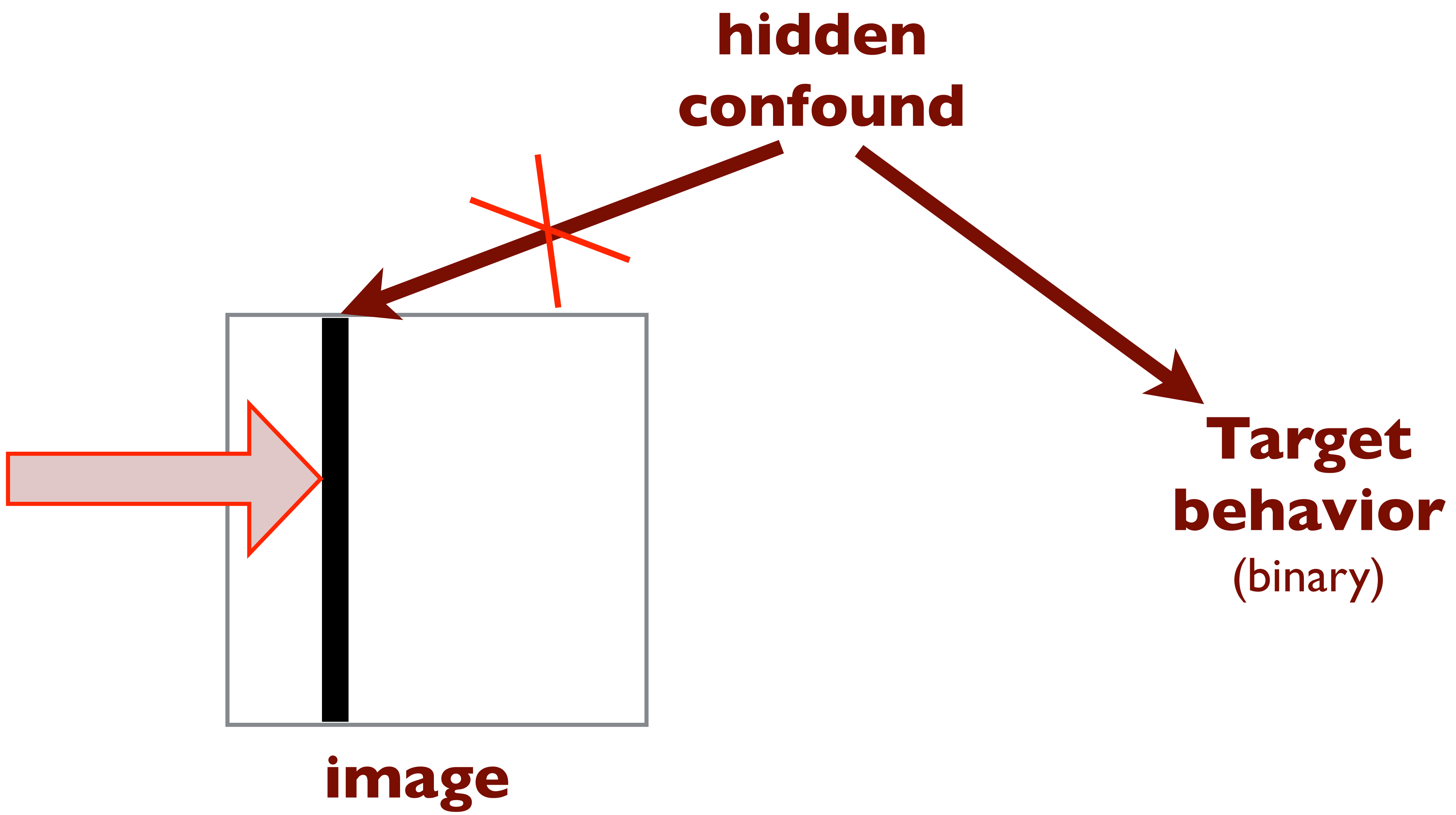
- ➡ **account of the construction of causal variables**
- ➡ **applicable to complex macro-level causes**
- ➡ **domain general**
- ➡ **supports an interpretation of causation as invariance under intervention**

- ➡ **merge states that have the same causal effect**
- ➡ **do not merge if an ambiguous manipulation would result**

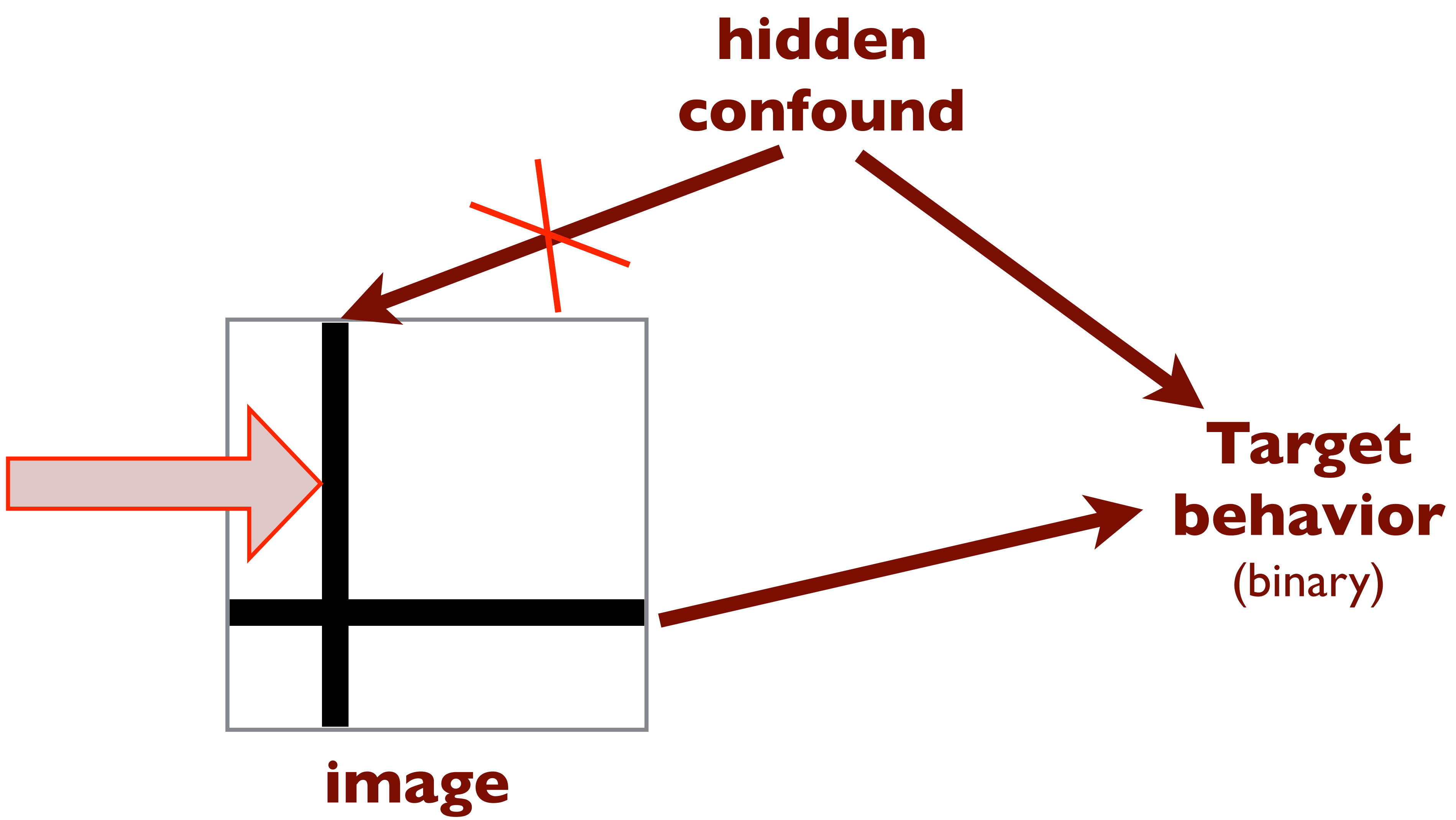
Toy example (discrete spaces)



Toy example (discrete spaces)

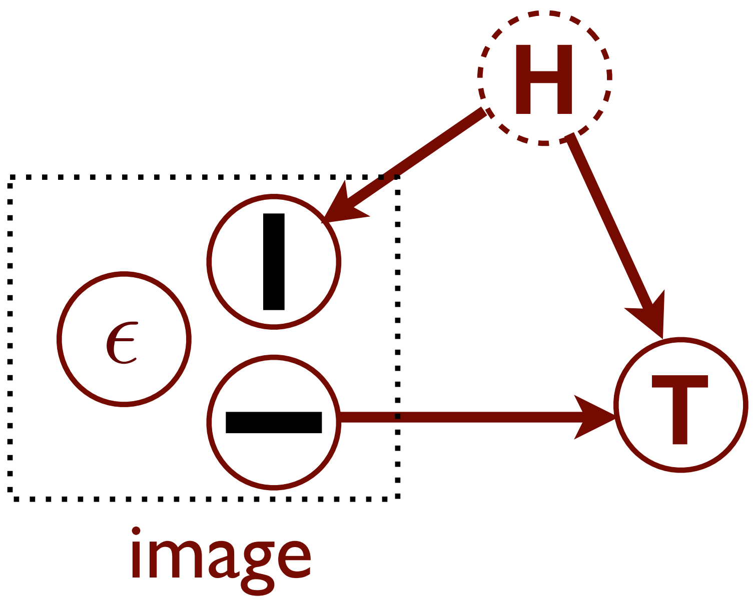


Toy example (discrete spaces)



The horizontal, but not the vertical bar, is causal of the target behavior, even though both are predictive of it.

True Macro-Causal Model



P(T=0 |) = 1

P(T=0 |) = .66

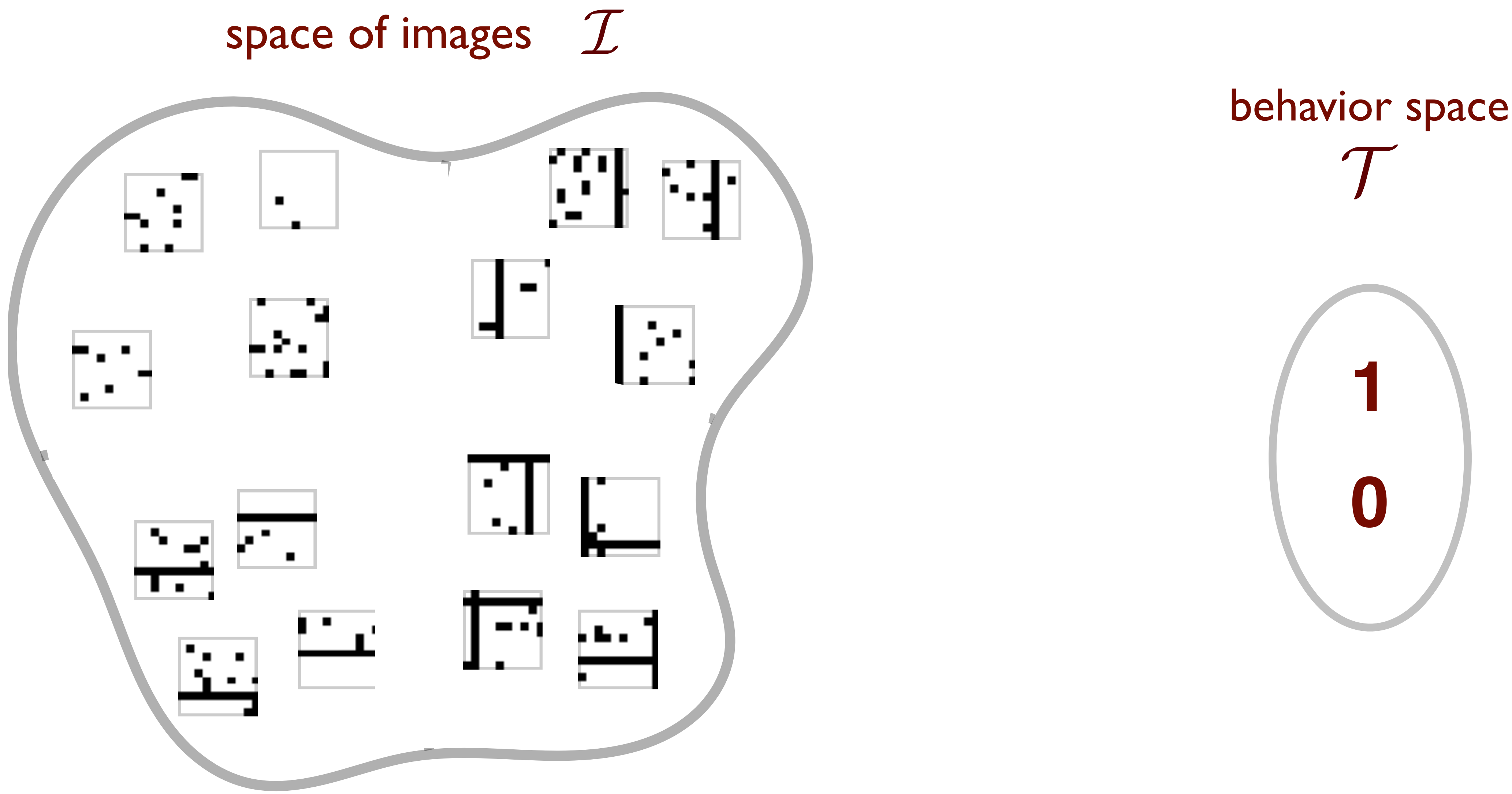
P(T=0 |) = .33

P(T=0 |) = 0

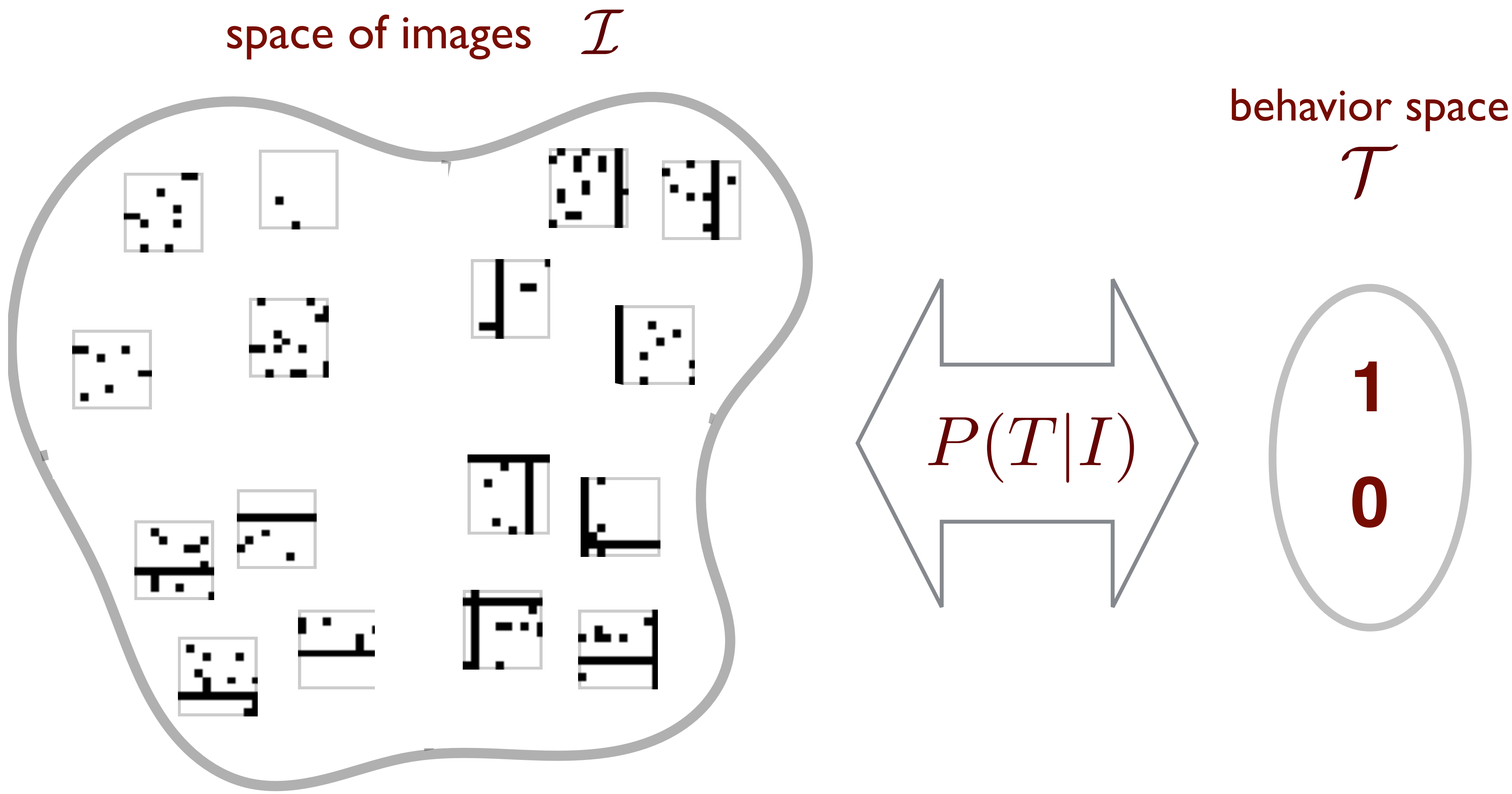
P(T=0 | do { }) = .83

P(T=0 | do { }) = .3

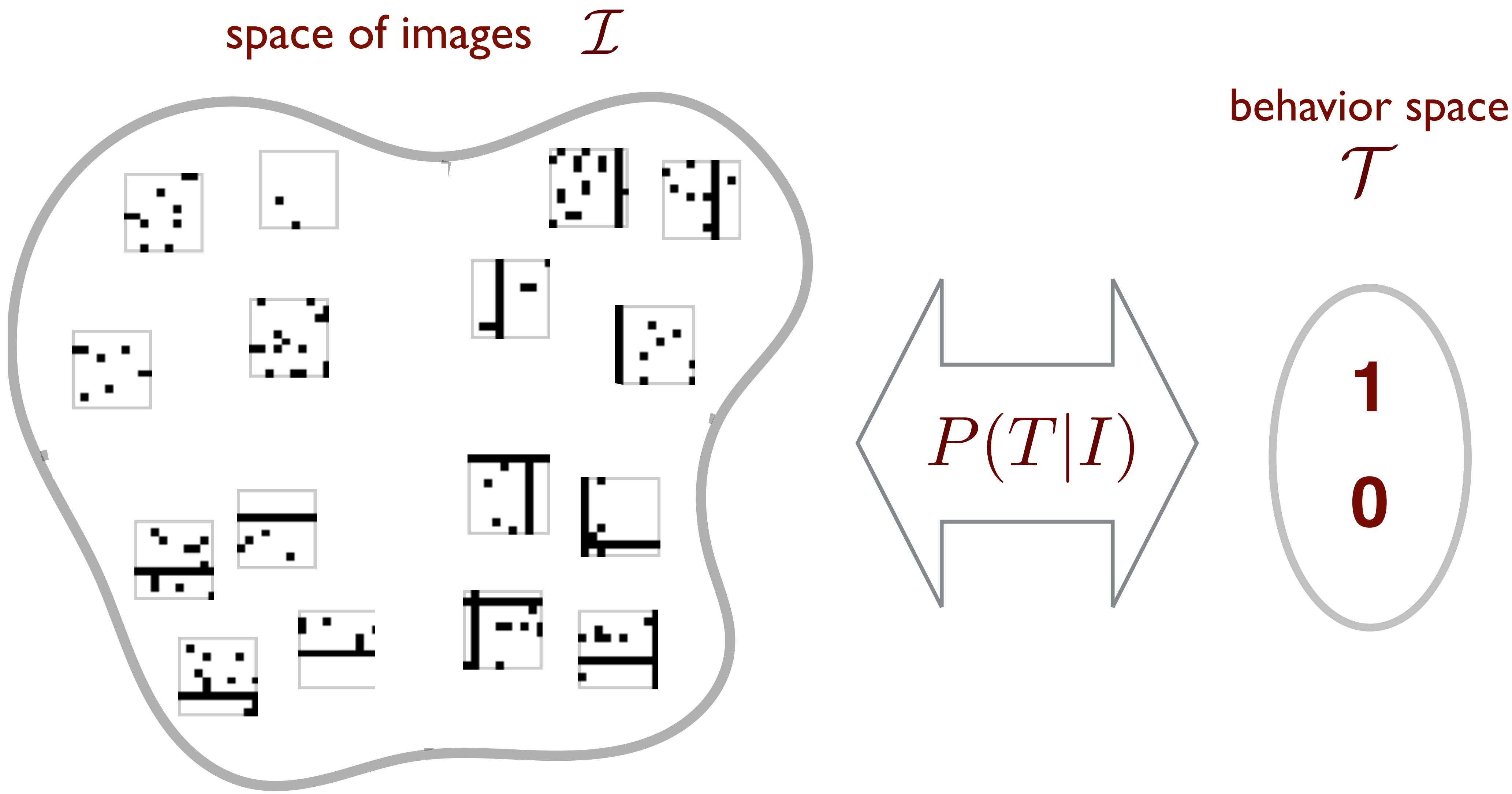
Observational Partition



Observational Partition

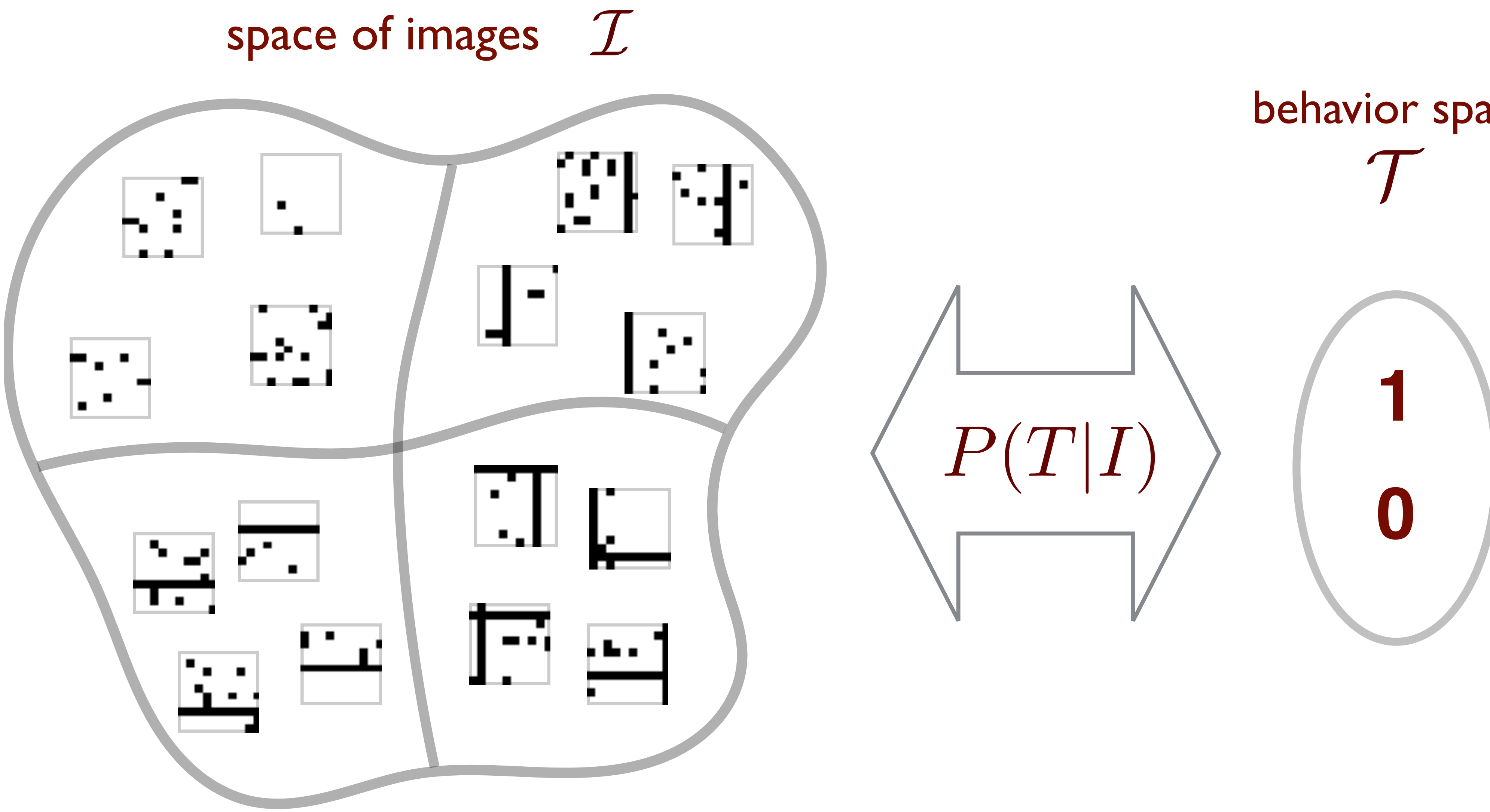


Observational Partition



- **observational partition:** partitions the space of images according to the equivalence relation induced by the conditional probability of the target behavior T given the image I

Observational Partition

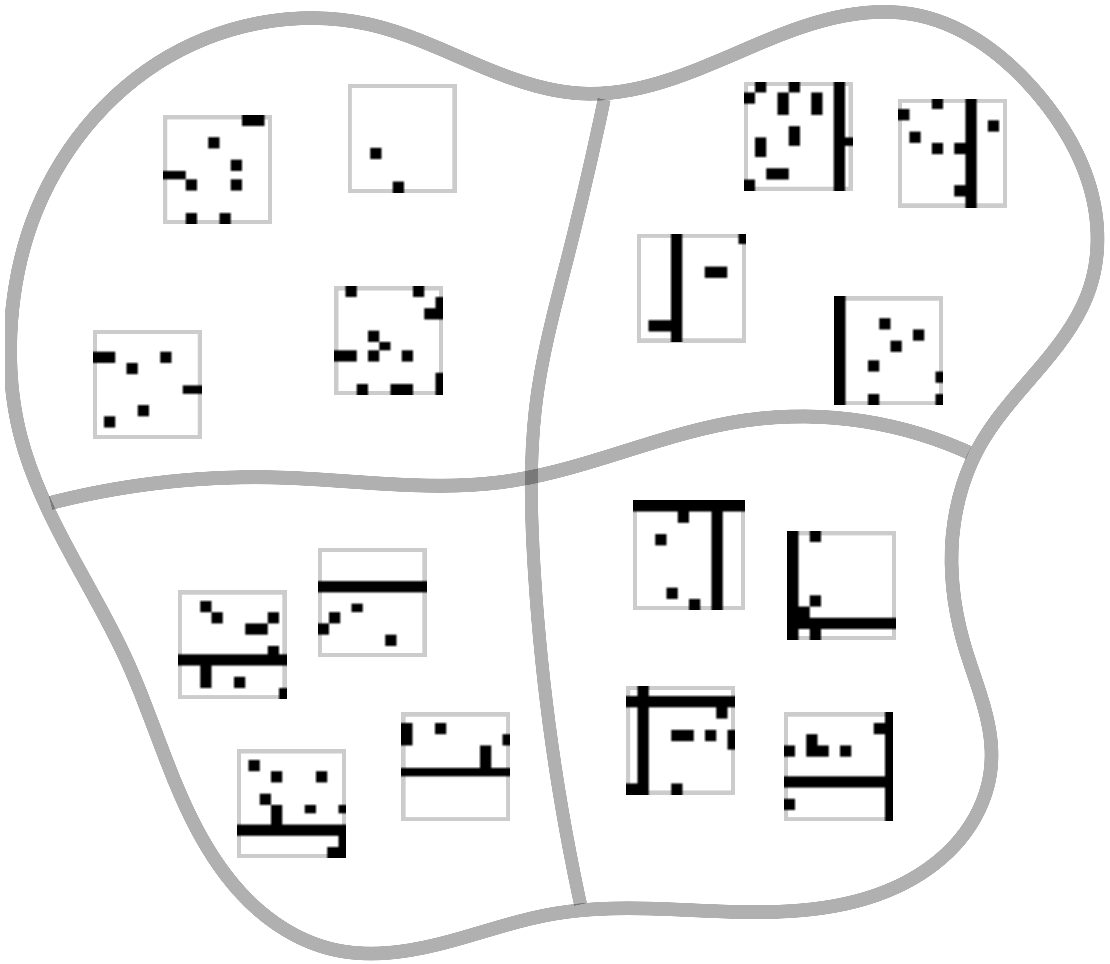


- **observational partition:** partitions the space of images according to the equivalence relation induced by the conditional probability of the target behavior T given the image I

$$i_1 \sim_I i_2 \quad \Leftrightarrow \quad \forall_{t \in \mathcal{T}} P(t \mid i_1) = P(t \mid i_2)$$

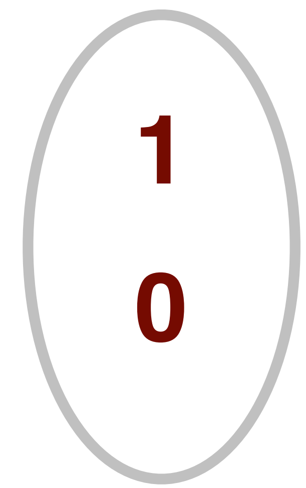
Causal Partition

space of images $\mathcal{I}$

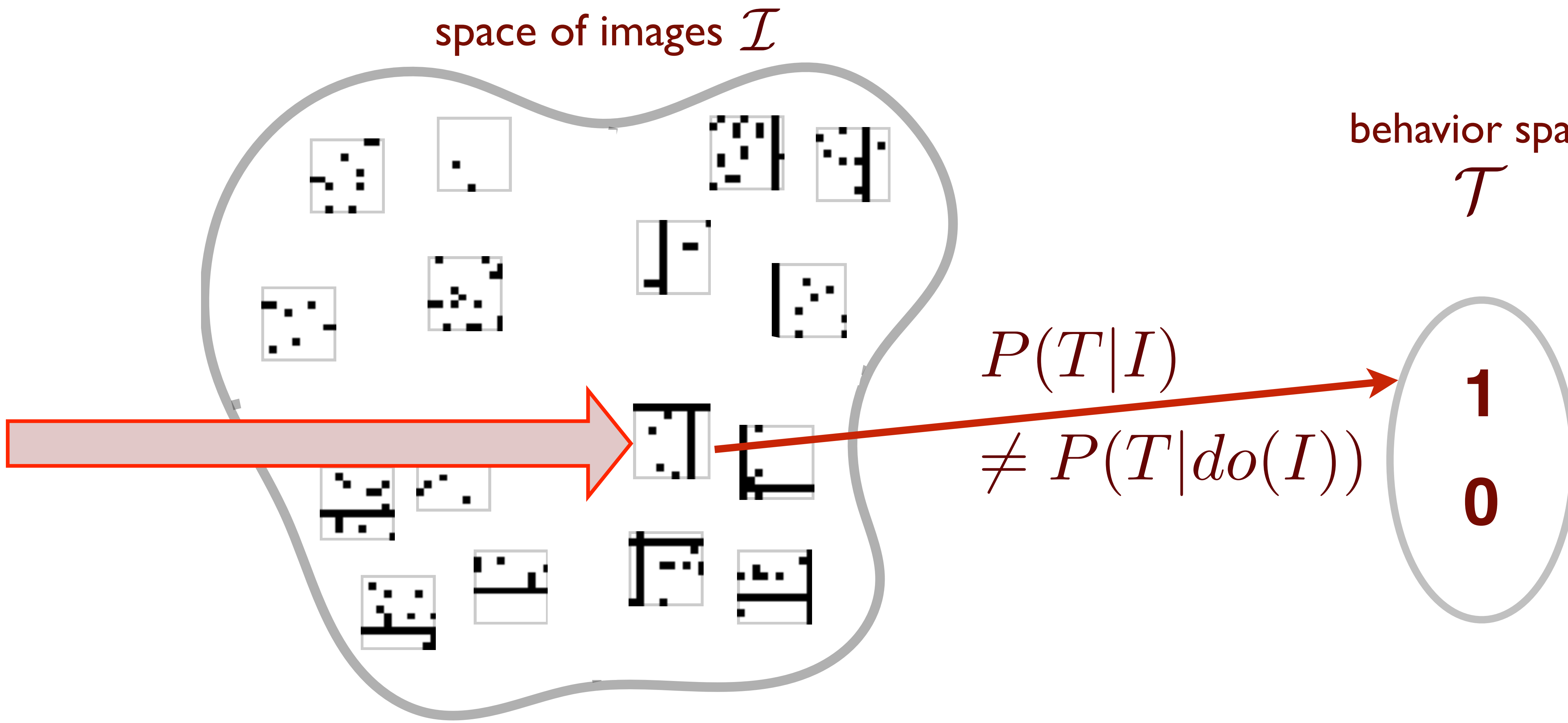


behavior space
$\mathcal{T}$

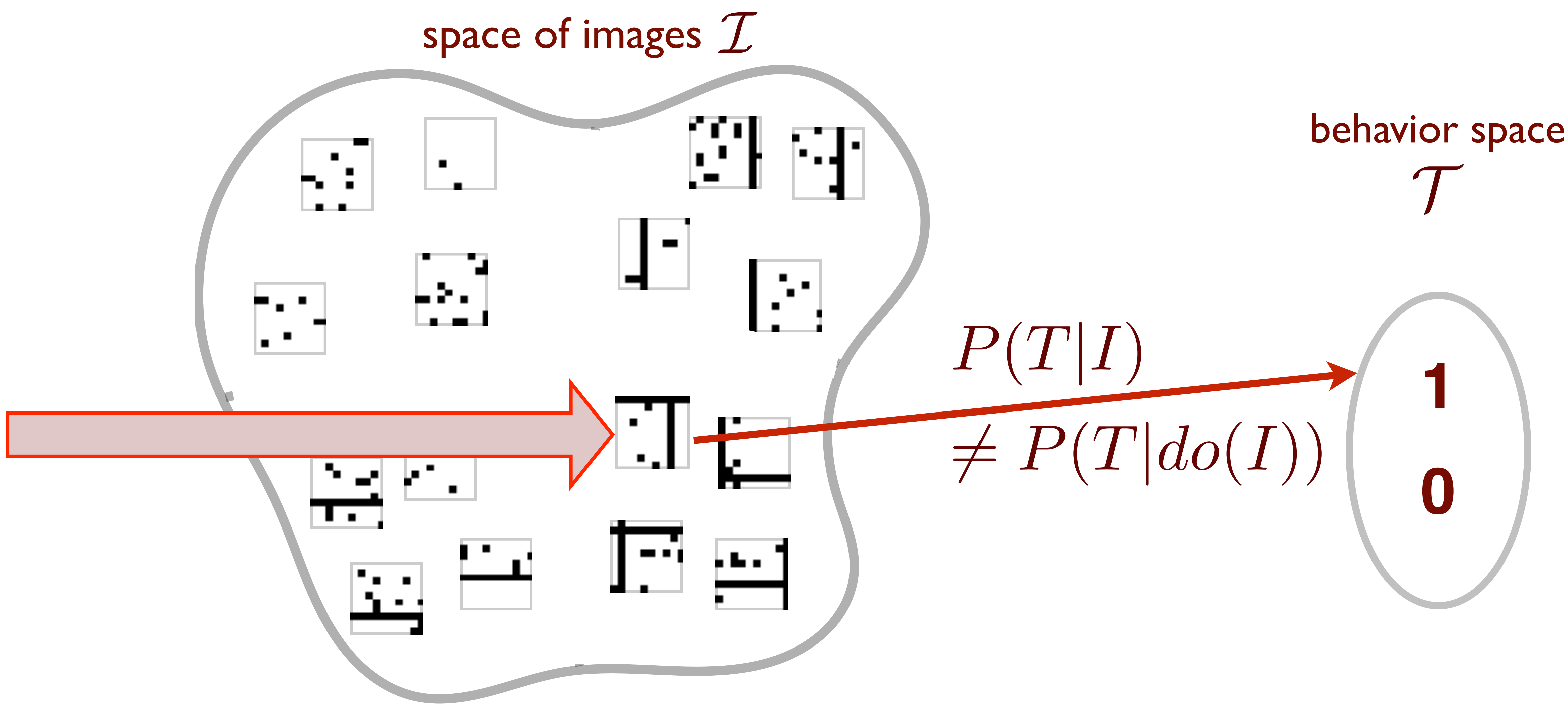
$$P(T|I)$$
$$\neq P(T|do(I))$$



Causal Partition



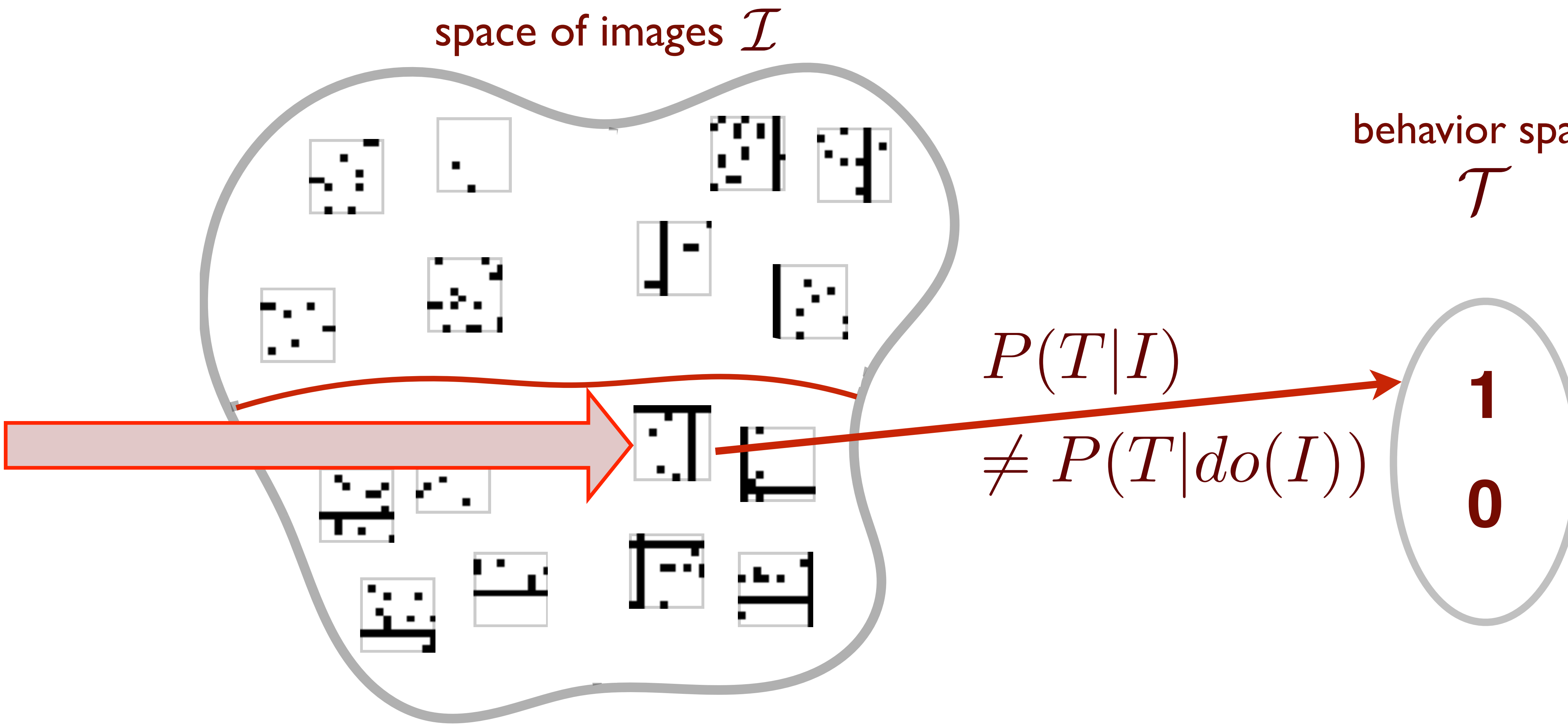
Causal Partition



- **causal partition**: partitions the image space according to the equivalence relation induced by the probability of the target behavior T given an ***INTERVENTION*** on the image

$$i_1 \sim_I i_2 \quad \Leftrightarrow \quad \forall_{t \in \mathcal{T}} P(t \mid \mathrm{do}(i_1)) = P(t \mid \mathrm{do}(i_2))$$

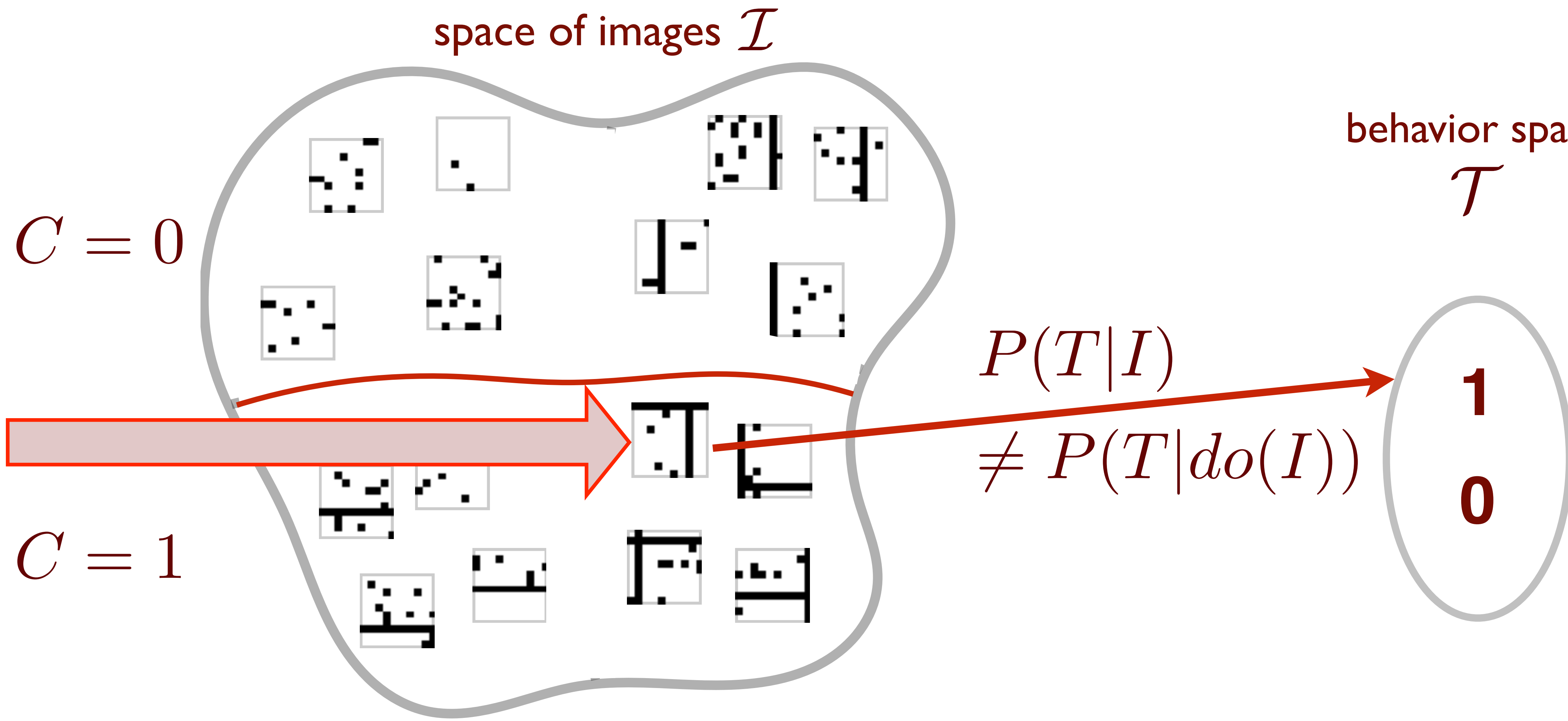
Causal Partition



- **causal partition**: partitions the image space according to the equivalence relation induced by the probability of the target behavior T given an ***INTERVENTION*** on the image

$$i_1 \sim_I i_2 \quad \Leftrightarrow \quad \forall_{t \in \mathcal{T}} P(t \mid \mathrm{do}(i_1)) = P(t \mid \mathrm{do}(i_2))$$

Causal Partition

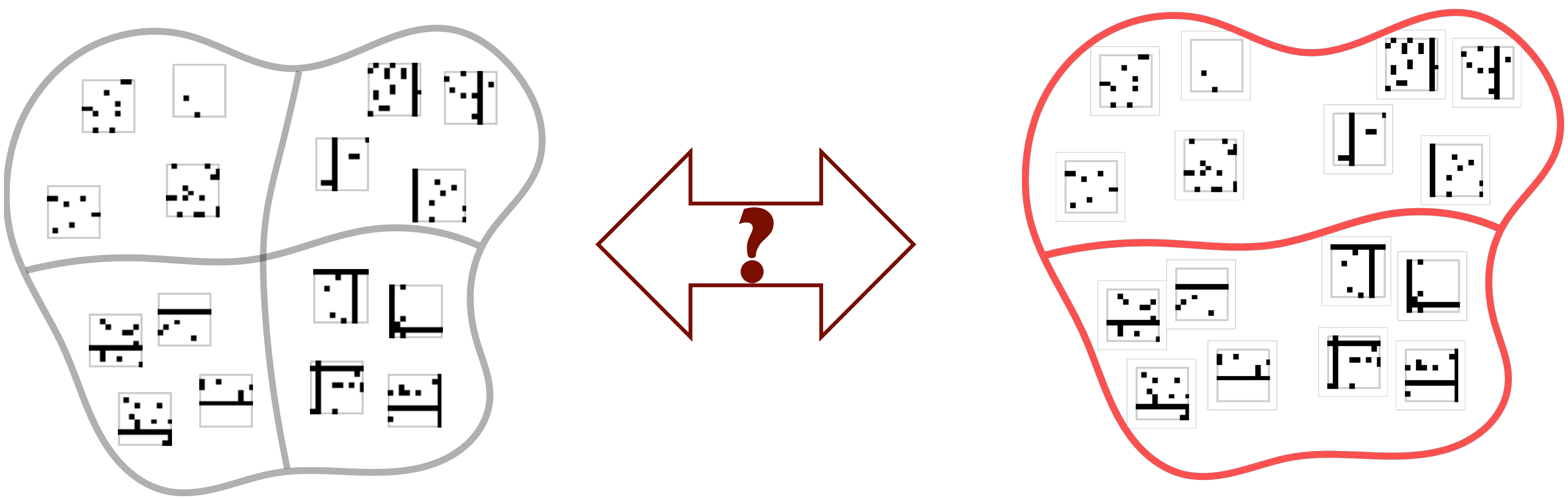


- **causal partition**: partitions the image space according to the equivalence relation induced by the probability of the target behavior T given an ***INTERVENTION*** on the image

$$i_1 \sim_I i_2 \quad \Leftrightarrow \quad \forall_{t \in \mathcal{T}} P(t \mid \mathrm{do}(i_1)) = P(t \mid \mathrm{do}(i_2))$$

- **macro cause:** the macro cause C of a target behavior T is a random variable whose value stands in a bijective relation to the causal class of the image

Observational vs. Causal Partition



observational partition of $\mathcal{I}$

$$P(T|I)$$

causal partition of $\mathcal{I}$

$$P(T|do(I))$$

Causal Coarsening Theorem

For

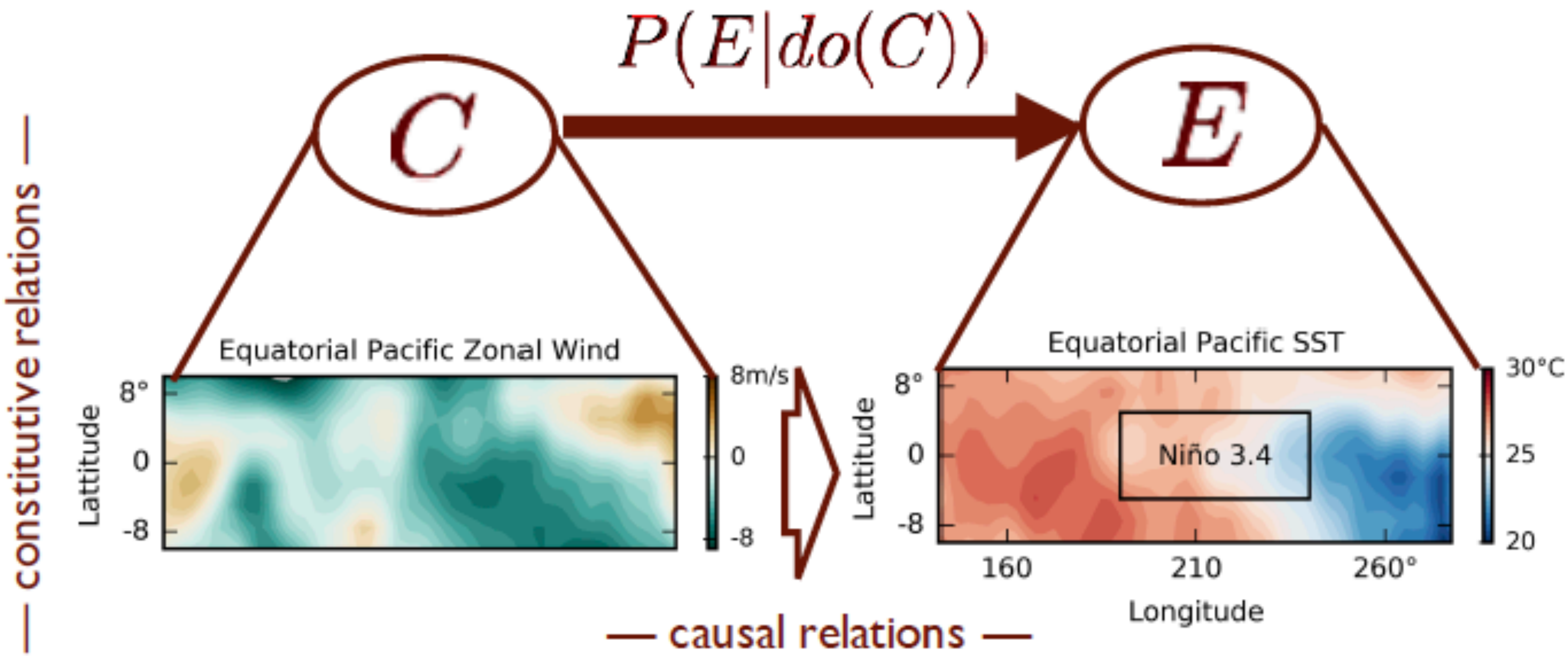
- *multinomial distributions*
- *no causal feedback*
- *[technical assumption about the nature of confounding]*

➡ *the subset of distributions that induce* ***a causal partition that is not a coarsening of the observational partition is Lebesgue measure zero.***

Applying the Causal Coarsening Theorem

- *learn the observational partition from non-experimental data*
- *under the assumptions of the theorem, the relevant causal distinctions are a subset of the detected distinctions*
- *test which distinctions are causal with a few experiments*

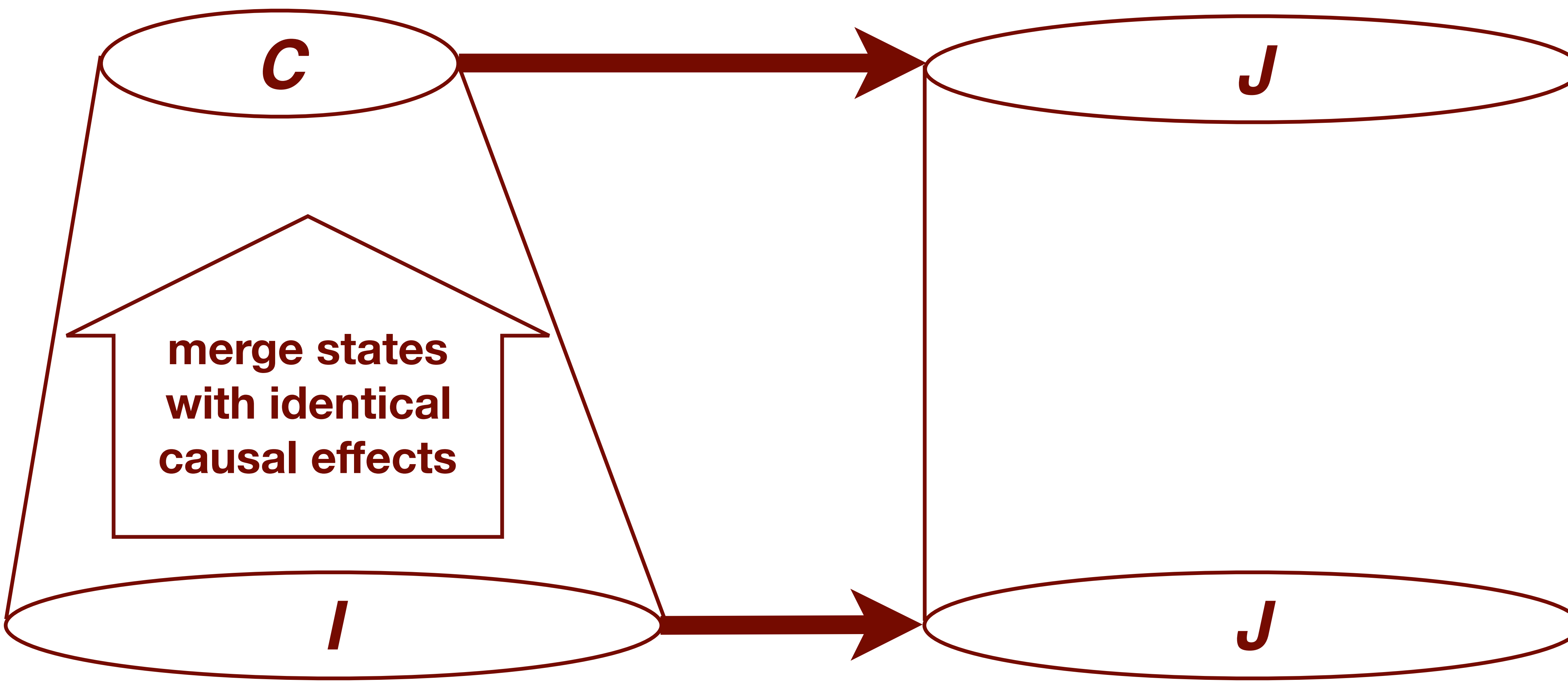
Causal Feature Learning



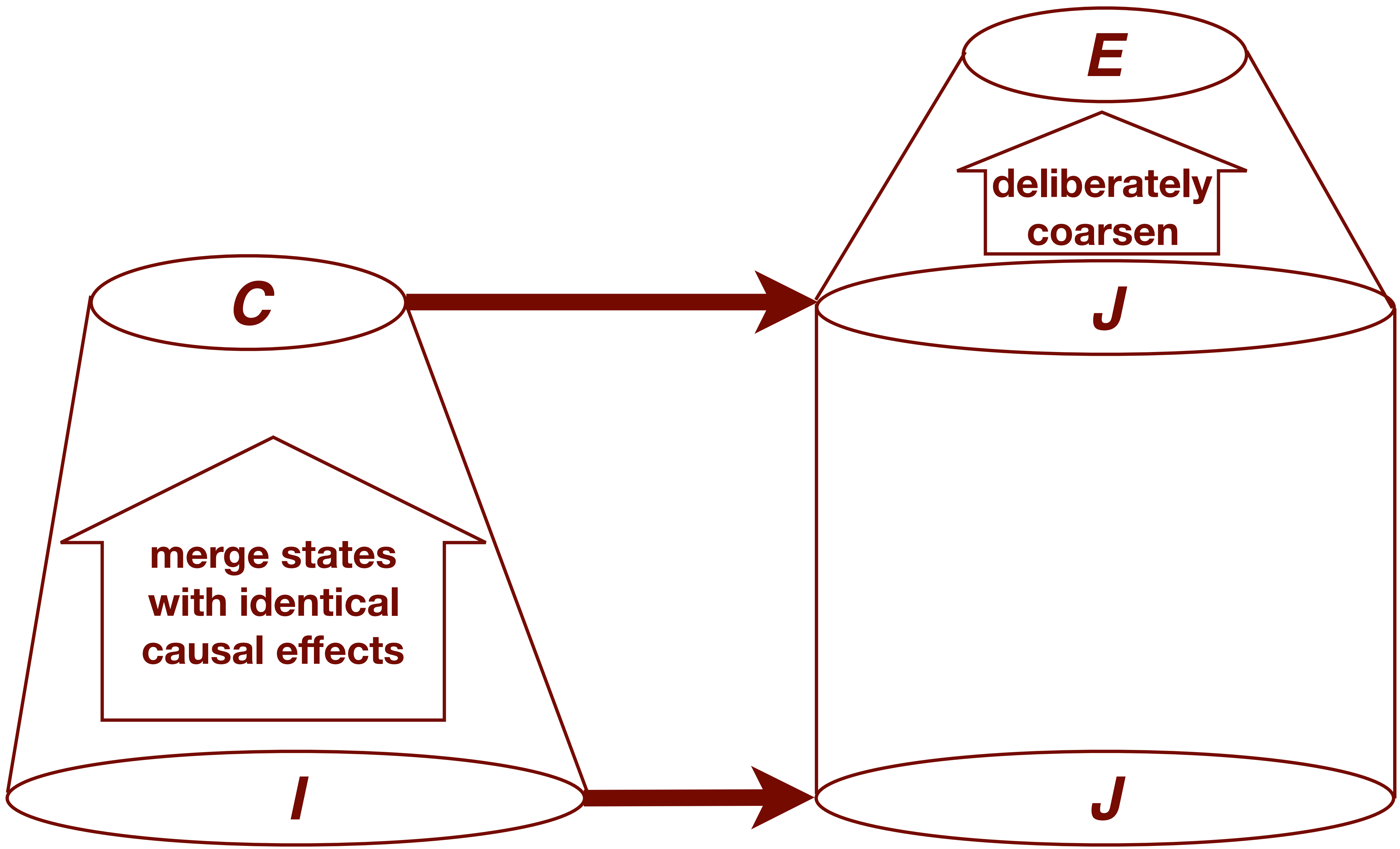
- we found the macro-level climate phenomenon of El Niño supervening on micro-level wind and sea surface temperature data of the equatorial Pacific in an entirely data driven (unsupervised) manner

Chalupka, Bischoff, Perona & Eberhardt (2016)

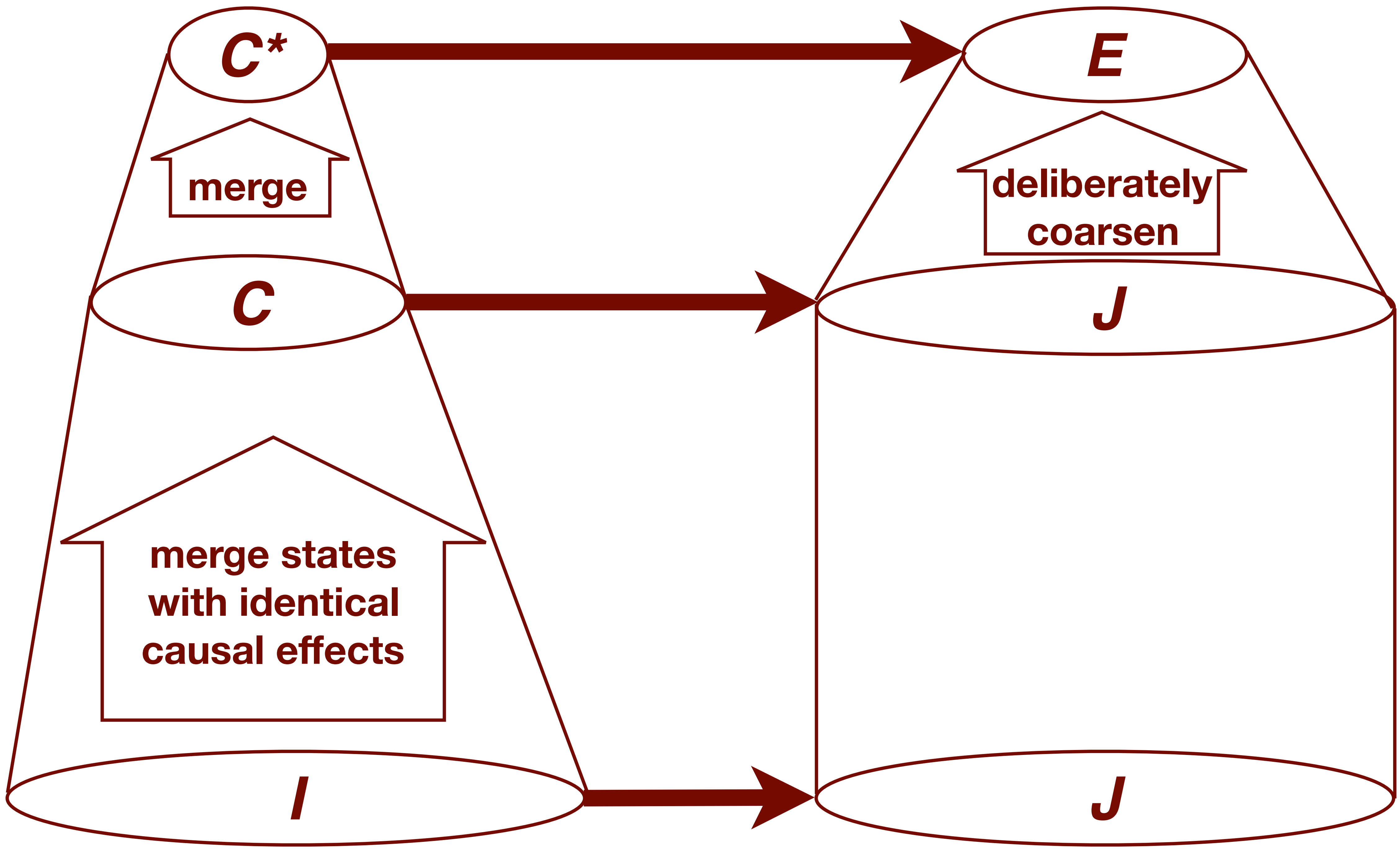
Multiple Levels of Causal Description



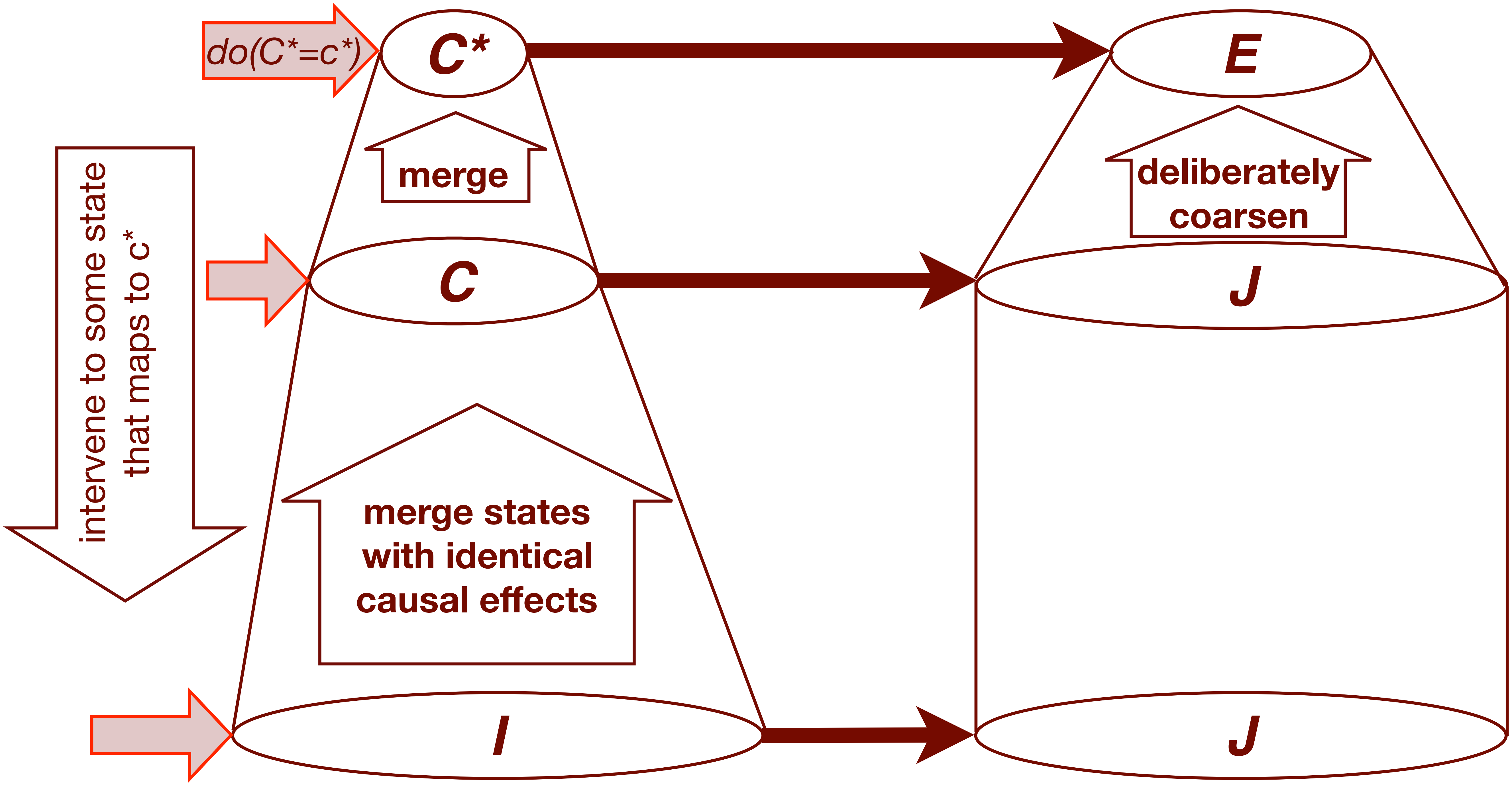
Multiple Levels of Causal Description



Multiple Levels of Causal Description



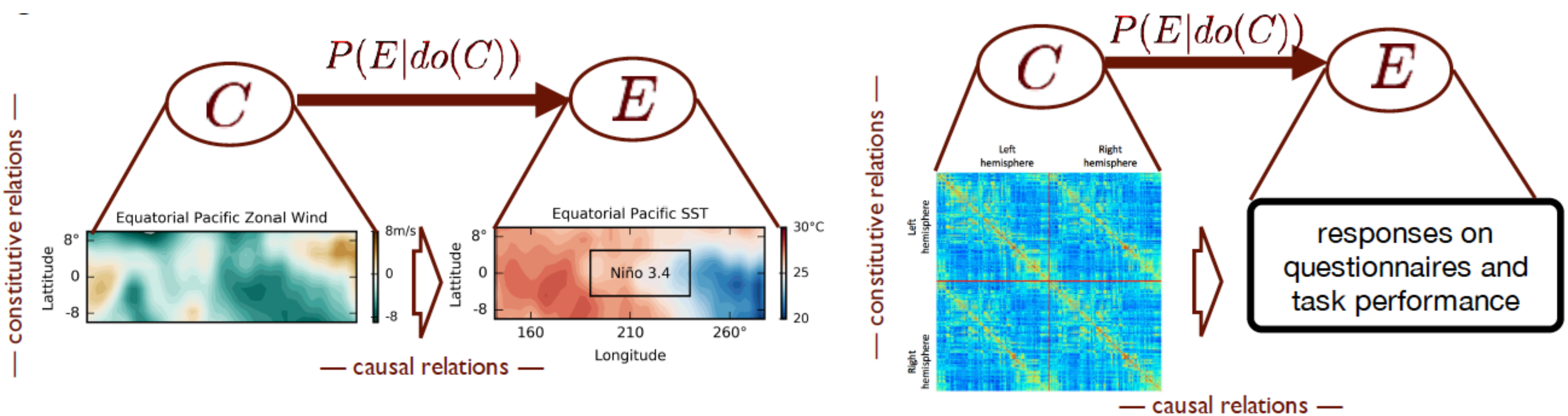
Multiple Levels of Causal Description



Causal Macro-Variables

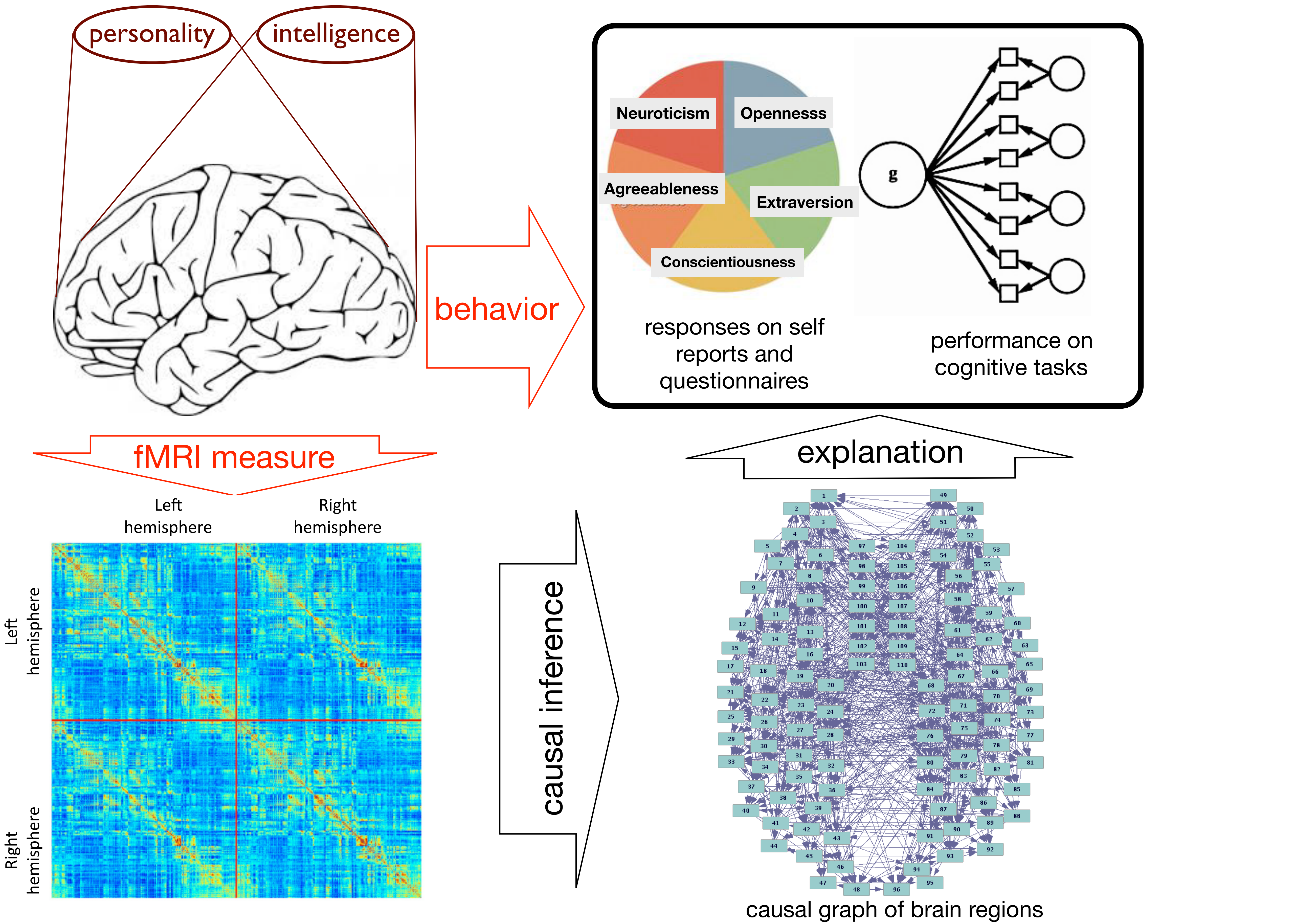
- account of causal macro-variables that
  - turns the question about the existence of causal macro-variables into an empirical question
  - identifies a privileged level of aggregation that retains exactly the causal information of the underlying micro-systems
  - supports a causal interpretation in terms of intervention (and avoids known problems of causal variable definition)
  - is domain-general
- algorithms that discover/construct such causal macro-variables
- applications as proof of concept

Causal Feature Learning



- can we use the same approach to search for macro-level neural features that are causal of behavior?

Neuroscience-based Psychology



Collaborators

Krzysztof Chalupka

Pietro Perona

K. Chalupka, P. Perona, and F. Eberhardt. Visual causal feature learning. In Proceedings of UAI, 2015.
K. Chalupka, P. Perona, and F. Eberhardt. Multi-level cause-effect systems. In Proceedings of AISTATS, 2016.
K. Chalupka, T. Bischoff, P. Perona, and F. Eberhardt. Unsupervised discovery of El Niño using causal feature learning on microlevel climate data. In Proceedings of UAI 2016.
K. Chalupka, F. Eberhardt, and P. Perona. Causal Feature Learning: an overview. Behaviormetrika, 2016.

All code available in python from Chalupka's webpage.

Collaborators

Krzysztof Chalupka

Pietro Perona

K. Chalupka, P. Perona, and F. Eberhardt. Visual causal feature learning. In Proceedings of UAI, 2015.
K. Chalupka, P. Perona, and F. Eberhardt. Multi-level cause-effect systems. In Proceedings of AISTATS, 2016.
K. Chalupka, T. Bischoff, P. Perona, and F. Eberhardt. Unsupervised discovery of El Niño using causal feature learning on microlevel climate data. In Proceedings of UAI 2016.
K. Chalupka, F. Eberhardt, and P. Perona. Causal Feature Learning: an overview. Behaviormetrika, 2016.

All code available in python from Chalupka's webpage.

Thank you!