Deterministic and stochastic causality

Michael Joffe
Dept of Epidemiology & Public Health
Imperial College London

Beware! – danger of systematic bias

- people tend to avoid stating the obvious, and focus on what is difficult
- in trying to explain how science works, to uncover the “laws” of discovery, this is ***dangerous***: if only the “difficult” examples are studied, this produces a systematic bias towards seeing how problematic it all is, and fosters ***relativism***
- hard cases make bad “laws”!

- "Philosophers agree that causal propositions cannot be proved, and find flaws or practical limitations in all philosophies of causal inference." (Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. AJPH 2005 Suppl 1, 95: S144-50.)

- As a scientist and non-philosopher (and non-statistician), I will approach this not from the viewpoint “how can we know anything?”, but rather, “we know quite a lot – *how* has this happened?” – i.e. like a scientist

3 kinds of scientific question

(1) WHAT CAUSES WHAT? (or, "*does* x cause y?")

(2) HOW? – explanation; mechanism

(3) HOW MUCH OF x CAUSES HOW MUCH OF y?

- Why these 3?
- What order? – I suggest it has to be (1) first, as the others depend on a "yes" to this
- An example of (1) and (2) would be, "does acupuncture relieve pain?", and if so, "how does it do this?"

1 kind of non-scientific question

(4) WHY? – meaning; values

- Superficially, (1) and (4) can be confused in everyday language, as in “why does my hand move?”

What causes my hand to move?

- An answer to this (type (1) question) would be, from proximal to distal:

(ii) biochemical changes in the muscles making some contract and others relax

(iii) the motor neurons cause the muscles to contract/relax

(iv) the neurons in the CNS cause the motor neurons to fire (enough)

(v) “I” cause the CNS neurons to fire in a pattern that brings about (iii) – ***agency***

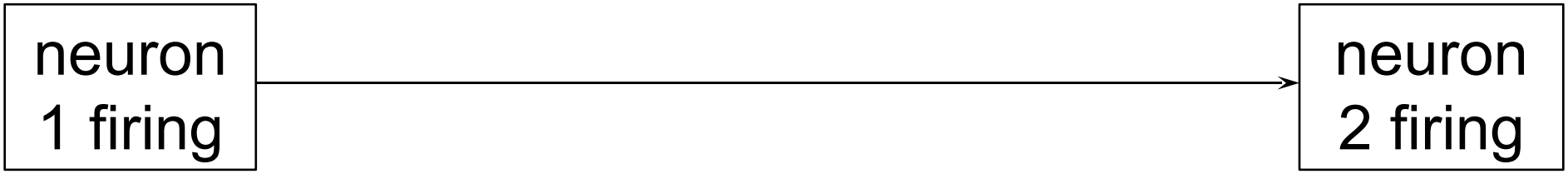
Agency

- to explain anything in this sphere we have to look into how I (my mind/brain) integrate information, and what purpose this movement serves – cf the “WHY” question earlier, number (4)
- humans are programmed to see things in terms of “why?” and agency; science involves ignoring them; psychology and social sciences need agency to be restored
- this is beyond the scope of this talk!

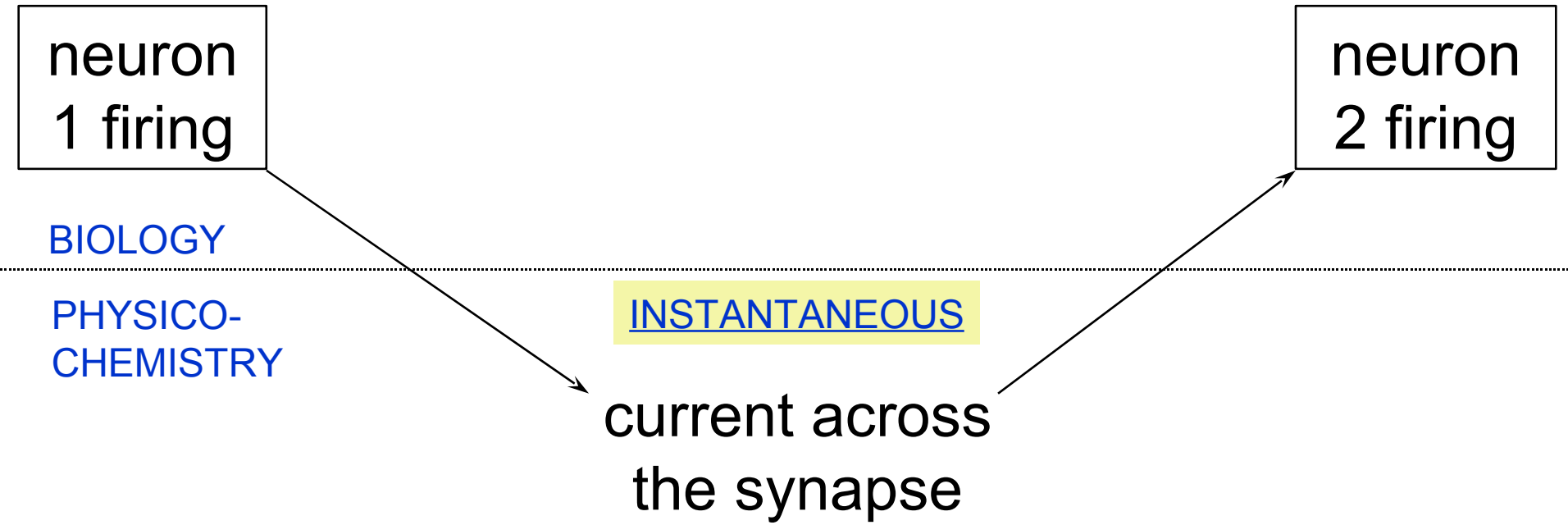
What causes my hand to move?

- I'll focus on (ii) and (iii), respectively the neuromuscular junction and the nerve-nerve junction (synapse) in the spinal cord
- back to the 3 kinds of scientific question: I've already answered (1) – but how do I know?
  - a severed nerve means I cannot move my arm
  - stimulating a nerve makes a muscle contract
- the 1st is observational, the 2nd experimental
- neither involves measurement (though we also would want to do this)

HOW DO NEURONS CONNECT?

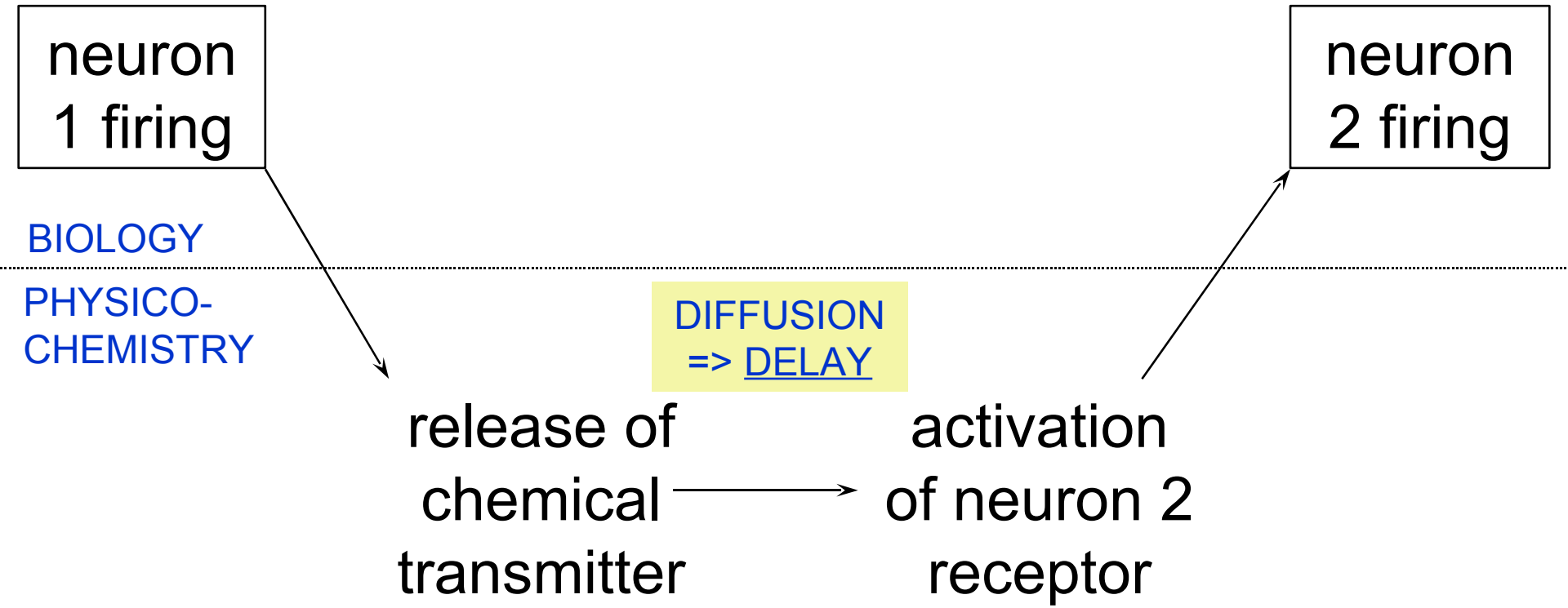


HYPOTHESIS 1: ELECTRICAL



EXPLANATION IS SEEN AS DESCRIPTION AT A DEEPER LEVEL – HERE AS PHYSICO-CHEMISTRY, BUT IT COULD BE e.g. MICROSCOPIC DESCRIPTION, OR MICROBIOLOGY – IT INVOLVES DESCRIBING THE KEY COMPONENTS OF THE SYSTEM AND THEIR PROPERTIES

HYPOTHESIS 2: CHEMICAL

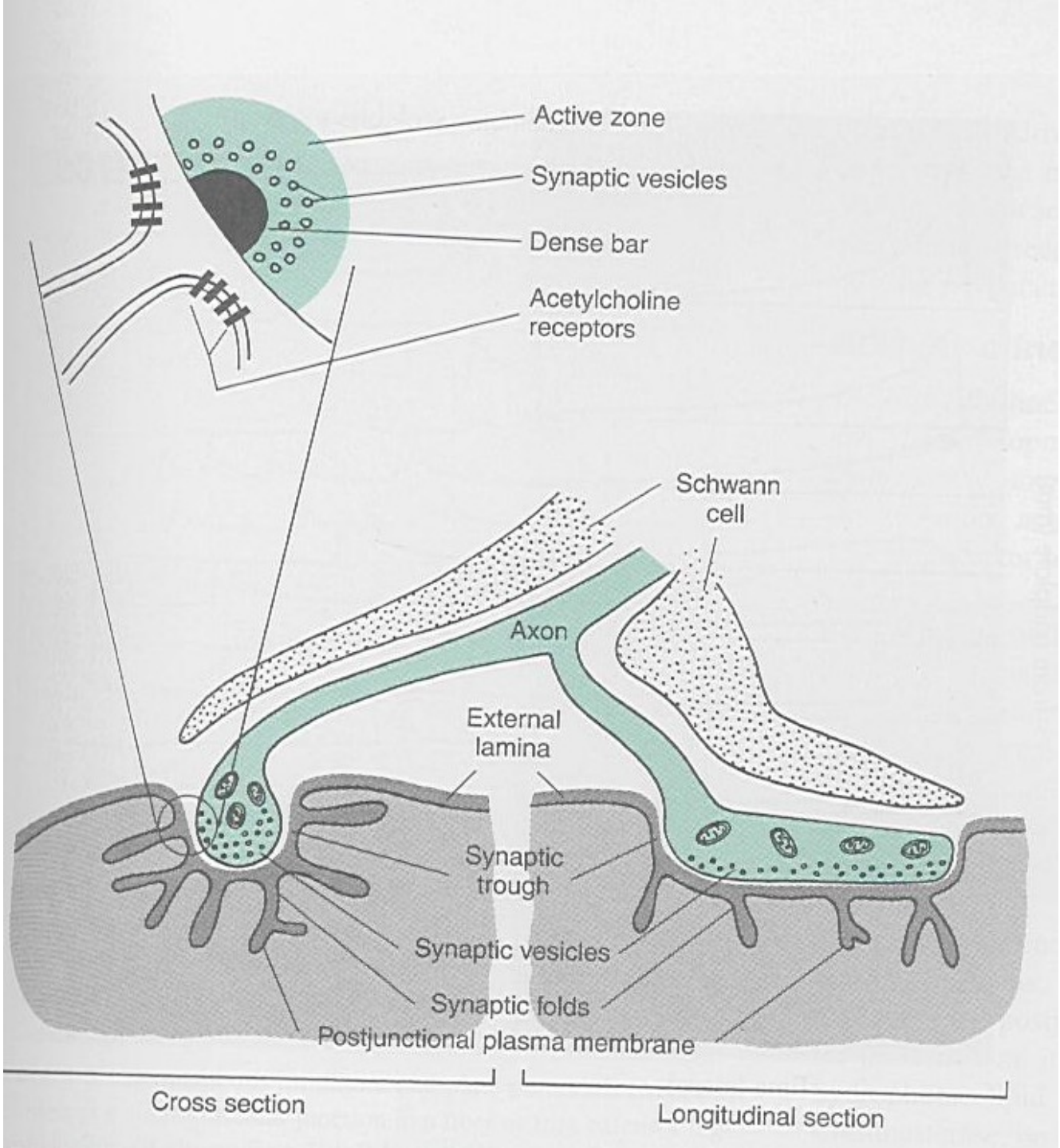


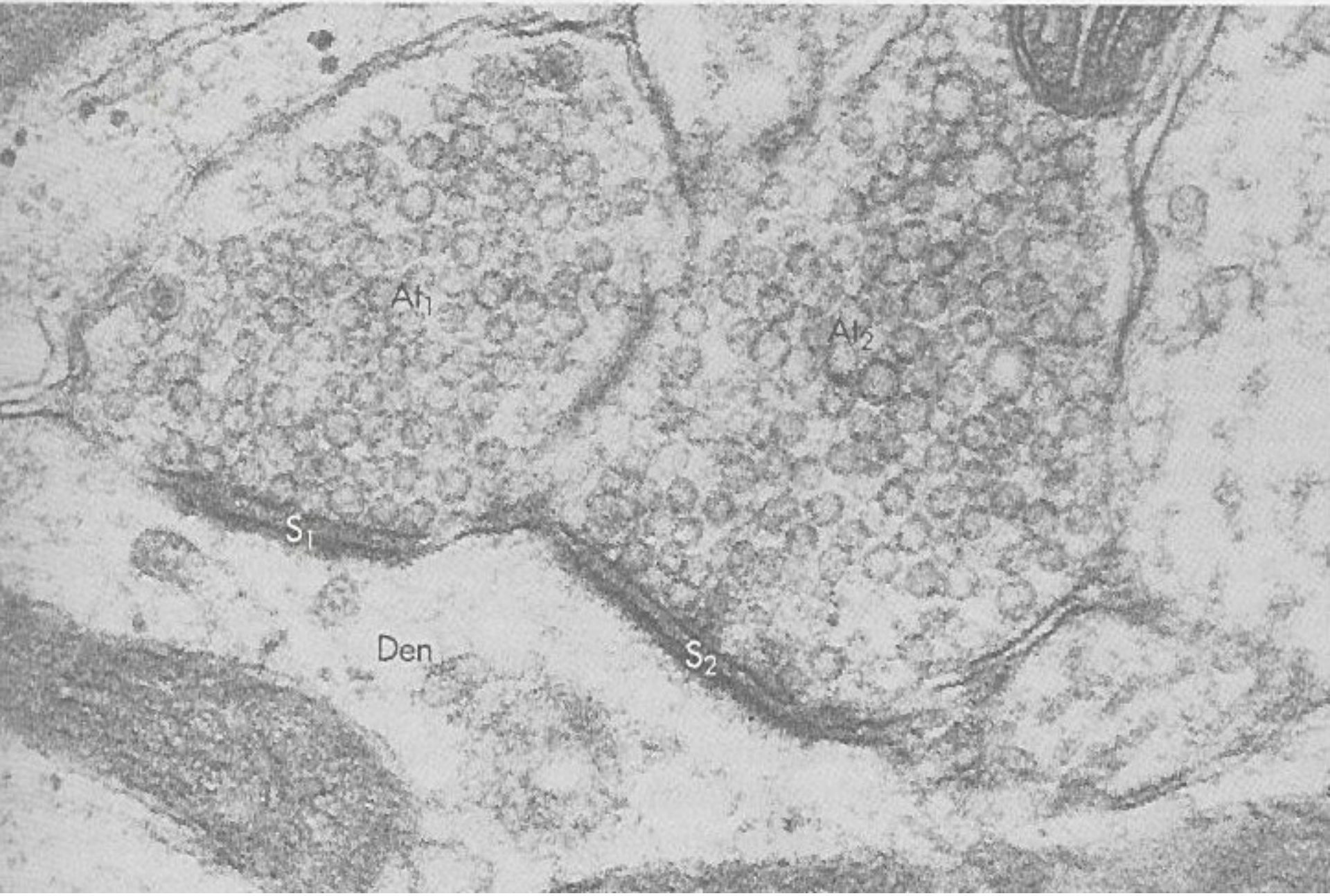
THESE ARE ACTUAL CHEMICAL PROCESSES – i.e. THE MOLECULES MOVE ACROSS THE SYNAPSE

THE CAUSATION IS DETERMINISTIC AND ALL-OR-NONE – AS LONG AS A THRESHOLD LEVEL IS REACHED

Electric *vs* chemical transmission I

- timing of the gap between neurons firing
- vesicles of transmitter seen in neuron 1 using electron microscopy
- chemical detection of transmitter
- presence of enzymes e.g. to synthesise the transmitter
- synthesis of transmitter and finding that it will activate neuron 2
- ALL EXCEPT THE LAST ARE BASED ON OBSERVATION NOT EXPERIMENTATION
- ONLY THE FIRST IS “MEASUREMENT”


Active zone
Synaptic vesicles
Dense bar
Acetylcholine receptors
Schwann cell
Axon
External lamina
Synaptic trough
Synaptic vesicles
Synaptic folds
Postjunctional plasma membrane
Cross section
Longitudinal section


Al1
Al2
S1
Den
S2

Electric *vs* chemical transmission II

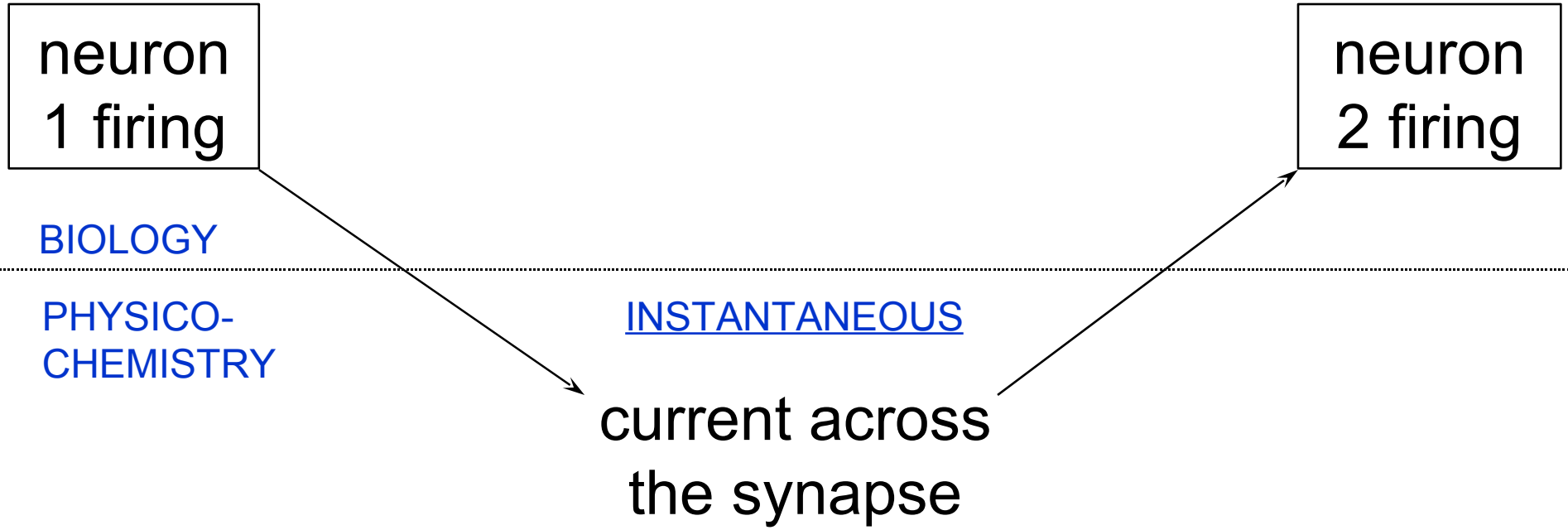
- we know that transmission is chemical, with absolute 100% certainty
- there are several interlocking pieces of evidence, all reinforcing the same story – our certainty that it is true relies heavily on this feature
- the story relies on three features (but we know much more):

4. the motor neuron produces acetyl choline
5. acetyl choline diffuses across short spaces
6. the muscle fibre responds to acetyl choline

Electric *vs* chemical transmission III

- however, even if a large number of types of synapse are tested, the possibility remains that there are others that are electrical and not chemical (Popper: “all swans are white”) – the ***impossibility of completeness***

HYPOTHESIS 1: ELECTRICAL



IT TURNS OUT THAT PURELY ELECTRICAL CONNECTIONS ALSO EXIST, CALLED GAP JUNCTIONS – NOT ALL SWANS ARE WHITE!

Electric *vs* chemical transmission IV

- all this is on the explanation for *propagation* – but what caused the first neuron to fire?
- the question of causation is often confused with that of ***initiation*** – see later
- this is typical of studying biological systems: biologists isolate one part of the process and study its mechanism without worrying about what starts it in motion – they can do this because it is a self-perpetuating system

INHERITANCE

affected by Huntingdon's disease

affected offspring

50% of gametes have affected gene

gamete is affected

THIS INVOLVES THE PHYSICAL MOVEMENT OF A CHROMOSOME, FOLLOWED BY REPLICATION

gamete is not affected

THE NON-DIRECTIONAL VERTICAL ARROW => "THIS MEANS THAT …"

THE CAUSATION IS STOCHASTIC IN ONE SENSE; MORE IMPORTANT THAN THE VALUE OF 50% IS THAT IT INDICATES PARTICULATE INHERITANCE, WHICH IS CRUCIAL TO THE NEO-DARWINIAN SYNTHESIS

Meiosis: the physical basis

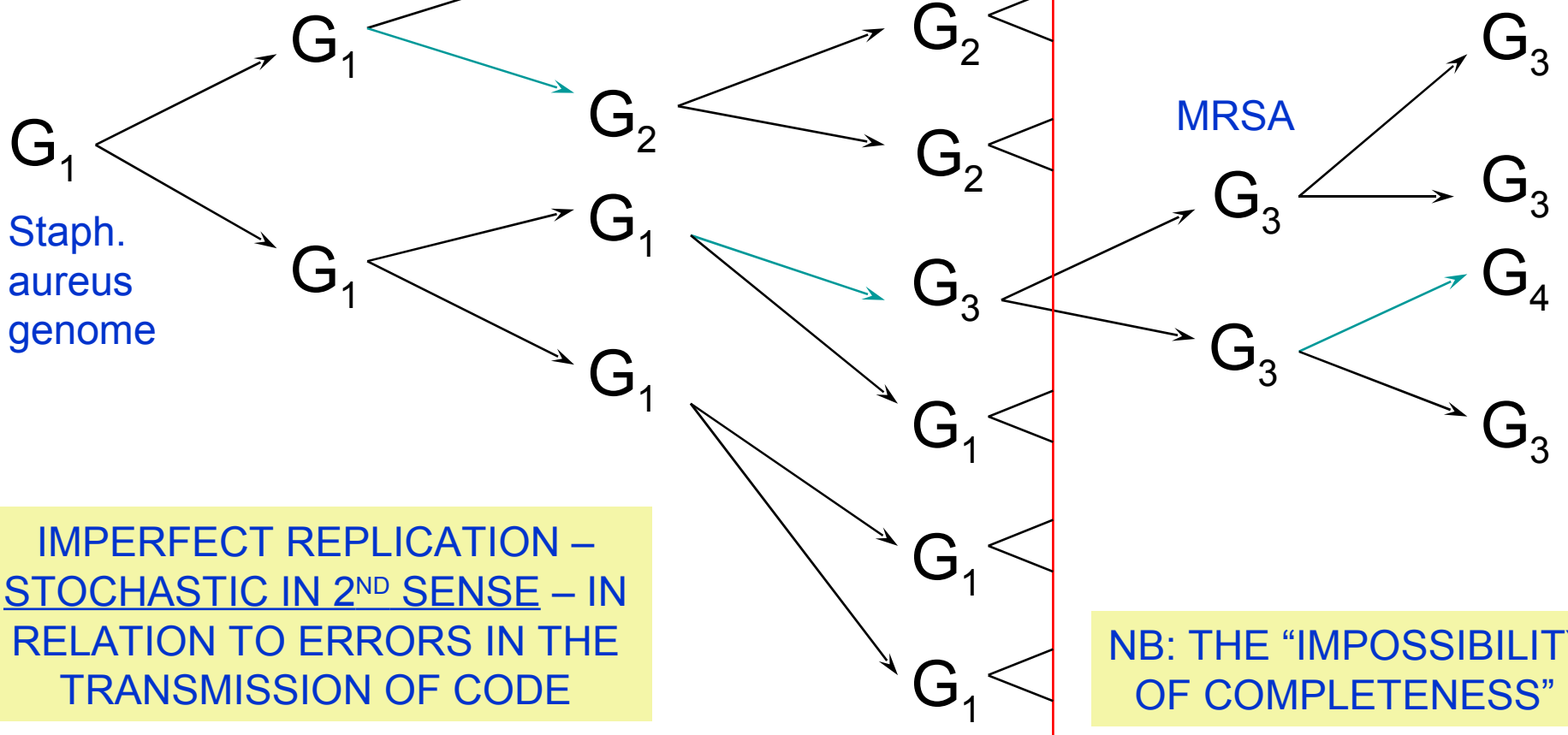
EVOLUTION

- IMPERFECT REPLICATION + INFORMATIVE ENVIRONᵀ

THE CAUSAL EXPLANATION IS NOT IN TERMS OF A "DEEPER LEVEL" NOW, BUT RATHER A SET OF HYPOTHETICAL PROPERTIES

METHICILLIN EXPOSURE

THIS IS A CHEMICAL EXAMPLE, BUT IT COULD BE e.g. A NEW PREDATOR



Staph. aureus genome

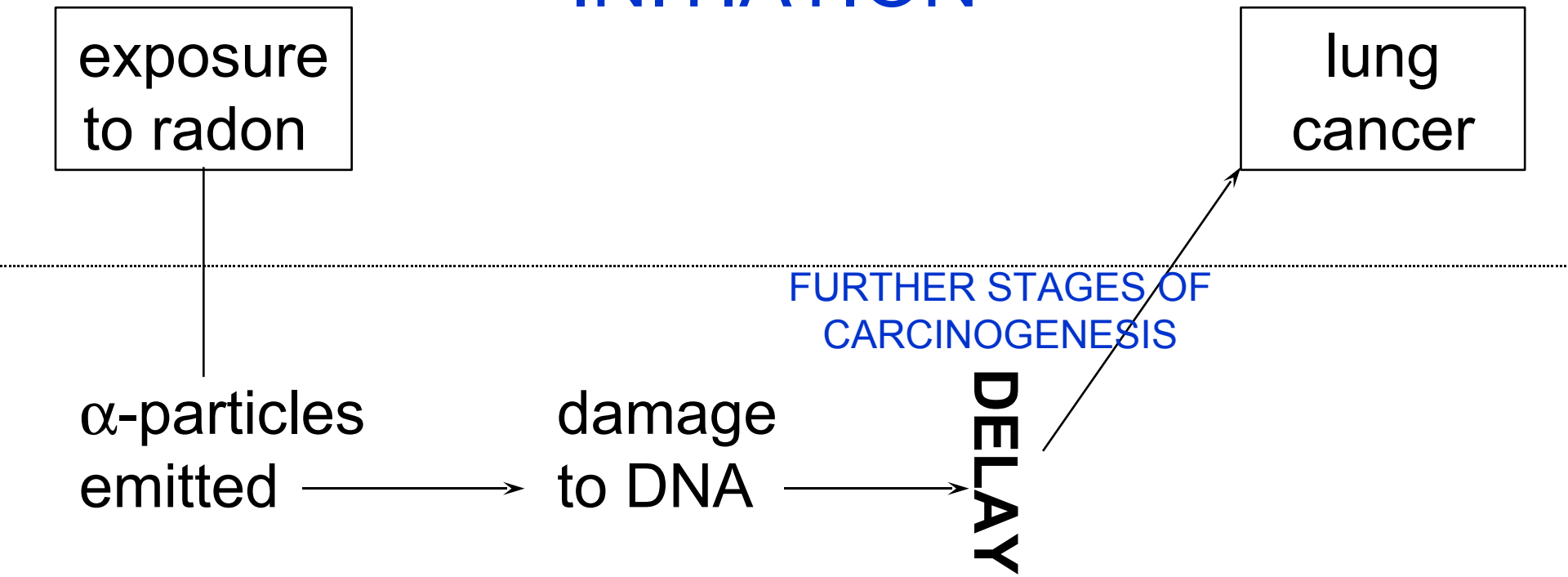
IMPERFECT REPLICATION – STOCHASTIC IN 2ND SENSE – IN RELATION TO ERRORS IN THE TRANSMISSION OF CODE

NB: THE "IMPOSSIBILITY OF COMPLETENESS"

So far:

- we have dealt with “impersonal causation” – as contrasted with “agency” – in the context of a self-perpetuating system, that of life
- the other causal systems are brought into existence by this evolutionary process – apparent teleology – or by human agency
- in self-perpetuating systems, nothing is “new” or unexpected, and we could call the causality “endogenous ” to the system
- this can be contrasted with what I will call “initiation” (exogenous causation)

INITIATION



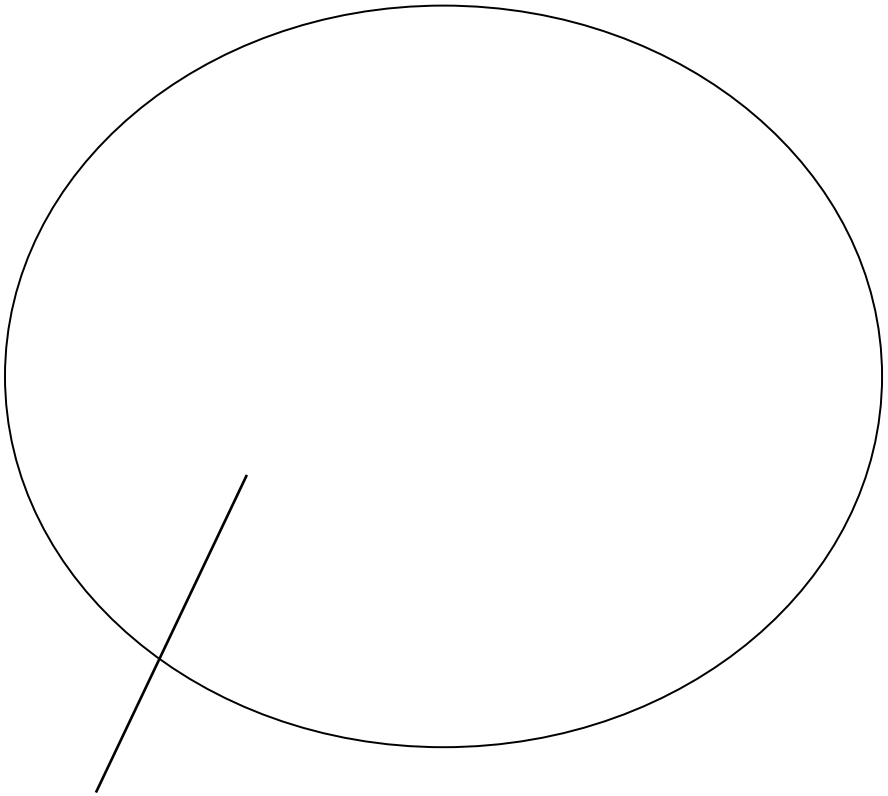
AN a-PARTICLE MAY OR MAY NOT HIT THE DNA – STOCHASTIC ALL-OR-NONE RESPONSE – THIS IS INHERENTLY ASYMMETRIC UNLIKE EQUATIONS, WHICH ARE SYMMETRICAL – BECAUSE THE PATTERN OF DAMAGE WOULD DIFFER ON DIFFERENT OCCASIONS, ALBEIT WITH THE SAME PROBABILITY DISTRIBUTION

STOCHASTIC CAUSATION IN A 3RD SENSE: A MATTER OF CHANCE WHETHER A PARTICULAR CELL'S DNA IS DAMAGED – ALSO, THERE IS IMPERFECT REPLICATION OF CODE (2ND SENSE)

Stochastic causation

- three types (not in this temporal order):
  - random segregation of chromosomes (Mendel)
  - imperfect replication (mutation etc)
  - an α-particle may hit and damage DNA – initiation, or exogenous causation
- other things are *treated as* stochastic:
  - measurement error
  - omitted variables

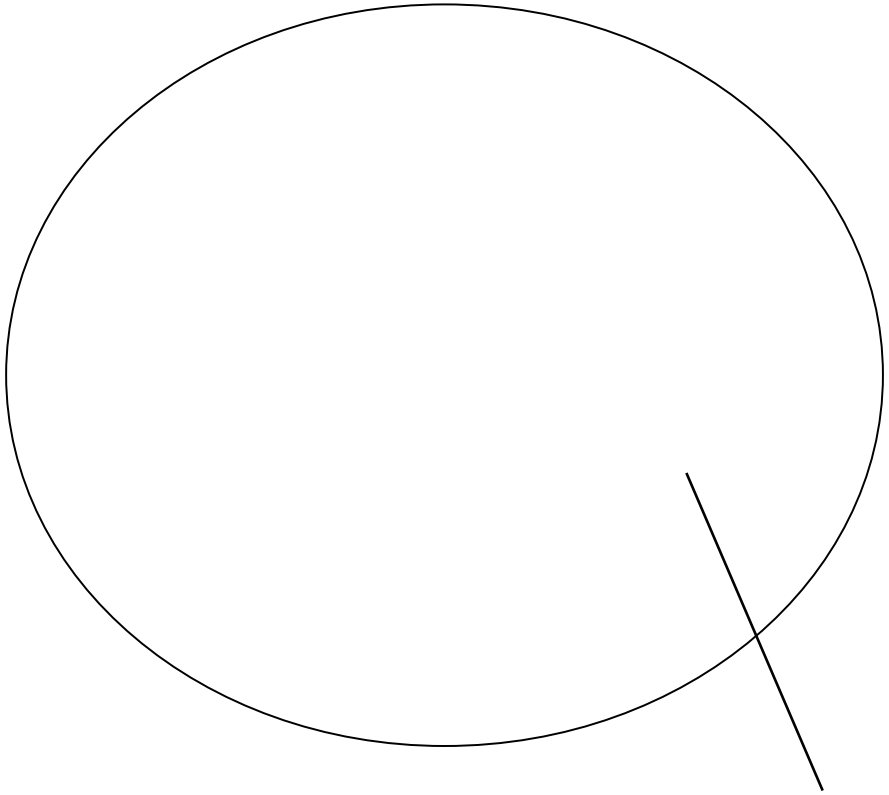
STOCHASTIC CAUSATION



ASSOCIATIONAL (JOINT DISTRIBUTION) ONLY – i.e. stochastic but without causation – symmetrical

AS IN SOME INTERPRETATIONS OF GRAPHICAL MODELS

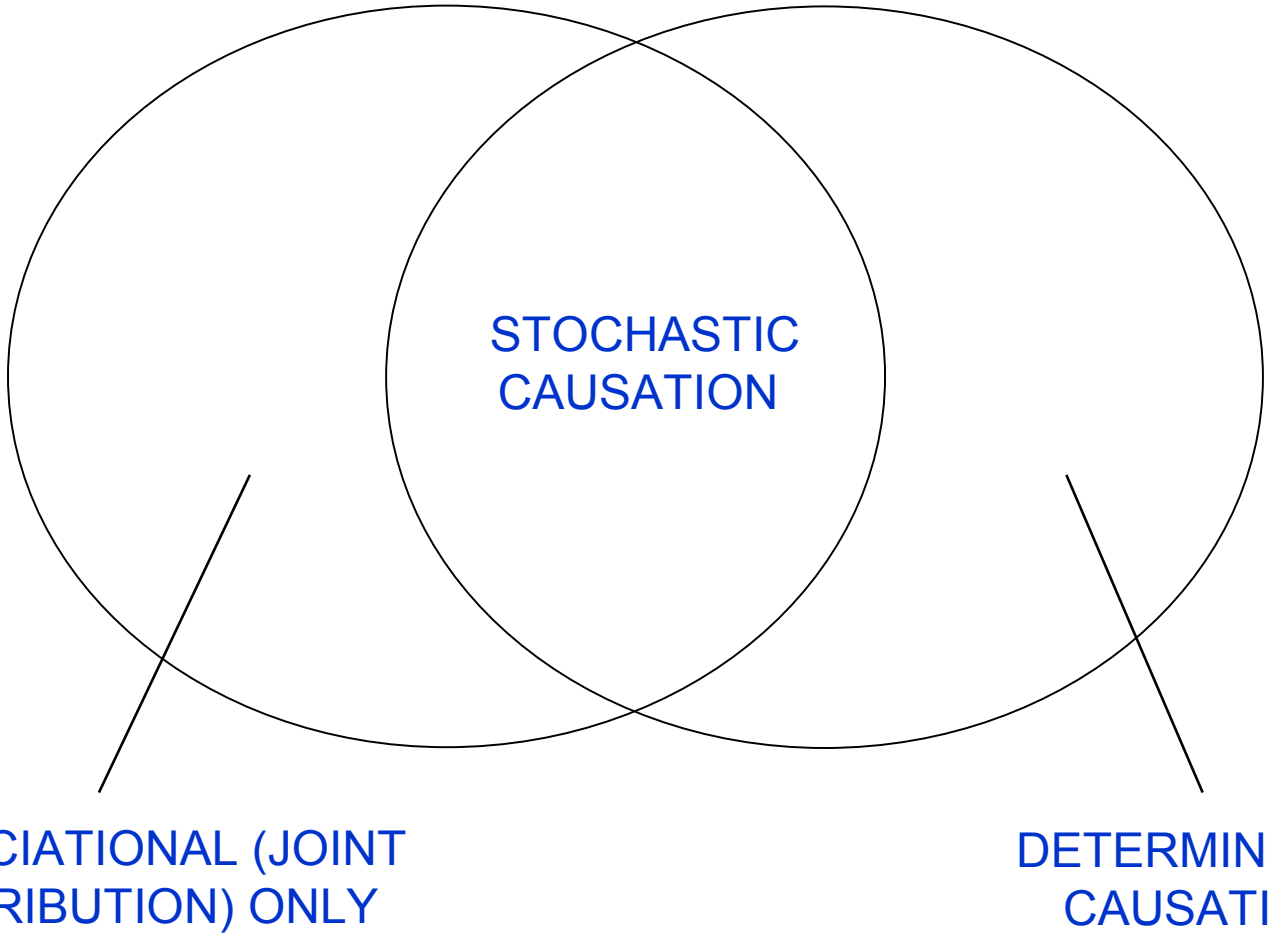
STOCHASTIC CAUSATION



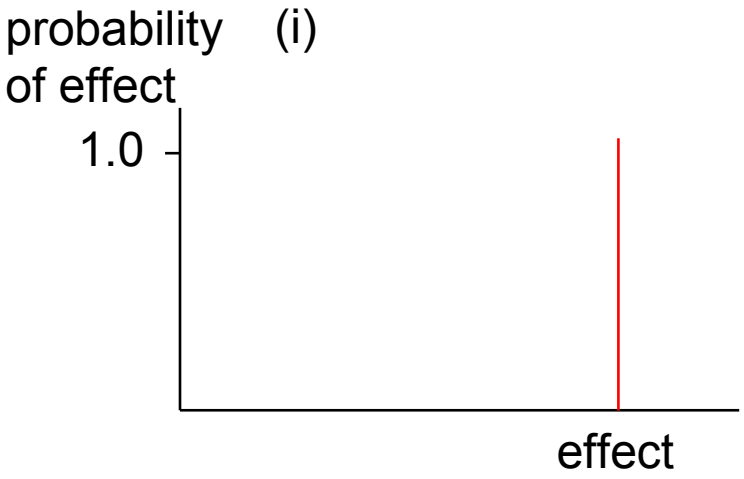
DETERMINISTIC CAUSATION – i.e. causation but not stochastic

e.g. THE NEUROMUSCULAR JUNCTION

STOCHASTIC CAUSATION



TYPES OF STOCHASTIC CAUSATION

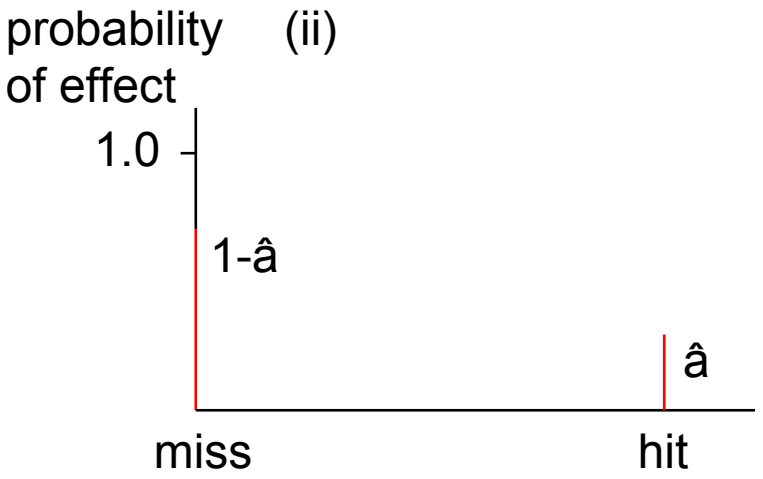


special case: determinism

$y = \alpha x + u,$

i.e. $\hat{a} = 1$

corresponds to neuron firing example



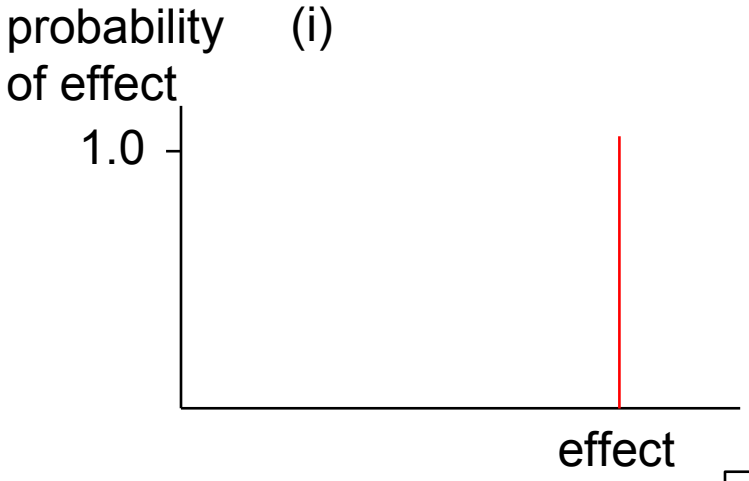
corresponds to radon example

$y = \hat{a}x + u$

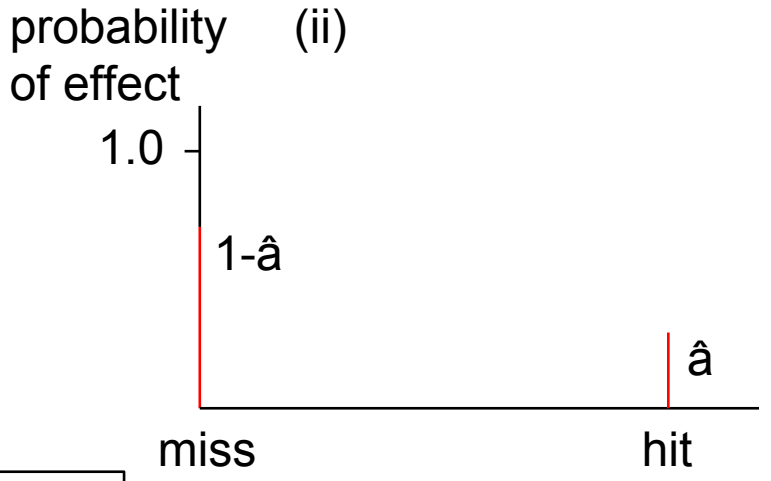
where â represents the probability of a "hit"

(following Cartwright)

TYPES OF STOCHASTIC CAUSATION

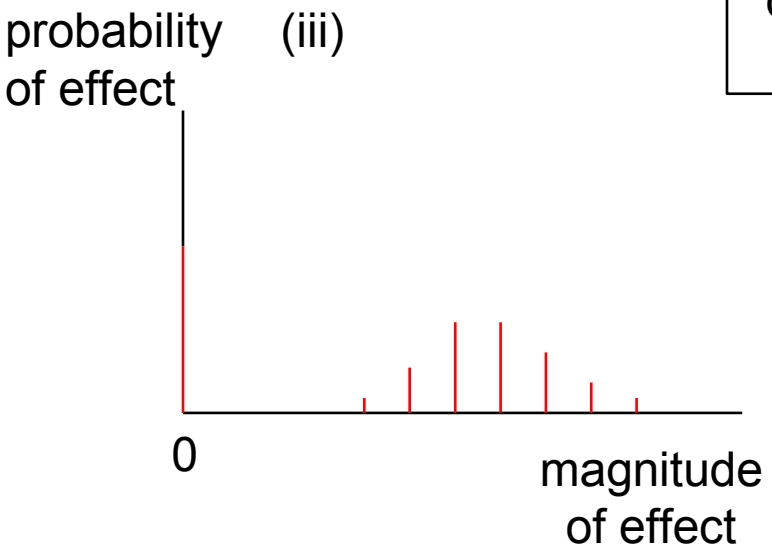


special case: determinism

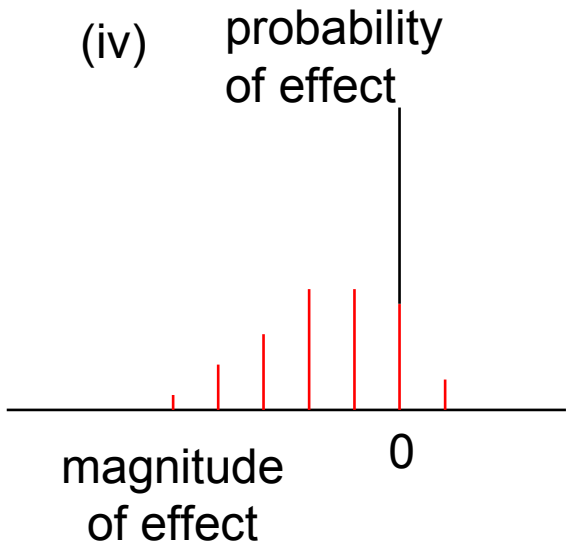


corresponds to radon example

4 different distributions of â

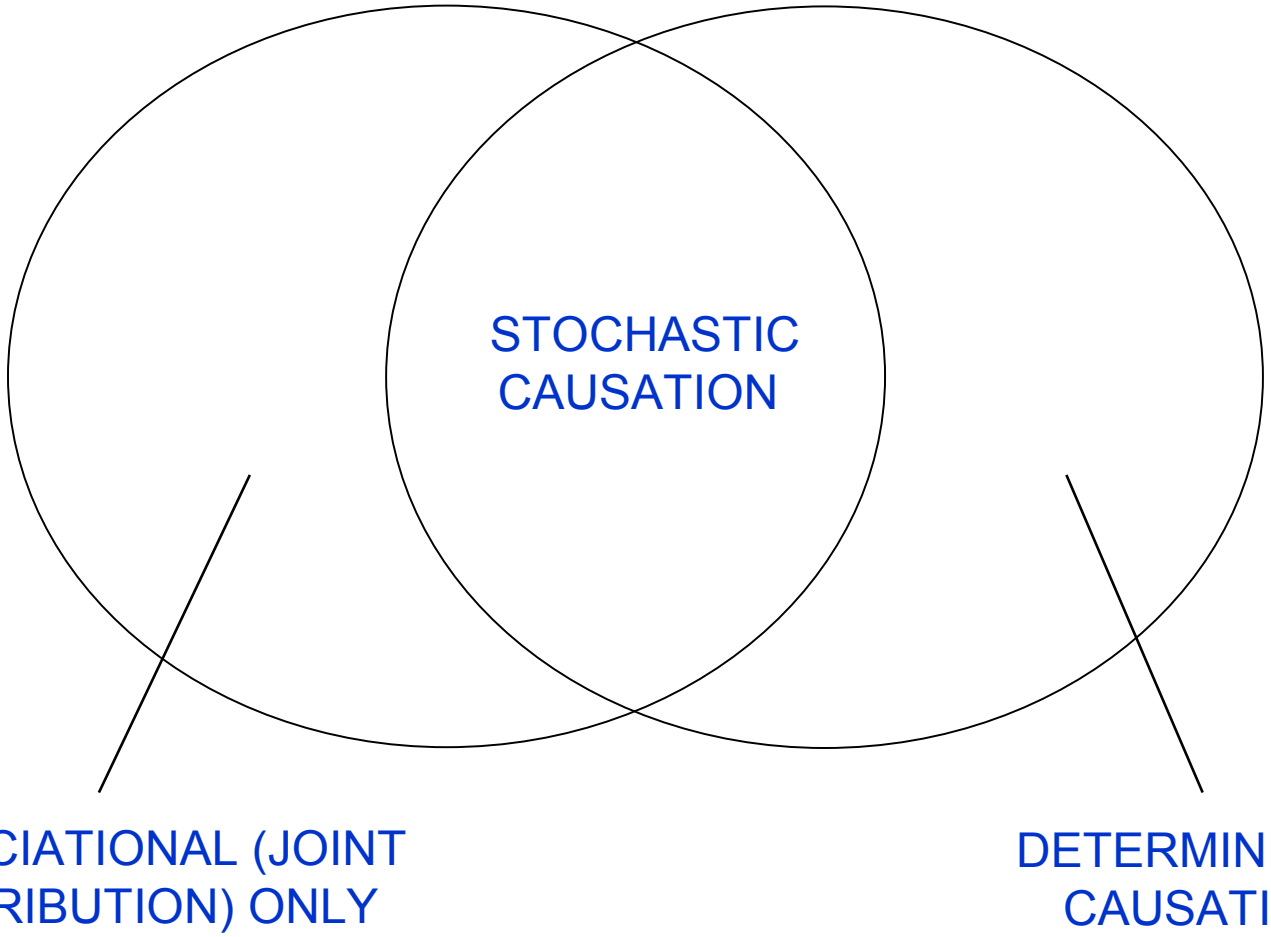


a quantitative case

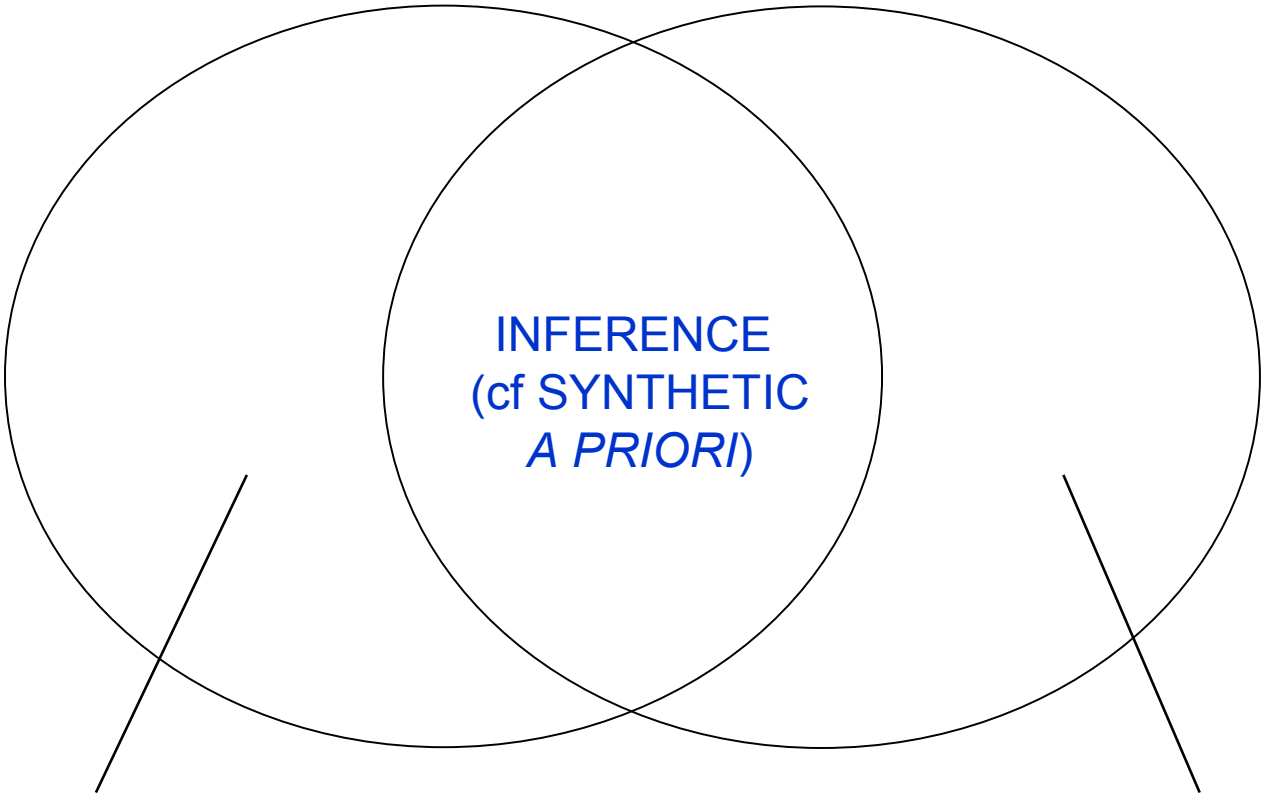


a negative quantitative case

STOCHASTIC CAUSATION



DISCOVERY



CHARACTERISTICS OF THE OBSERVED ASSOCIATIONS (JOINT DISTRIBUTION)

TIME ORDER; DEEP UNDERSTANDING (ACCUMULATED KNOWLEDGE)

Pearl: causal & statistical languages

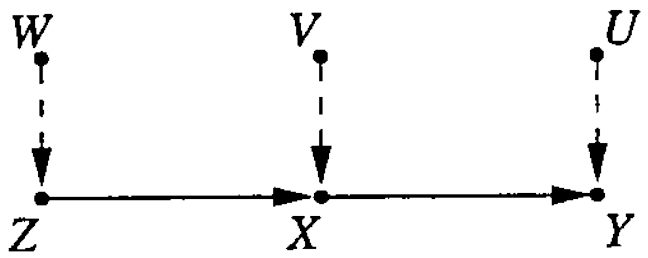
associational concept:
can be defined as a joint distribution of observed variables

- correlation
- regression
- risk ratio
- dependence
- likelihood
- conditionalization
- “controlling for”

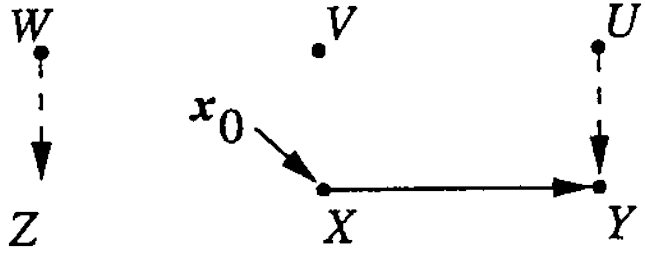
causal concept:

- influence
- effect
- confounding
- explanation
- intervention
- randomization
- instrumental variables
- attribution
- “holding constant”

Pearl: intervention & causation



(a)



(b)

$$z = f_Z(w)$$
$$x = f_X(z, v)$$
$$y = f_Y(x, u)$$

$$z = f_Z(w)$$
$$x = x_0$$
$$y = f_Y(x, u)$$

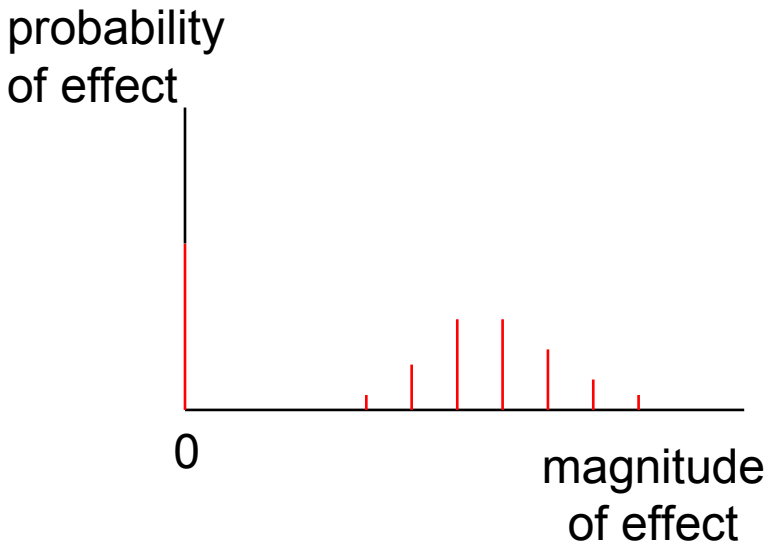
this diagram contains 7 causal assumptions

$$P(z, y|do(x_0))$$

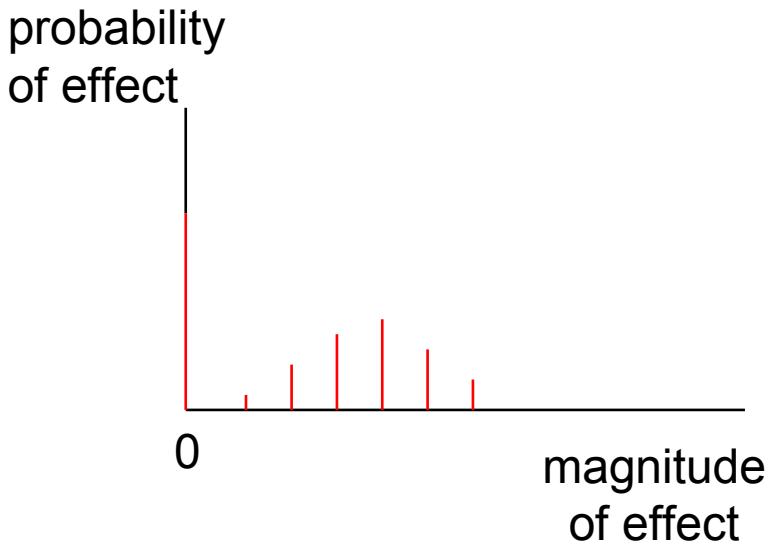
the ***totality*** of these assumptions
can be tested using *d-separation*

CHANGE DUE TO INTERVENTION

Consider example (iii) as now representing the risk and severity of injury or death on a particular stretch of road. The two diagrams show the situation before and after an intervention. (In practice one would have to use a stronger design than an uncontrolled before-after study, and take account of the possibility of regression to the mean.)



original road layout



improved road layout

“Change” models: advantages

- *Parsimony*: the immense complexity of the pathways can be greatly reduced by focusing on changes, especially in the absence of effect modification;
- *Philosophy*: causality is more readily grasped when something is altered, e.g. a particular road layout rather than “roads” as a necessary condition of “road deaths”;
- *Pragmatism*: changes in the determinants of health determinants link naturally to policy options (cf “natural experiments”).

Maths/statistics as a language

- maths may provide an accurate description or prediction of a phenomenon “without passing through the causal process” (Cartwright 1989) – it’s more flexible
- one response could be, constrain the maths to do that, => causal diagrams are primary (graphical models), equations are secondary
- do statistical models represent objective reality (ontology) or subjective beliefs? –
  - if the latter, what language is appropriate for representing ontological statements?

Causal diagrams as a language

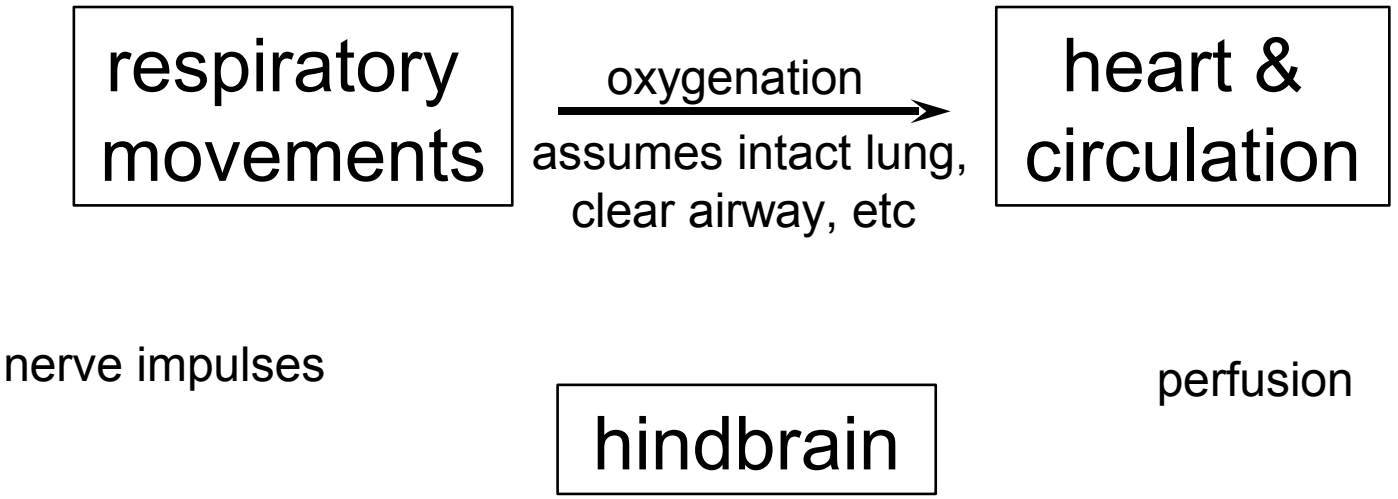
- causal diagrams express time order plus our ideas of *what causes what*, and *how* things happen; they can also include *quantification*, as well as feedback (e.g. self-perpetuation)
- for ontological statements, I suggest that causal diagrams are primary, whereas other “languages” including maths are secondary
- as evidence grows, the status of a diagram evolves from hypothesis to established fact
- we need a better syntax for causal diagrams

THANK YOU!

What is life?

What is life? – short-term aspect

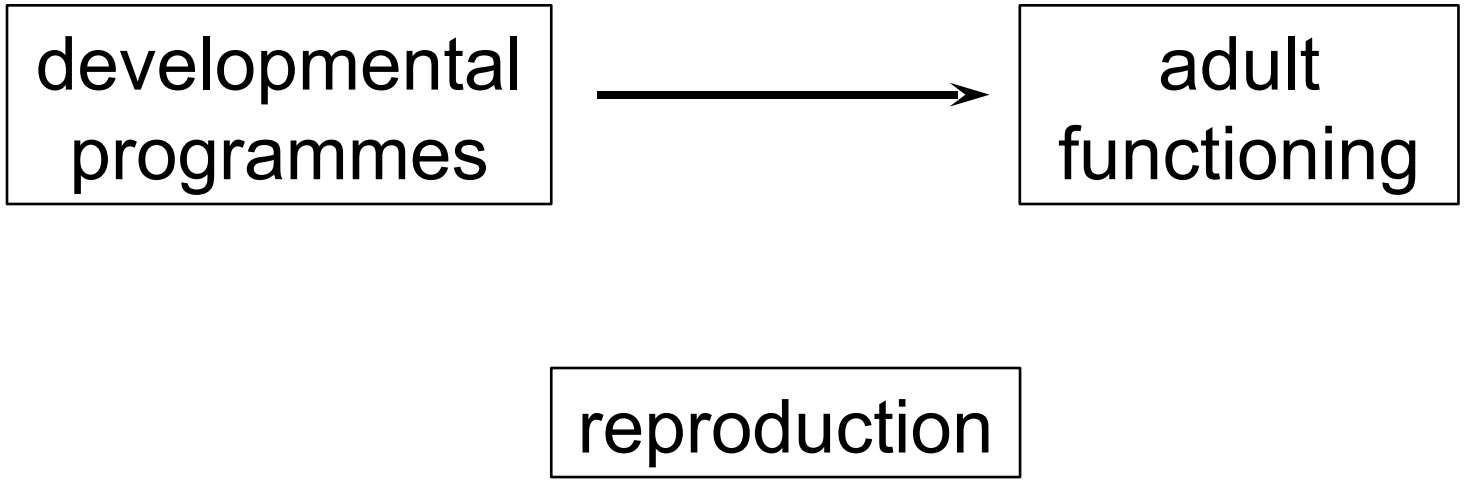
- life is a series of *self-perpetuating systems*
- an *individual* is alive iff a sequence of inter-related causal pathways remains intact – for a mammal the irreducible minimum is:



- this needs amplification with an energy source, and other things with a longer-term perspective

What is life? – longer term

- the systems fulfilling these functions are brought into being by a sequence of developmental causal programmes:



- a *population* survives if enough individuals are alive and able to reproduce at any one time – also a *self-perpetuating system*

Medical diagnosis

- causal inference at the individual level – the particular case (as with legal cases) rather than understanding a general process
- the question is now, “what is the cause of this person’s symptoms/illness?”, rather than “what biological processes cause what range of symptoms?”, or “how does this biological process work?”
- again, the answer is typically in terms of the deeper (physicochemical) level e.g. type I diabetes : lack of insulin, which has known fx

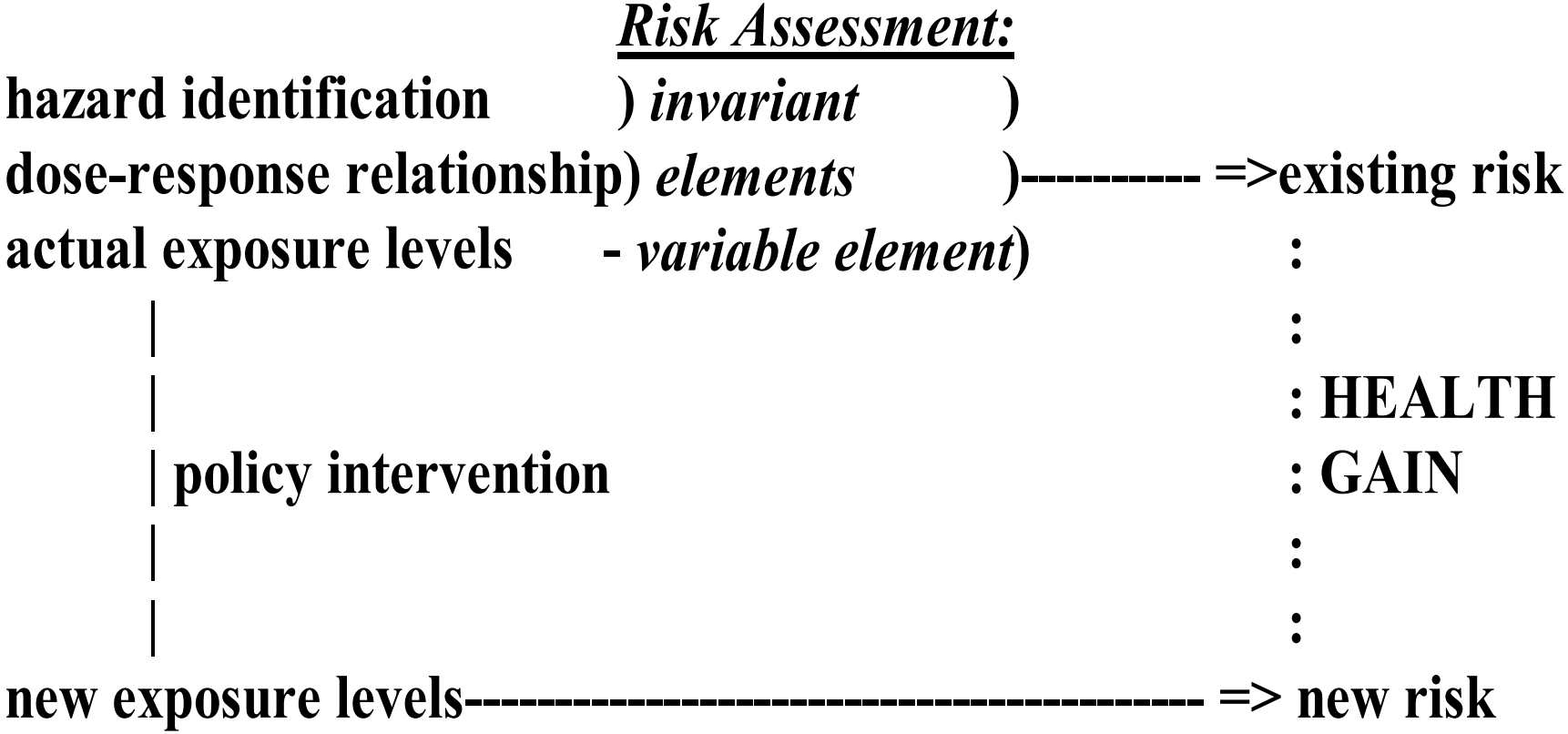
Boundary compatibility

- essential, in explanations involving other disciplines
- emergent properties
  - the higher level must be *compatible* with the deeper level, but the deeper level does not *predict/determine* what happens at a higher level
  - however, it must facilitate it, e.g. a vertebrate or insect circulatory system needs the enzymes, developmental processes, etc, etc
  - the higher-level phenomena come into being through an evolution-type process (or by human agency)

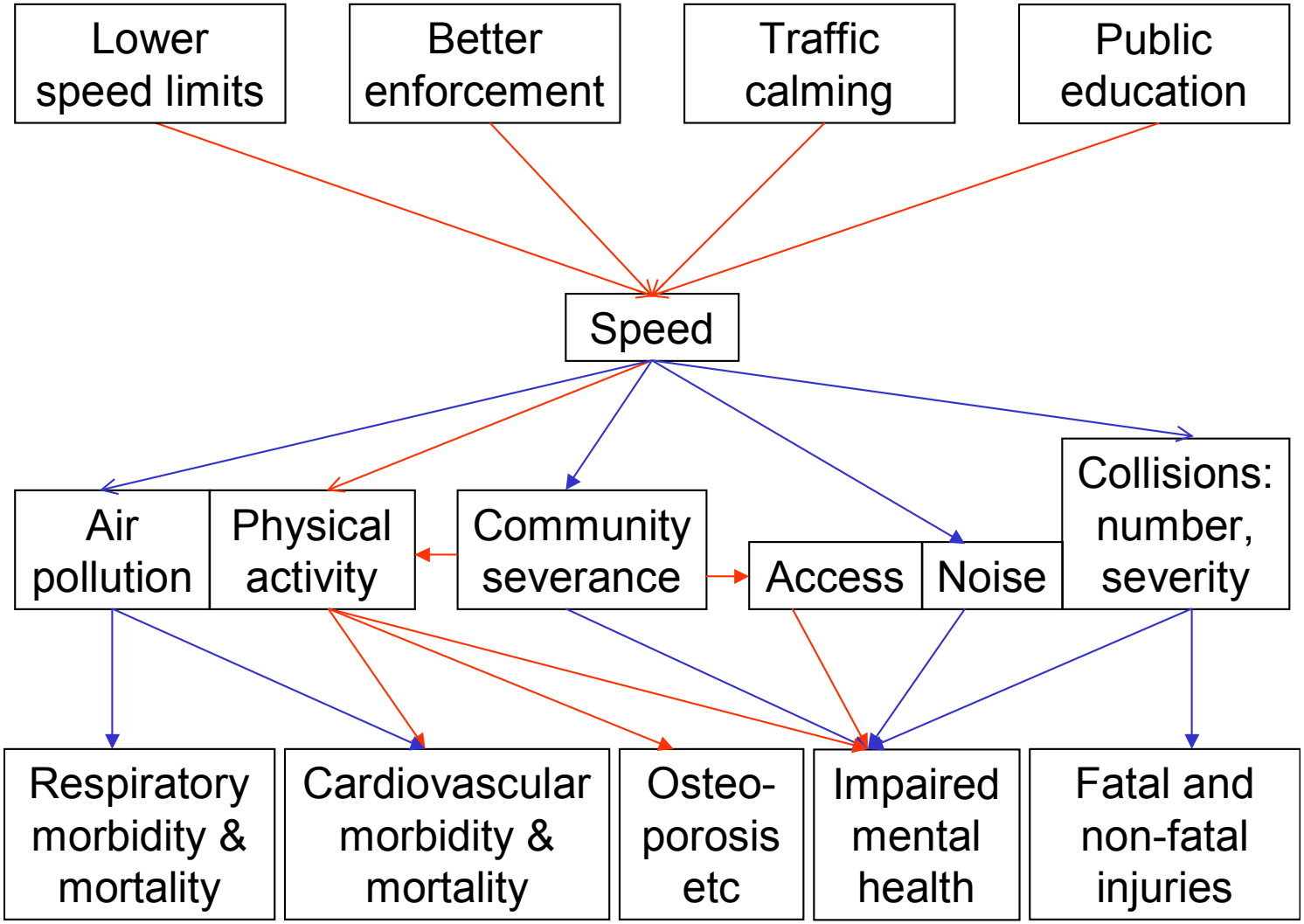
Traffic law enforcement: case-crossover study of road deaths

- fatal crashes identified (‘88-’98), Ontario
- “exposure” is motoring conviction
  - additional information e.g. penalty points
- comparison is a period just before the crash with a period e.g. a year earlier
- 35% reduction in RR of a crash, lasting for a few months
  - especially if penalty points were received
- individual level, each case his/her own control

Policy/risk assessment model (PRAM)



Speed control and health gain



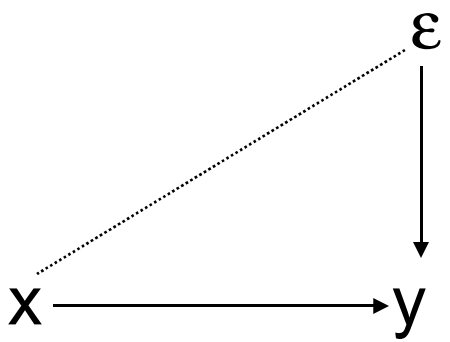
Instrumental variables I

In the classical regression model

$$\mathbf{y = x\beta + \varepsilon}$$

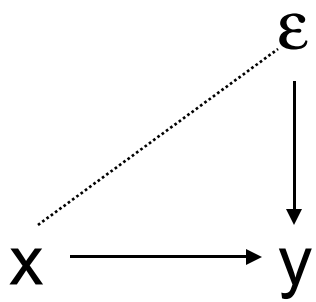
[where x, β can be vectors],

if x is correlated with ε, the OLS estimate of β is biased; in the case of a positive correlation, β is overestimated and its variance is underestimated.

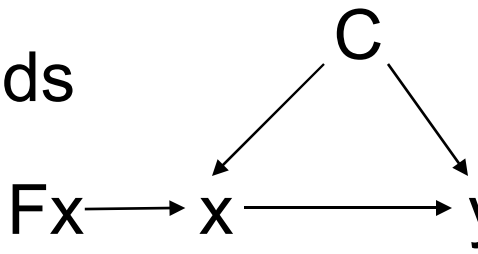


Instrumental variables II

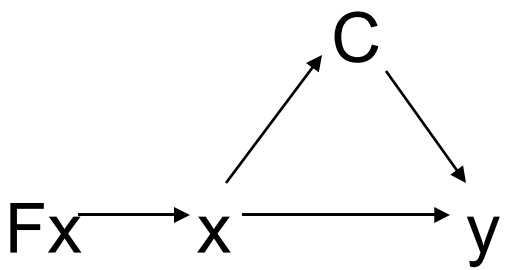
In epidemiology, we usually express the correlation of x and ε in terms of confounding:



corresponds to

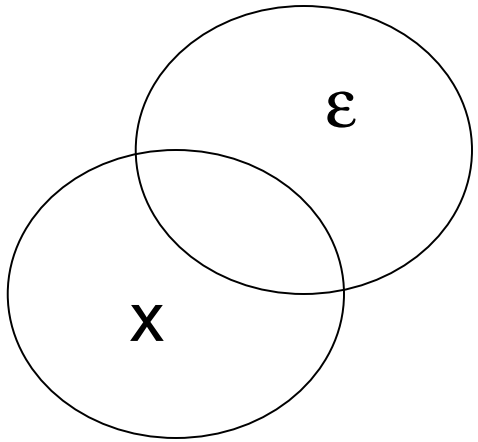


but not to



(Fx => "what would happen if I fix the value of x?"
- Fx is independent of C in the 1st but not the 2nd)

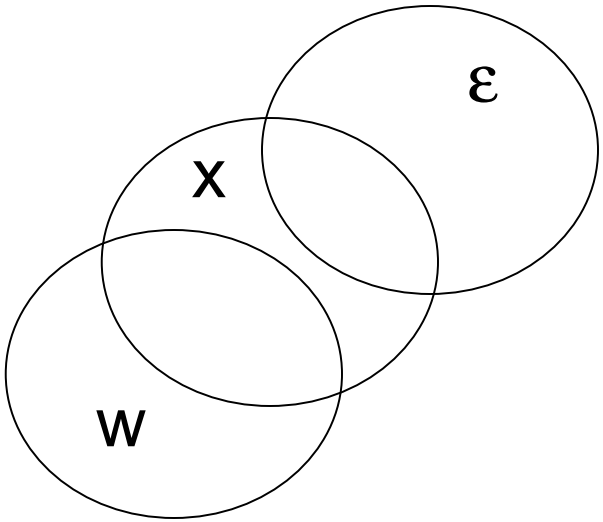
which we can represent by:



Instrumental variables III

If a variable can be found that is **correlated*** with x but is uncorrelated with ε, we have:

** this is not necessarily causal*



The overlap between x and w can be used to provide an estimate of β. In the case of an exactly linear function, this estimate is unbiased; if not, it still provides some information. w is called an *instrumental variable* (IV). As the information set has shrunk to $x \cap w$, the variance of the IV estimator is considerably higher than the OLS estimator, => the larger the overlap, the better.

Instrumental variable