Theory based impact evaluation: advice for getting better at it

Benedict Wauters

21-10-2015

Prague

Evaluation in 2014-2020: Challenges and Opportunities
First annual conference of the National Coordination Authority's Evaluation Unit
Czech Republic

OPERATIONAL PROGRAMME TECHNICAL ASSISTANCE

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

NATIONAL CO-ORDINATION AUTHORITY

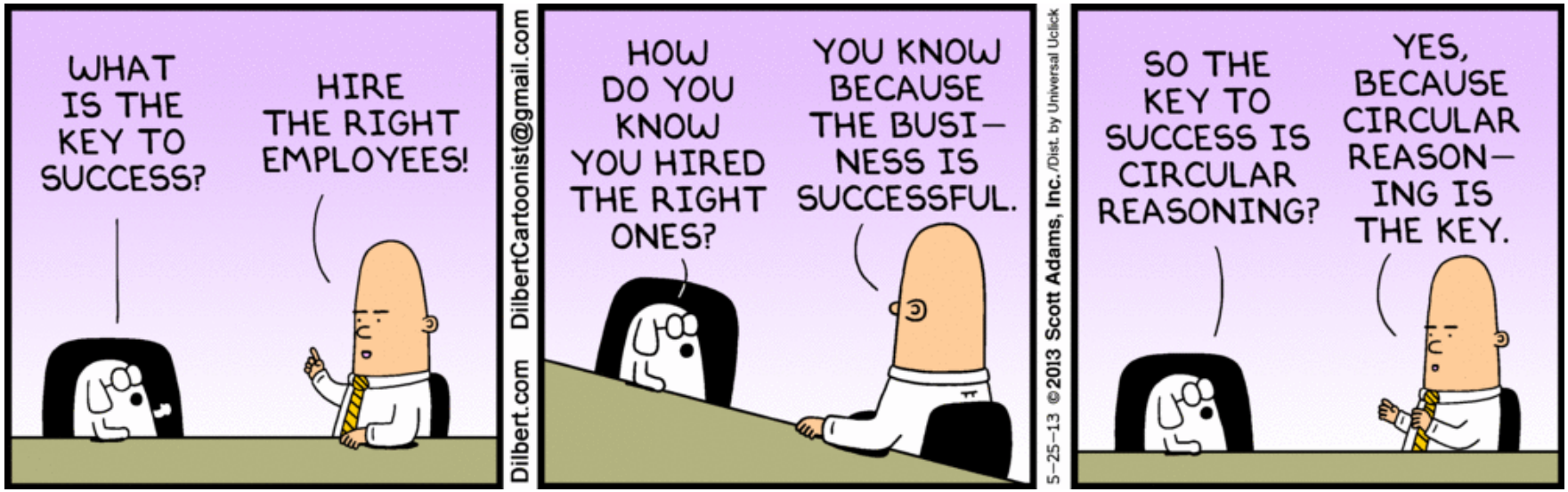
MINISTRY OF REGIONAL DEVELOPMENT CZ

Advice nr 1

- Understand the difference between deductive and inductive inference or between conclusive and exploratory research

Typically…

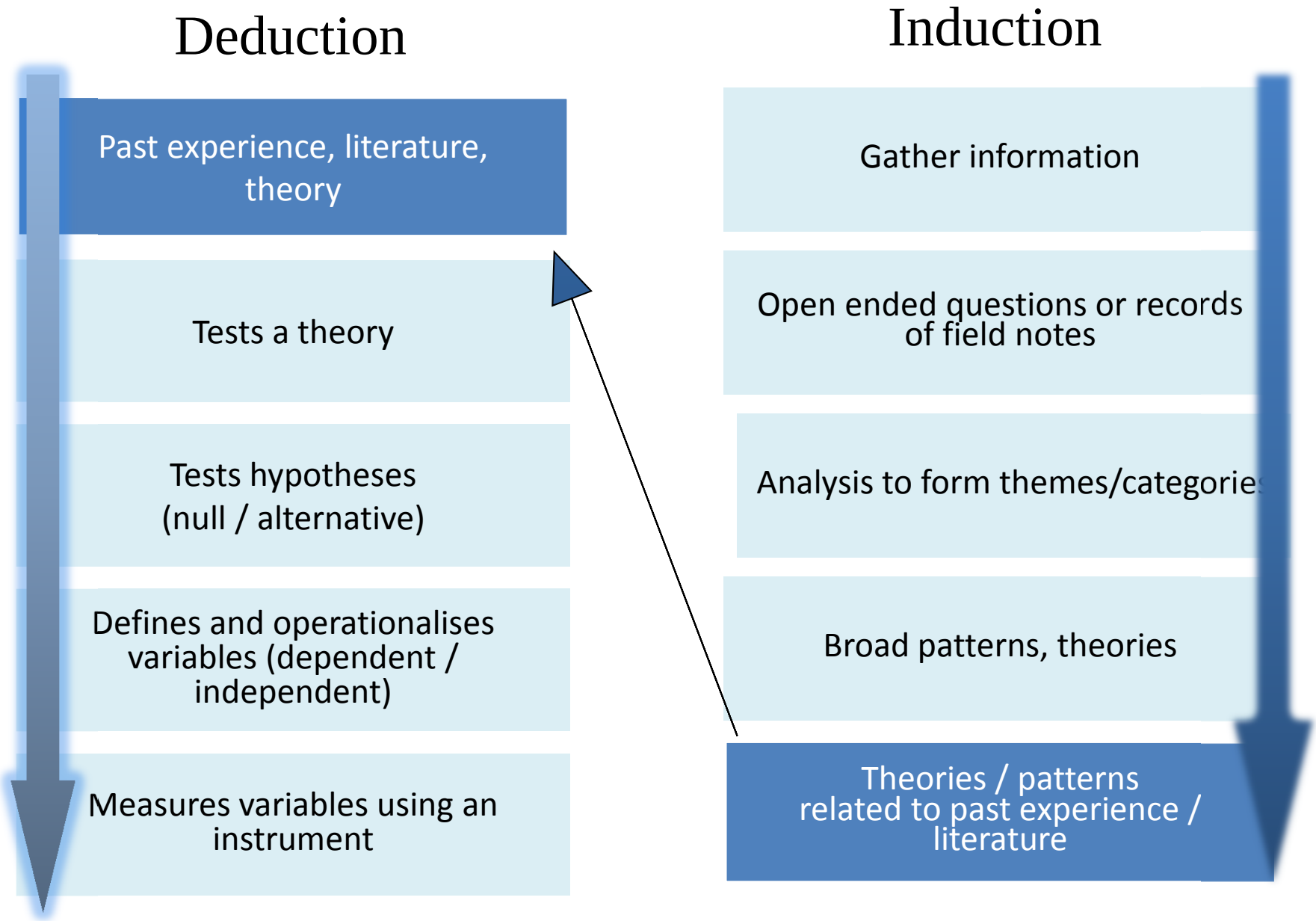
- “Theory based evaluators” go into the field where they do stakeholder interviews, collect some documents, perhaps send out a questionnaire, have a focus group…
- Then they try to make sense of it (“analyse”)…
- …and behold, we have a “theory of change”…
- …of course supported by evidence, hence valid,…
- …right?



In research: you CANNOT validate a hypothesis based on the same data that you used to generate it!

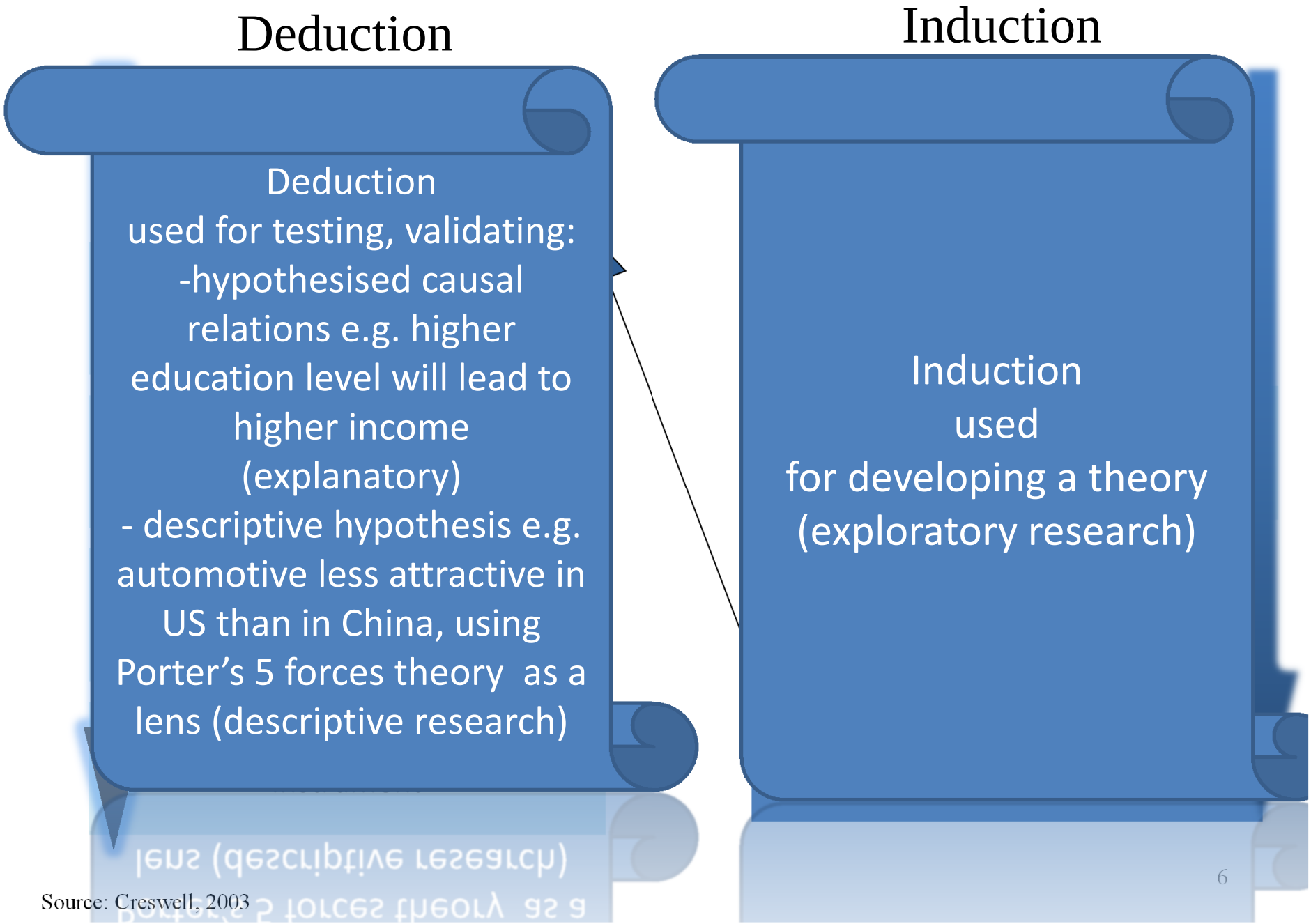
OBVIOUSLY, the same data will hardly contradict the theory…

Two ways to inference



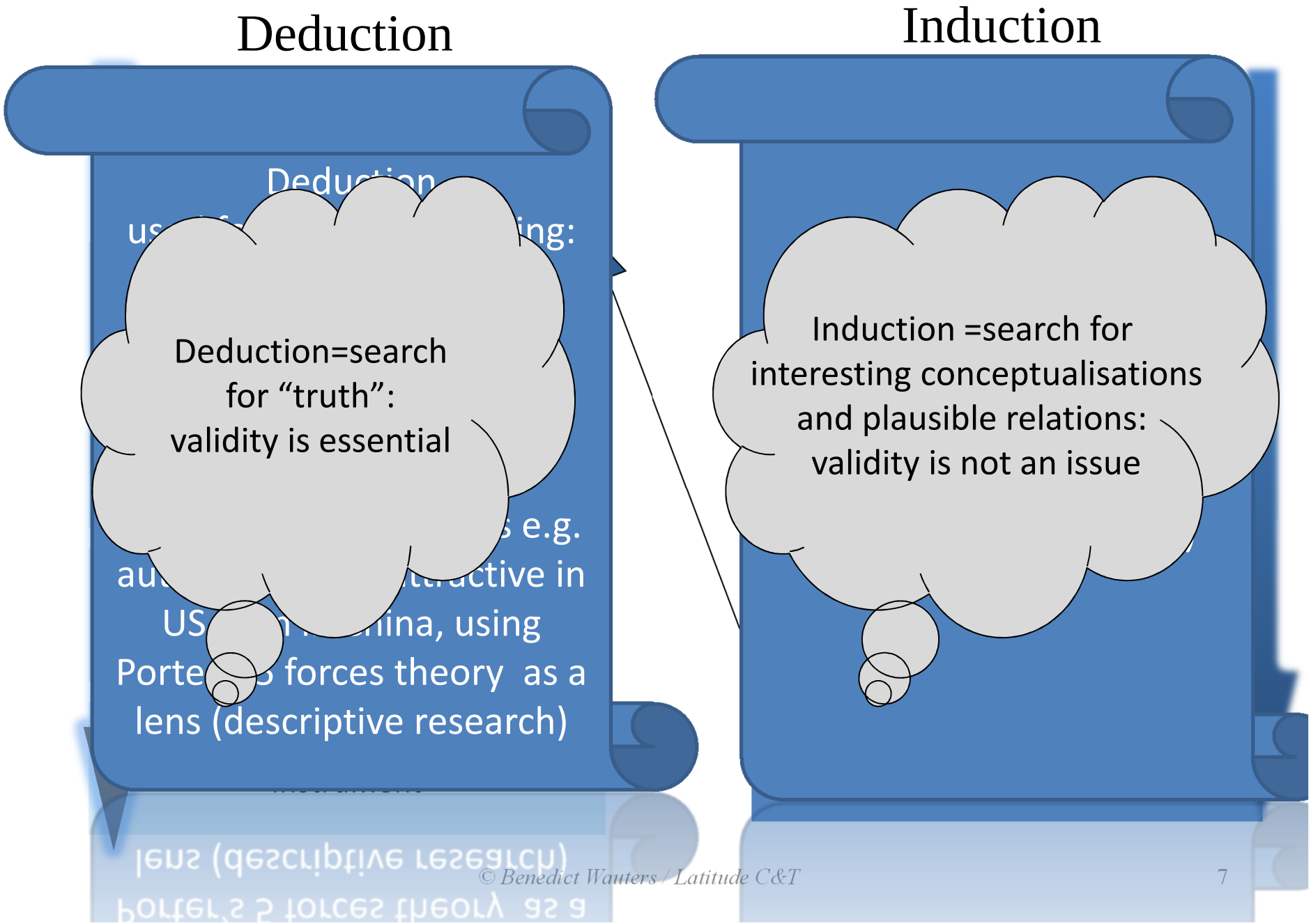
Source: Creswell, 2003

Two ways to inference



Source: Creswell, 2003

Two ways to inference

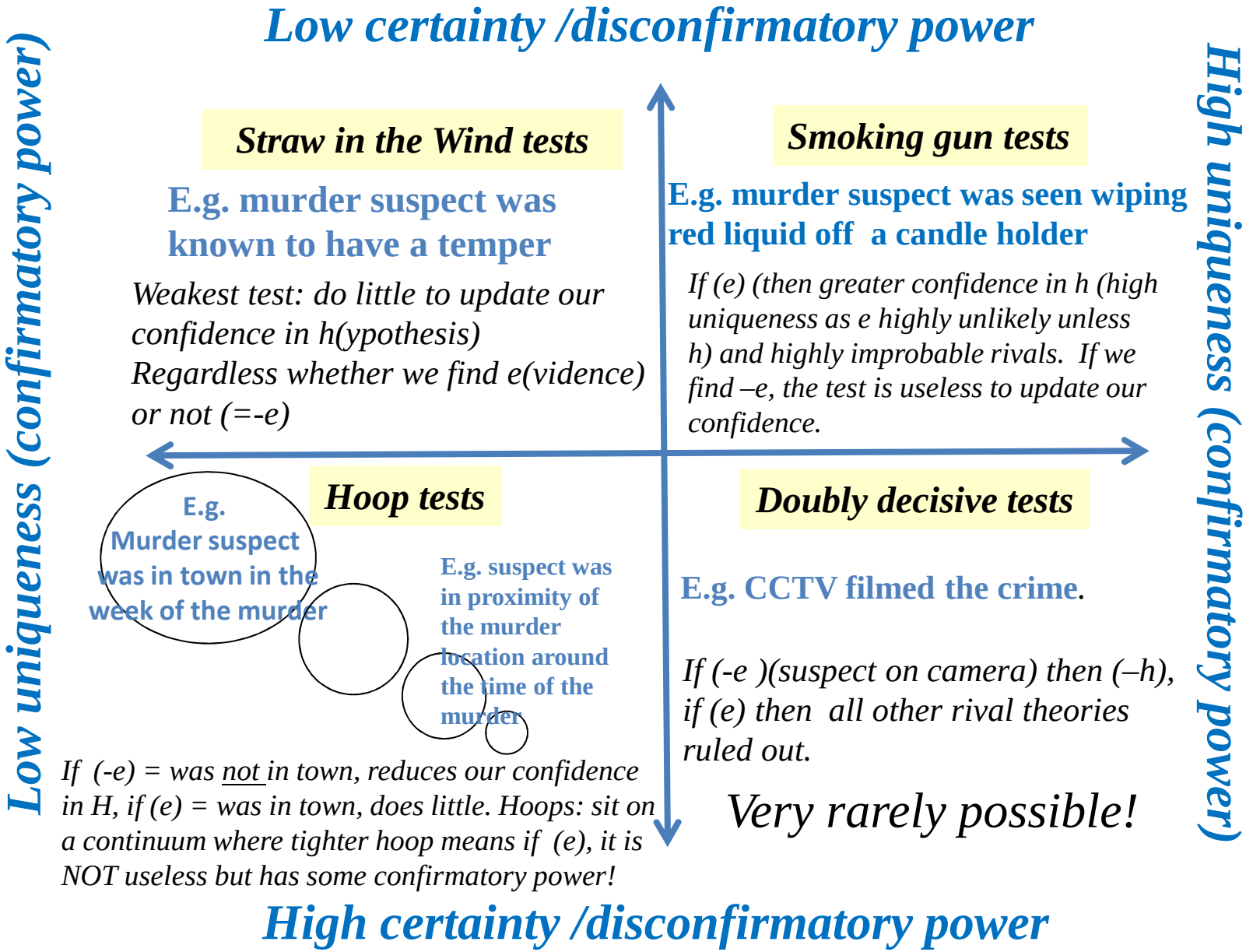


Advice nr 2: learn from Sherlock Holmes

Imagine…

- A murder was committed…
- …someone's head was bashed in with a candle holder…
- …after a bit of "exploration", Sherlock Holmes suspects a fellow called Mr Blow did it!
- Mr Blow is the killer = hypothesis
- Now we have a theory…
- …but how do we prove it?

We need to think about “observable implications” for the theory: what would we have to be able to see in the real world to be able to confirm or disconfirm the hypothesis?



Adapte d from: Beach and Pedersen

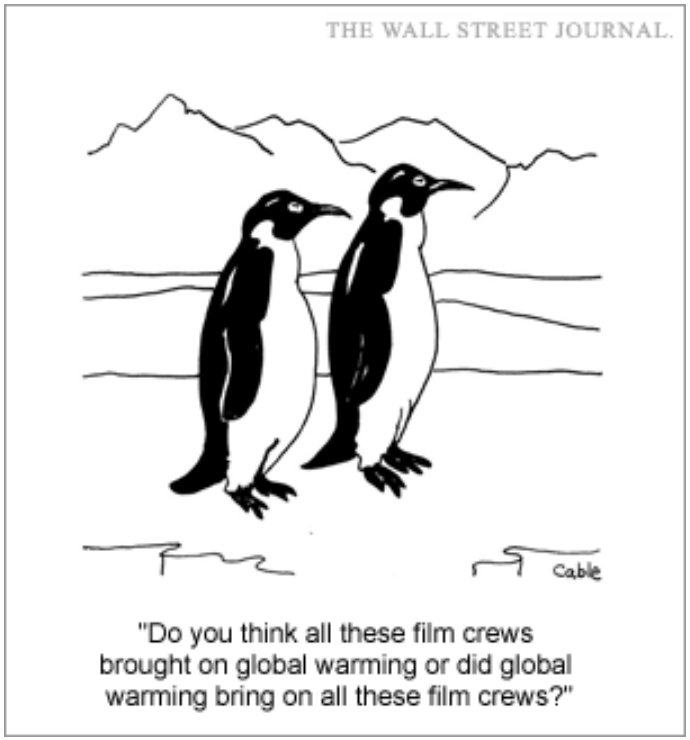
Considered normal in CIE!

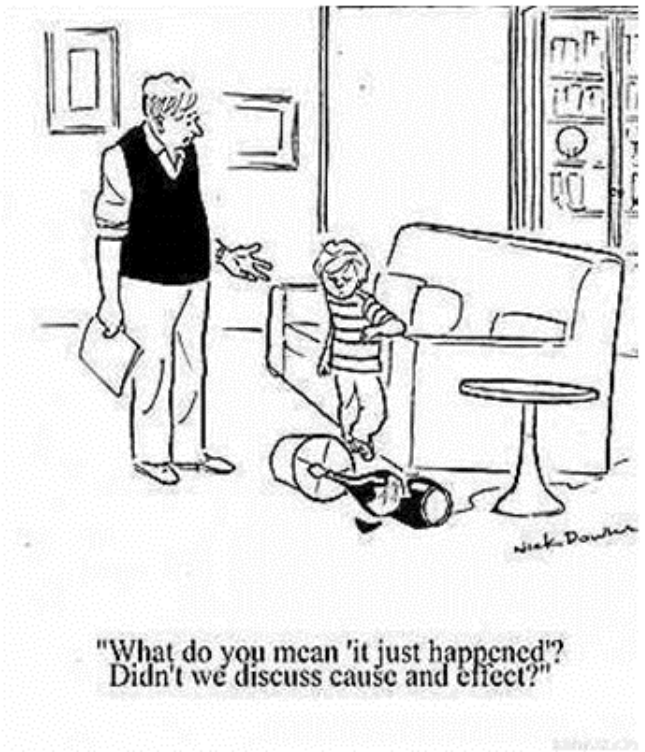
In Reality / What We Conclude	**H_0 (null -no relation- hypothesis) true** **H_1 (alternative hypothesis) false** **In <u>reality</u>...** **•There is *no* relationship** **•There is *no* difference, no gain** **•Our theory is *wrong***	**H_0 (null hypothesis) false** **H_1 (alternative hypothesis) true** **In <u>reality</u>...** **•There *is* a relationship** **•There *is* a difference or gain** **•Our theory is *correct***
We accept the null hypothesis (H_0) **We reject the alternative hypothesis (H_1)** **We <u>say</u>...** **•"There is no relationship"** **•"There is no difference, no gain"** **•"Our theory is wrong"**	**1-α** (e.g., .95) **THE CONFIDENCE LEVEL** The odds of saying there is no relationship, difference, gain, when in fact there is none The odds of correctly not confirming our theory *95 times out of 100 when there is no effect, we'll say there is none*	**β** (e.g., .20) **TYPE II ERROR** The odds of saying there is no relationship, difference, gain, when in fact there is one The odds of not confirming our theory when it's true *20 times out of 100, when there is an effect, we'll say there isn't*
We reject the null hypothesis (H_0) **We accept the alternative hypothesis (H_1)** **We <u>say</u>...** **•"There is a relationship"** **•"There is a difference or gain"** **•"Our theory is correct"**	**α** (e.g., .05) **TYPE I ERROR** **(SIGNIFICANCE LEVEL)** The odds of saying there is an relationship, difference, gain, when in fact there is not. The odds of confirming our theory incorrectly *5 times out of 100, when there is no effect, we'll say there is one* We should keep this small when we can't afford/risk wrongly concluding that there is a relation	**1-β** (e.g., .80) **POWER** The odds of saying that there is an relationship, difference, gain, when in fact there is one. The odds of confirming our theory correctly *80 times out of 100, when there is an effect, we'll say there is* We generally want this to be as large as possible

Trochim, The Research Methods Knowledge Base , 2006

Advice nr 3:

- understand how we can “know” about causality





Epistemology of causality

- David Hume:
  - Contant conjuction (regularity):
    - "Of two events, A and B, we say that A causes B when the two always occur together, that is, are constantly conjoined"
    - Contiguous in time and space; in succession (B always after A); regular (more than once)
    - Not clear how many times is enough!
  - Counterfactual:
    - "We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed"
    - More broadly: what other explanation can there be for the second object, apart from the first one?
  - In CIE: is the observed regularity (treatment-result) explained by the intervention or by selection bias (rival theory)?

Epistemology of causality

- In TBIE regularity + counterfactual =
  - Does theory A provide a better explanation in comparison with B, C…?
  - Causal inference is always RELATIVE, NOT ABSOLUTE
  - Requires fully formed theoretical alternatives, with rich set of distinctive hypotheses

Example: PDP in Flanders

- X= Personal development process (PDP)
  - Supported process to develop employed persons
  - Discussions within the process can consists of two to six conversations between a facilitator and a participant, totalling 2 and a half to 8 hours
  - Not meant to be used to evaluate someone (although info from evaluations can be used in the PDP)
- Y= competency development AND regular engagement in self-reflection
- Will in the long run contribute to being more pro-active in shaping their career and hence more self-reliant in the face of misfortunes such as lay-offs

How does a personal development process with a coach move participants towards taking charge of their career development?

Rational choice propositions :

- Participants in the PDP gain more insight in their own competences;
- They gain more insight in their personal interests and in what they value in work;
- They increase their understanding of possible future career paths;
- Based on the gained insights, participants will make informed choices regarding the development issues they need to address;
- Participants will then draw up action plans that address these identified development issues;
- Participants execute these action plans and acquire or strengthen the necessary competences;
- Participants apply for other jobs or execute their current jobs better
- …

How does a personal development process with a coach move participants towards taking charge of their career development?

"Opportunity space" theory (+/- White's vacancy chain):

- Highly motivated employees (HMEs) in the organization who want to advance their careers already have a good idea of how they want to develop their careers but could not proceed due to a lack of structured opportunities;
- The PDP triggers the organization to set up internal mobility processes;
- HMEs will swiftly volunteer to participate in a PDP to take advantage of this opportunity;
- HMEs will move more rapidly (in the PDP) through the reflection stage regarding what they want and their competences without having to be coached much,
- HMEs will execute their action plans more systematically and faster than other employees;
- HMEs will respond and apply more rapidly for new or vacant positions
- …

Programme - prisoner education (PE)

(T_1) PE might provide qualifications to allow ex-inmates to compete for jobs

(T_2) PE might boost confidence and provide social skills to reduce aggressive outbursts in ex-cons

(T_3) PE might increase cognitive skills and allow ex-prisoners to reason through their difficulties

(T_4) PE might increase presentational and reasoning skills enabling them to become clever criminals

(T_5) PE might provide shelter from violent prison culture for more vulnerable prisoners

etc.

Based on Pawson and Tilley

Programme - prisoner education (PE)

(T_1) PE might provide qu
inmates to comp

(T_2) PE might boo
skills to reduc

(T_3) PE might incr
prisoners to rea

(T_4) PE might i ing
skills en ing them to be riminals

(T_5) PE might provide shelter from violent prison culture for more vulnerable prisoners

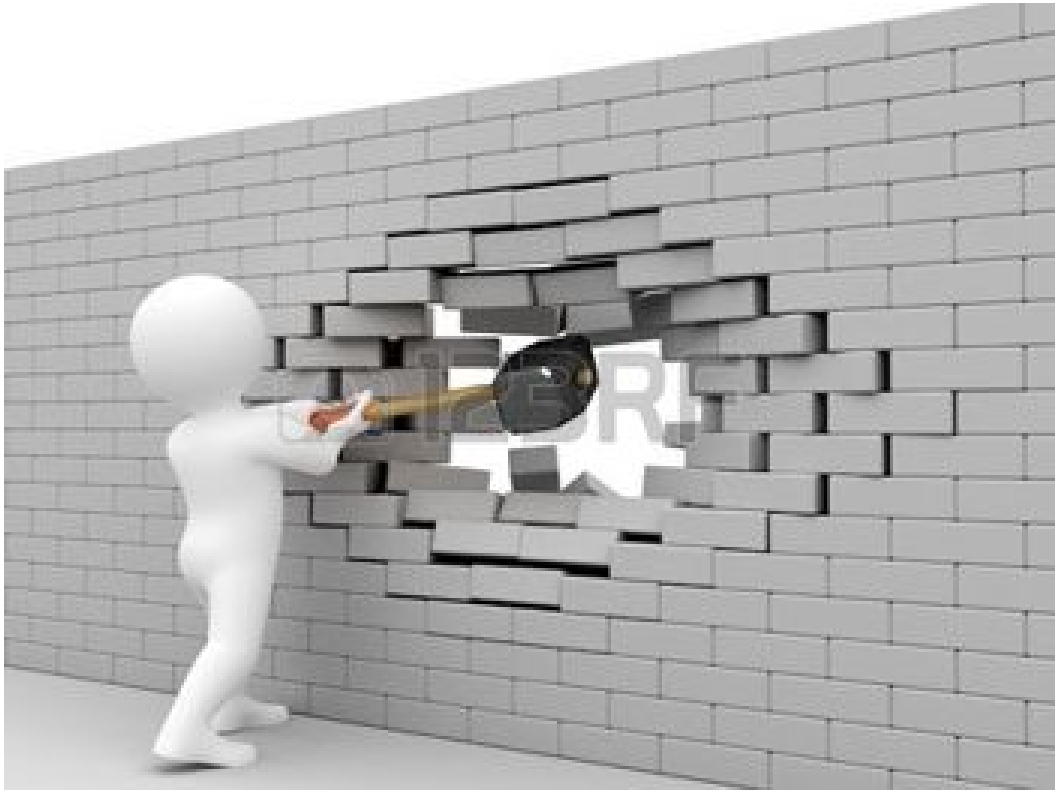
etc.

Most theories are NOT mutually exclusive but can complement and even overlap

Based on Pawson and Tilley

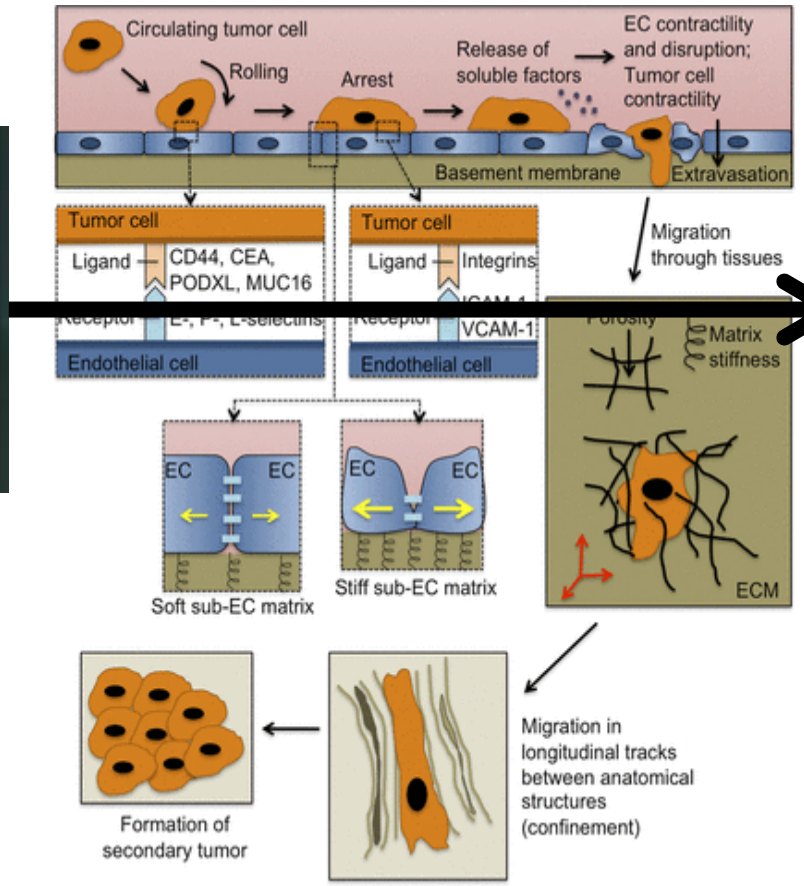
Epistemology of causality

- Roy Bashkar and critical realism:
  - "mechanistic" causality is different from regularity and counterfactuals
  - Can happen only once
  - X only needs to "produce" Y through transmission of causal force

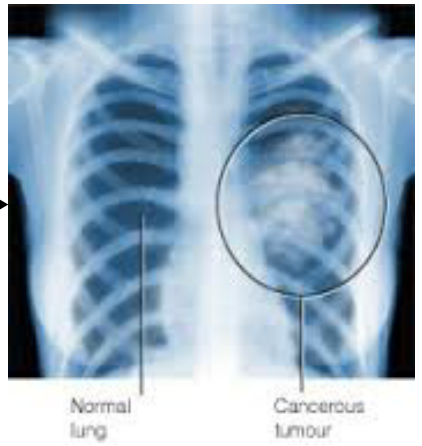


If we can observe an **unbroken chain of action and reaction of agents** we need to observe this only once to know there was cause and effect; also no need to think about “rival theories”

X

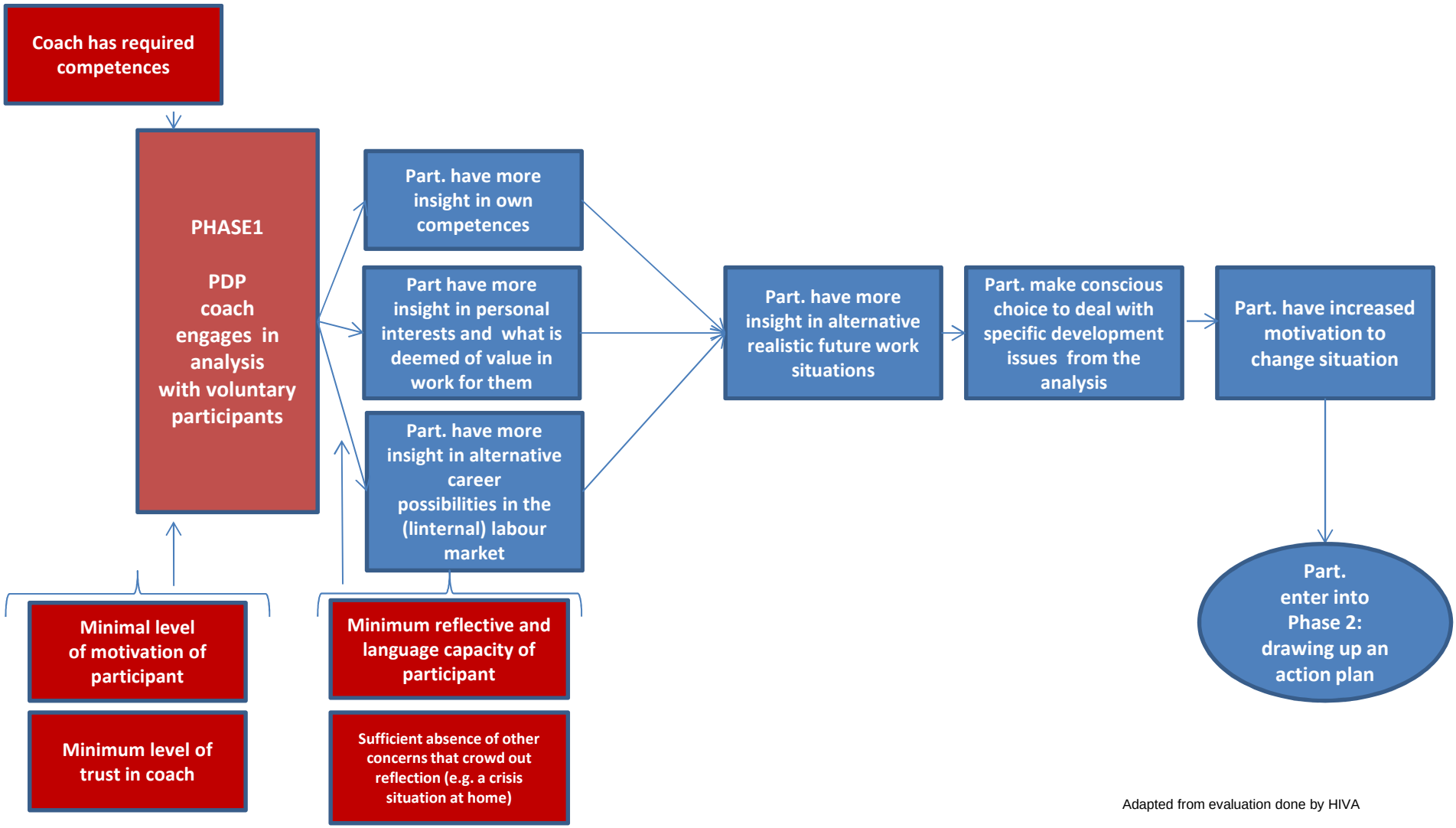


Y



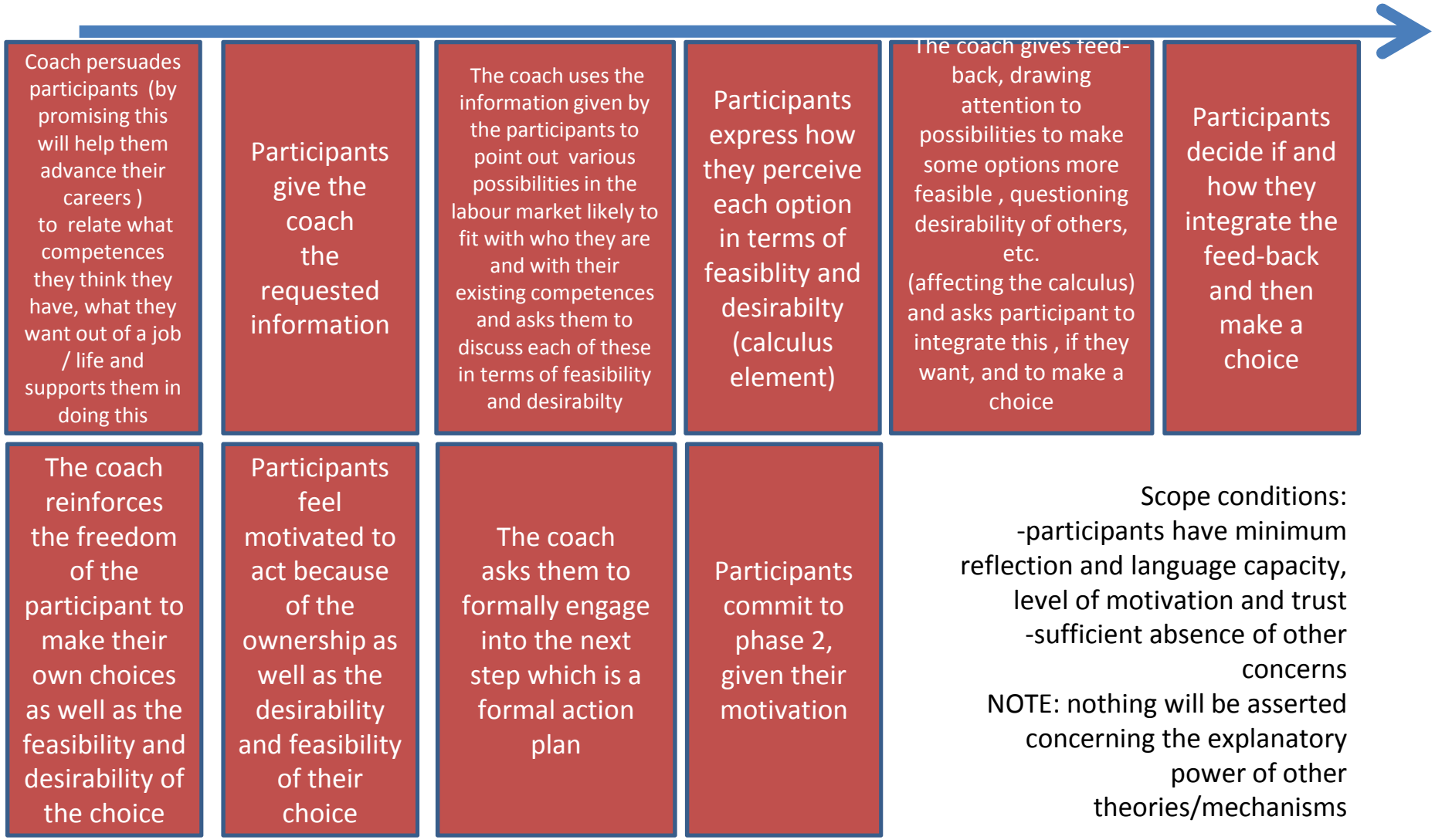
Cecullar biology = mechanism

NOT AN UNBROKEN CHAIN!



Adapted from evaluation done by HIVA

UNBROKEN CHAIN!

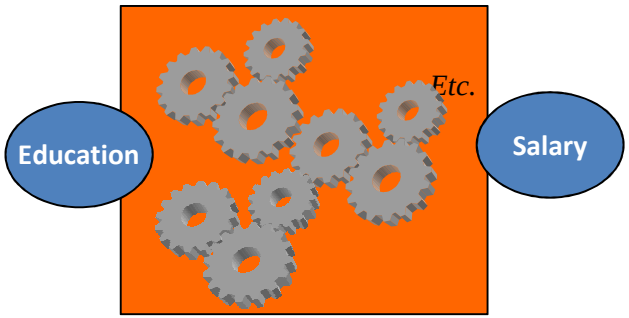


Scope conditions:
-participants have minimum reflection and language capacity, level of motivation and trust
-sufficient absence of other concerns
NOTE: nothing will be asserted concerning the explanatory power of other theories/mechanisms

2 basic **within** case analysis methods

E.g. one companies' PDP process as executed by many participants

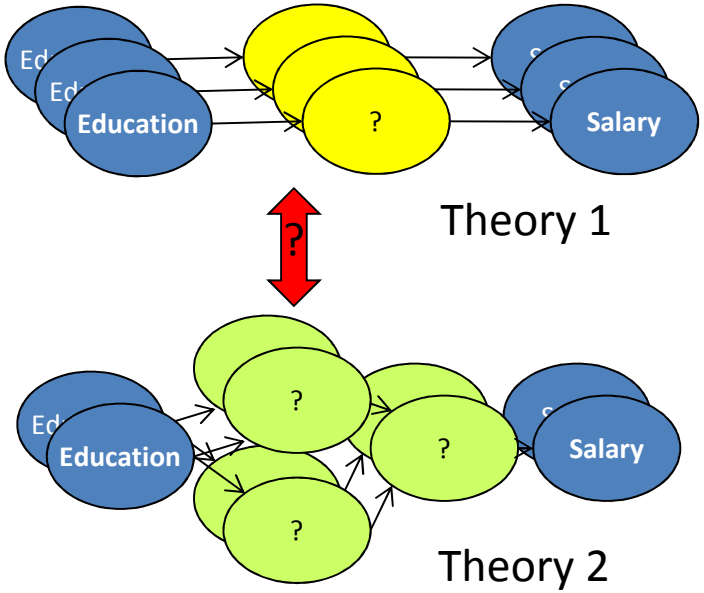
Process tracing / realist evaluation



Casual inference based on unbroken chain of action and reaction between education and salary

≠

Congruence / contribution analysis



Causal inference based on assessing evidence for patterns (regularities) for different theories that account for the influence of education on salary

… there ain’t nothing else…

Advice nr 4:

- Remember that TBIE = case study research

No TBIE makes sense if it is not within a case study framework as we need to understand context!

If cars do not start, does this mean cars are no good to drive us somewhere?

If none of **these** start… we'd have some cause to think so…

...but combustion engines are not supposed to work without air, do they?

If 200 potted plants are randomly assigned to either a treatment group that receives daily water, or to a control that receives none,

and both groups are placed in a dark cupboard,

the treatment group does not have better outcomes than the control.

Possible conclusions: Watering plants is ineffective in making them grow.

Example courtesy of Patrica Rogers

If 200 potted plants are randomly assigned to either a treatment group that receives daily water, or to a control that receives none,



and both groups are placed in a dark cupboard,

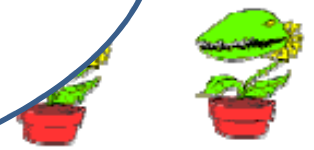

the treatment group does not have better outcomes than the control.

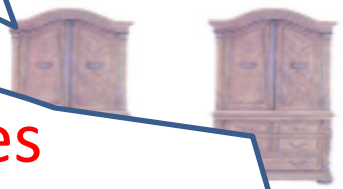

Possible conclusions: Watering plants is ineffective in making them grow.

Better conclusion: Water is not sufficient.

Example courtesy of Patrica Rogers

Context of the intervention matters! Is there also light (condition)? Otherwise water (intervention) will not work.





In a CIE, these contextual differences (more / less light) are randomised out = decontextualisation. The plants with less light are equally distributed in control vs treatment groups = NO selection bias, but smaller effect of water than if all plants had been in good light!

And finally, something to think about…

- Ray Pawson (Realist evaluation) also cautions:
  - “We cannot contemplate, let alone observe and control, every supposition that will find its way into a programme…. enlightenment describes rather well the working relationship between research and policy (slow dawning…)… I think the aim should be to produce a sort of ‘highway code’ to programme building, alerting policy makers and practitioners to the problems that they might expect to confront and some of the safest measures to deal with them. … remember A, beware of B, take care of C, D can result in both E and F, if you try G make sure that H is in place….”

Pawson (2003)

Thank you!

Key challenges for the future

- Building broader and more strategic programmes of intervention research where we expand our knowledge about the underlying mechanisms that CAN generate results for different kinds of citizens in different contexts in different ways rather than looking for the holy grail of universal best practice for all mankind
- Focusing on impact in terms of citizen well-being and human development rather than on an arbitrary “measure” such as “having a job or not 6 months after an intervention”
- Moving away from the current climate of distrust where evaluation is becoming part of the arsenal of audit / control rather than an instrument for learning
- Connected to that, countering the trend of impact evaluation as being perceived as a “technical” and “complex” undertaking best left to independent external evaluators, rather than “impact” and “evidence-based” being part of everyone’s mind-set, from policy-makers to implementers.