KU LEUVEN

On the bumpy road of doing experimental research in accounting

By Prof. Dr. Eddy Cardinaels
EAA PhD forum

Presentation EAA, Valencia

Core concept of experiments

“experimentalists are control freaks”

KU LEUVEN

Life of an experimentalist is difficult

Decision making in real life is complex !

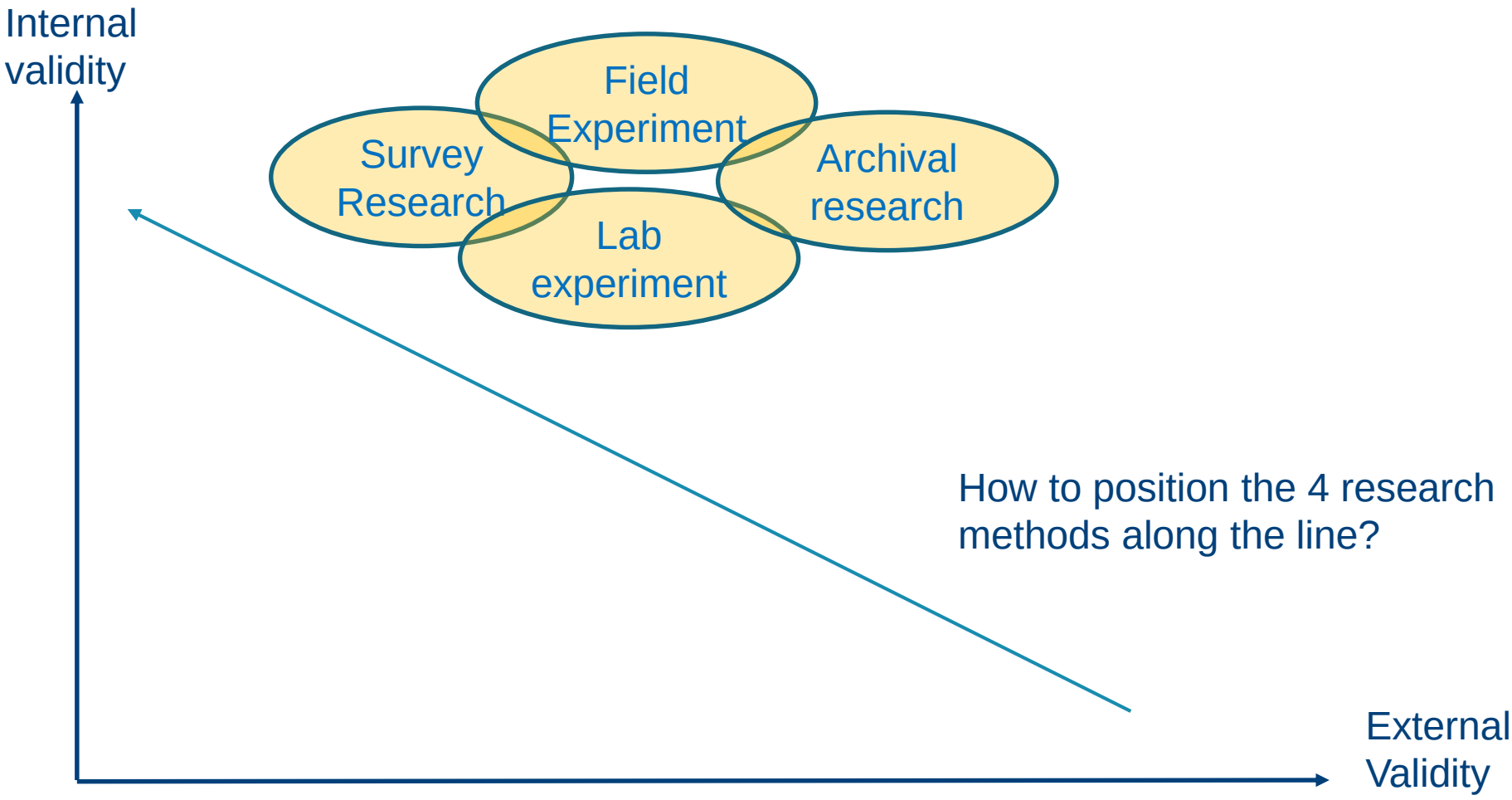
The environment is changing on many dimensions

Can we be sure that the results be driven by the process you specify? Could your results not be driven by … ?

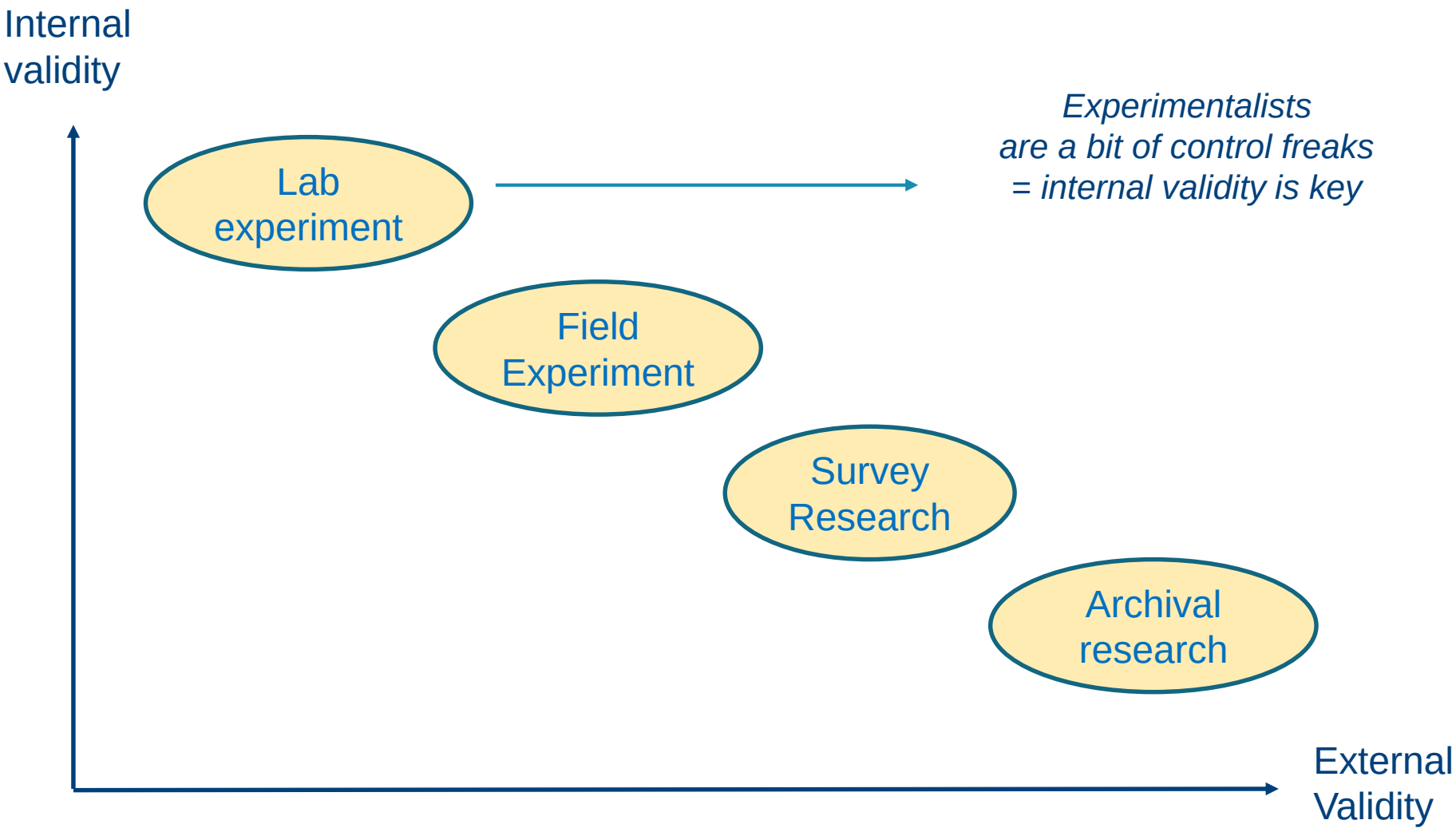
External validity
= The extent to which results can be generalized beyond specific tasks, measurement methods and participants employed in the study

Internal validity
= The extent to which we can accurately Infer that the independent and dependent variables are causally related

Life of an experimentalist is difficult



Life of an experimentalist is difficult



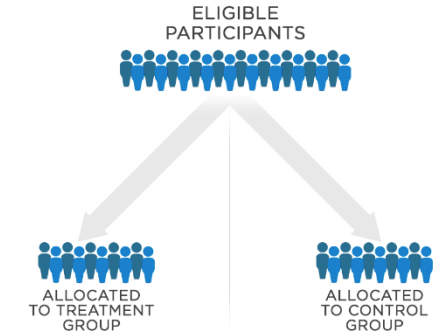
Experimentalists are a bit of control freaks = internal validity is key

KU LEUVEN

Internal validity is our reason of existence

Experiments: The best method to establish causality

Y= TREATMENT + [Assume all other things held constant]



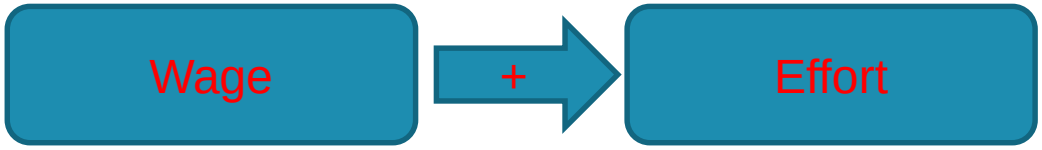
Incentive

Fixed wage

Internal validity is key: Construct vs. Operationalization = closely aligned

Internal validity is our reason of existence

Falk and Heckman (2009): Gift exchange hypothesis.



Hard in real life:

- ✓ workers effort is not always observable or measurable
- ✓ Workers have other incentives
- ✓ Firm size, productivity differences, self-selection of employees into firms
- ✓ In reality, there are reputational concerns too.

Better in experiments: causality between wage offer and effort provision

- ✓ People in roles of principal (firm) and workers → via random assignment
- ✓ Principal makes (binding) wage offers that workers can accept
- ✓ If workers accept, they decide about 'costly' effort provision

Internal validity threats

… REALITY BITES …
It is not always that easy

Do we measure the right construct?

- *Construct validity is "the degree to which a test measures what it claims, or purports, to be measuring."*
- *Manipulations should be as clean as possible !*
- *Theoretical construct = Captured in the operationalization*
- *Causality: An effect IV should be causally related to the DV*

Internal validity threats

Banker et al. (2004, The Accounting Review):

Strategy map, next to a BSC present (treatment) vs. absent (control)

Idea: Strategy map enhances strategy comprehension, which in turn affects performance evaluations; people do not forget about strategy linked measures in a BSC

Manipulation successful or not?

Internal validity threats

Panel A: Strategy Comprehension			
1. The strategy of The Women's Store is to generate greater sales through its *existing* infrastructure rather than invest in new stores.	3.23	3.73	−0.50**
2. To grow sales, The Women's store must successfully introduce new lines of clothing to its existing customers.	2.48	3.38	−0.90**
3. The strategy of The Family Store is to grow by adding new stores.	2.99	3.54	−0.55**
4. The Family Store needs an innovative marketing group because its growth plans depend upon the success of its advertising campaign in attracting new customers.	2.96	3.67	−0.71**
Strategy comprehension scale[d]	2.91	3.58	−0.67***
Panel B: Understanding of Task			
1. The performance measures were usefully categorized in this case.	2.64	2.80	−0.16
2. The emphasis on financial measures was appropriate.	1.94	2.21	−0.27
3. The two business units, The Women's Store and The Family Store, used some different performance measures.	3.15	3.50	−0.35*
4. It was appropriate for the two business units, The Women's Store and The Family Store, to employ the different performance measures.	3.33	3.90	−0.57**
5. The case was easy to understand.	2.40	2.90	−0.50**
6. The case was difficult to do.	−1.02	−1.04	0.02
7. The case was realistic.	2.68	3.21	−0.53**

→ 1. MANIPULATION PICKS UP NOISE

KU LEUVEN

Internal validity threats

Panel B: VBC System

	GRP	TRN		
	PLANT	BLUE	RED	YELLOW
# of Subassemblies Completed	15	6	4	5
× $1.00 = **TOTAL REVENUES**	$15.00	$6.00	$4.00	$5.00
COSTS:				
× $ .70 = STANDARD COSTS	$10.50	$4.20	$2.80	$3.50
Difference Between Actual and Standard	$1.50	$(.20)	$1.20	$.50
TOTAL COSTS	$12.00	$4.00	$4.00	$4.00
NET PROFIT	$3.00	$2.00	$0.00	$1.00
Divided by Three = **Compensation per Person**	$1.00			

Under group incentive workers were shown first column

Under TRN incentive workers were shown column two, three and four

Drake et al. (1999):

Does payment under TRN vs. GRP incentives make a difference?

2. MANIPULATION PICKS UP DIFFERENT THEORETICAL CONSTRUCT

- Not only rewards differ
- Information set also differs !!

Better to show all columns and Reward differently ?

Internal validity threats

3. ESTABLISHING CAUSATION: PEOPLE DO NOT RUN SAME HORSE RACE

Kachelmeier and Towry (2002): TP& Fairness based concessions

- **FN 6:** Computerized experiment (footnote six) required additional training phase with bid ask software; face-to-face negotiation did not receive such training.
- Computer = Bid-ask acceptances vs. face-to-face = negotiation?

Try to avoid manipulating different constructs in one manipulation

4. DEMAND EFFECTS: Experimenter gets desired behavior (Zizzo, 2008, ssrn)

Binmore et al.'s (1985) more self-interest than in Guth et al. (1982):
Instructions "How do we want you to play? YOU WILL BE DOING US A FAVOUR IF YOU SIMPLY SET OUT TO MAXIMIZE YOUR WINNINGS."

Internal validity: Solutions

Y= Treatment + Assume all other things constant

↓

Manipulations (PEQ): avoid alternative interpretations

→ Manipulations are often extreme polarizations of reality: absent / present
see Kachelmeier & Towry, (2002): often needed to test theory

→ Effect size; profit impact: lab experiments are less well suited
depends on parameters used in experiment

EXPERINENTAL INSTRUCTIONS should be carefully designed

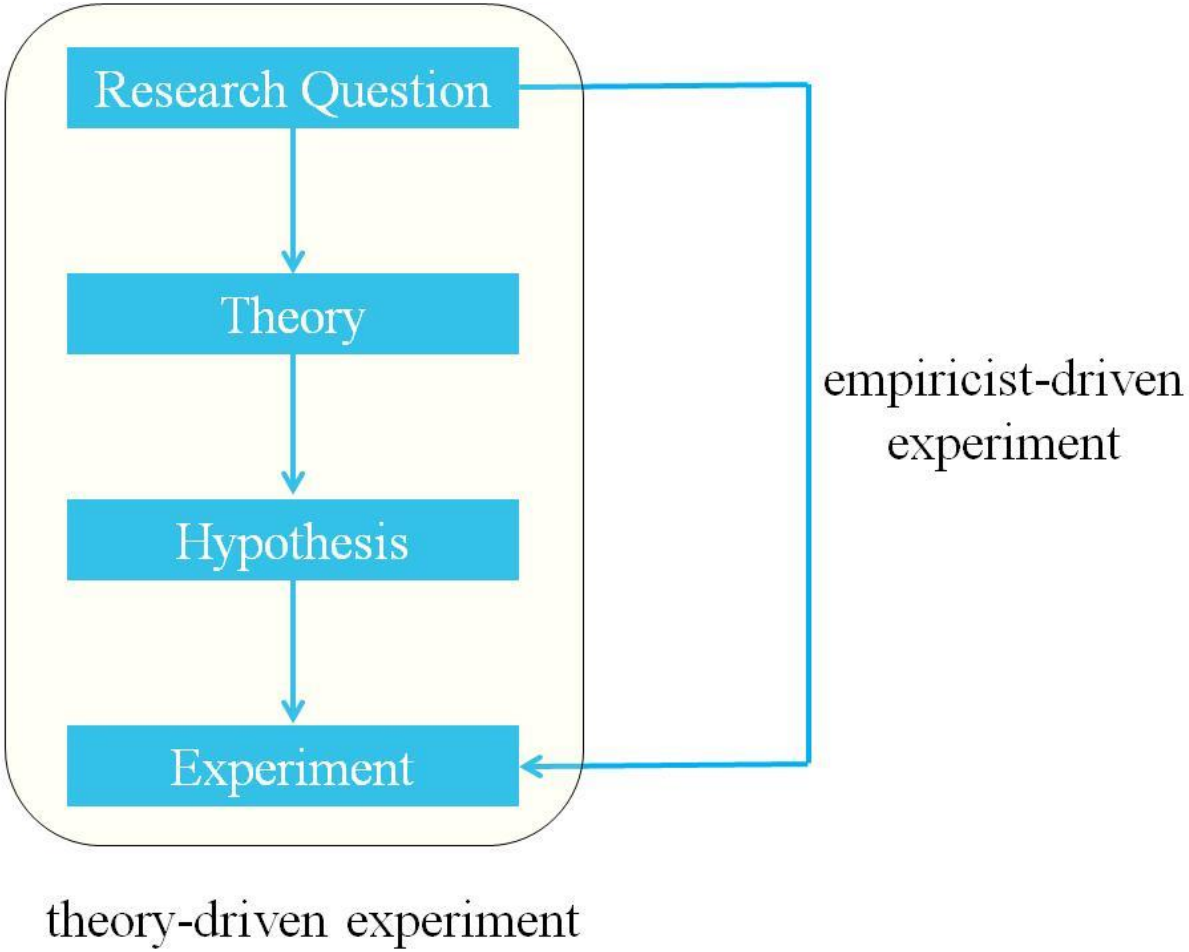
Measure co-variates & mediators: Try to rely on validated scales

The Drawing Table
Design of Experiment

“Theory matters; and complexity should be avoided”

KU LEUVEN

The experimentalist has a feel for theory



The experimentalist has a feel for theory

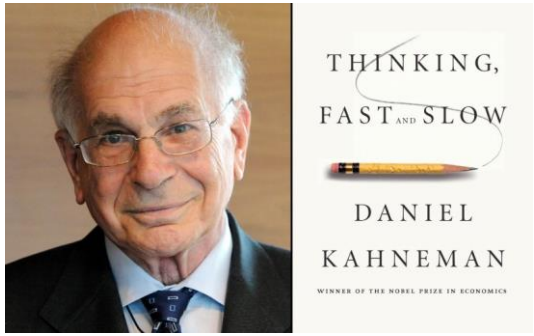
All too often we see experiments defined from practice.
RQ: ABC, assuming better information, leads to higher profits

Problem:

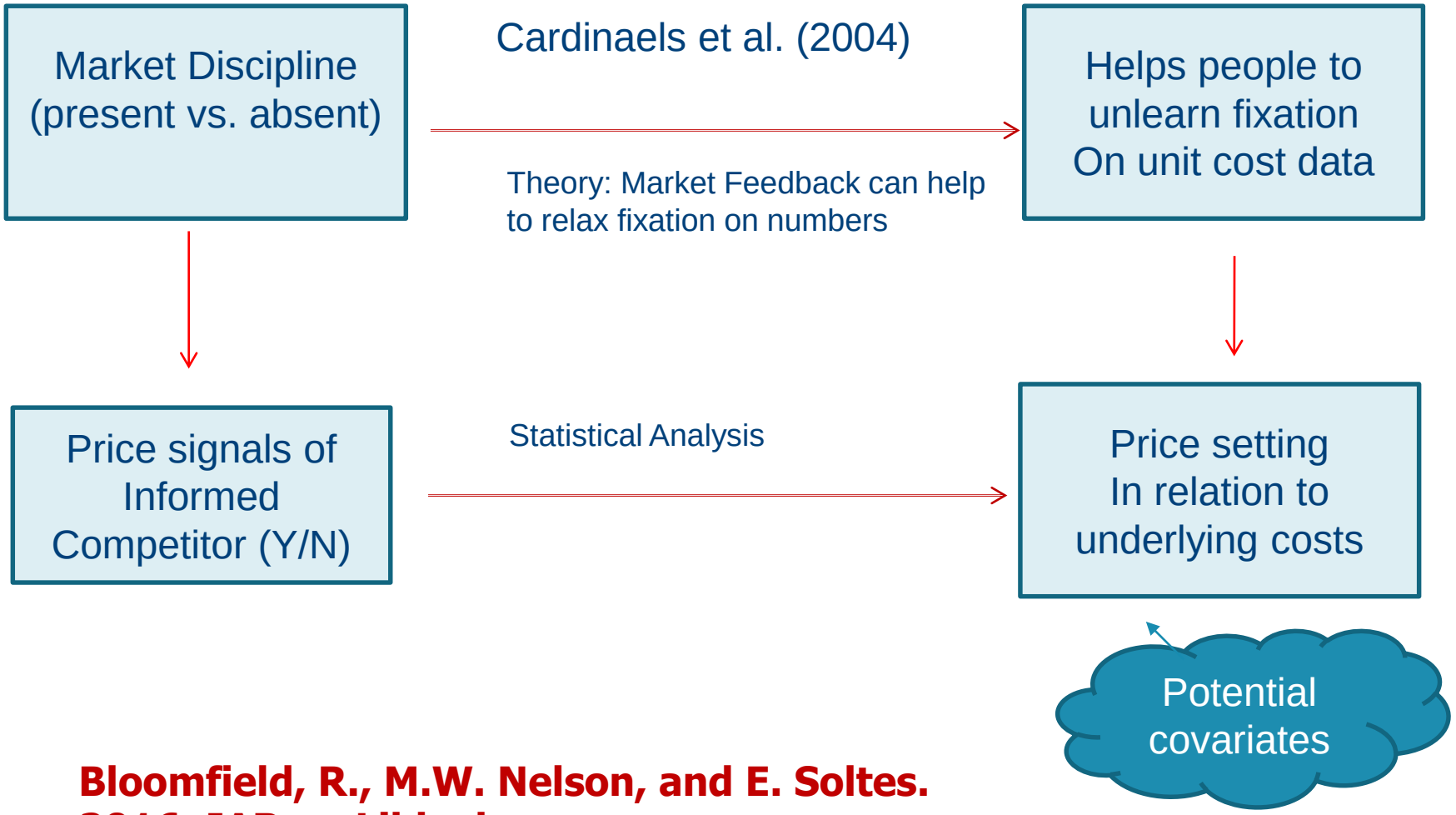
- ❑ A core theoretical construct is absent
- ❑ Tension is not clear
- ❑ The research question stays very broad

Similar idea: but now more focused: Theory driven experiment

- ❑ People anchor on numbers and ABC vs. VBC provide different anchor points → Tvesky and Kahneman: Anchoring and adjust
- ❑ How can we unlearn/debias this fixation !! → LIBBY BOXES



Libby boxes are key to operationalization



Bloomfield, R., M.W. Nelson, and E. Soltes. 2016. JAR on Libby boxes

Libby boxes help to defend validity

The experiment is not realistic. In reality, you also have….

- ❑ The goal of an experiment is not to be a reflection of reality.
- ❑ An experiment manipulates the construct of interest (all else held constant)

<u>Theoretical construct</u> determines the relevance of lab experiment.

- ❑ Managers will say that numbers matter: fixation happens all the time !!

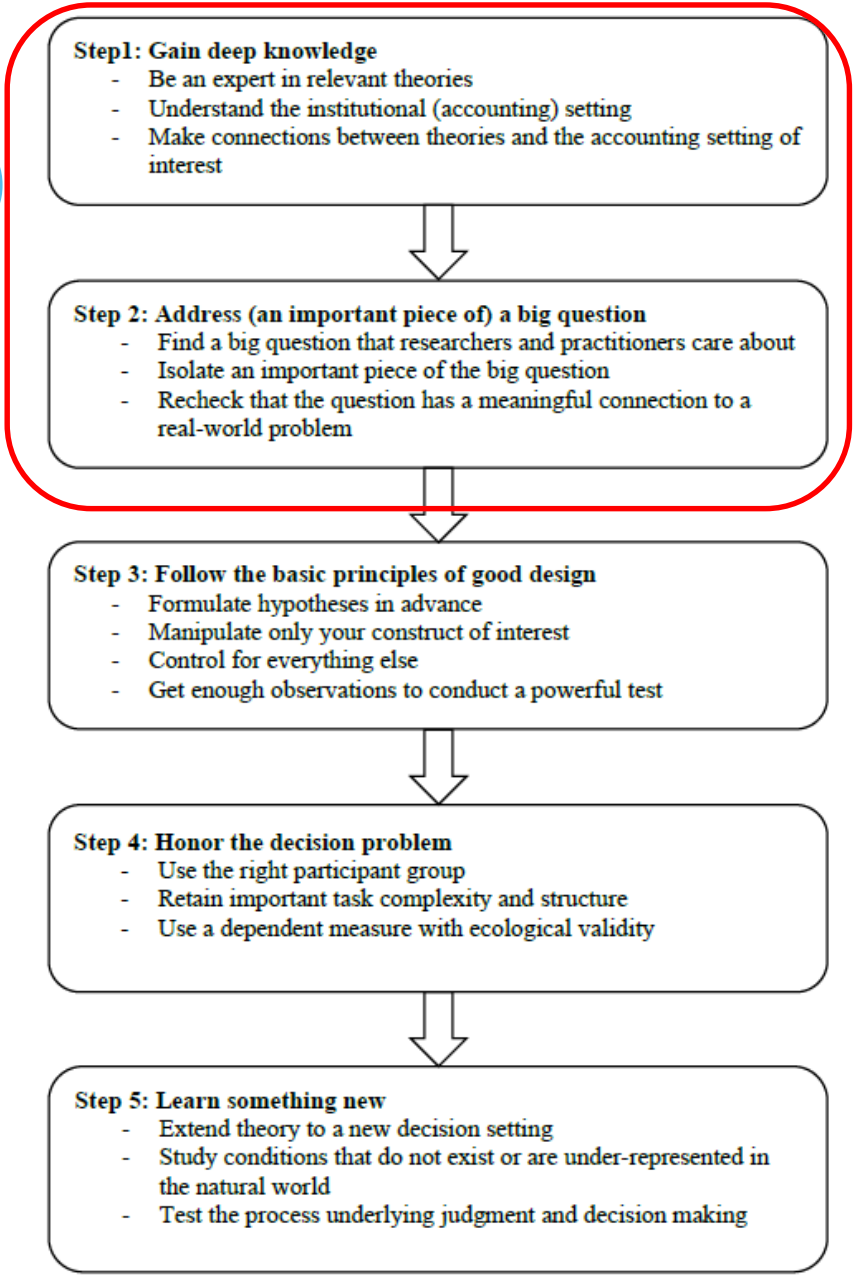
In real life, managers take such decisions, not students…

- ❑ students are not influenced by culture, experience, within organizations etc.
- ❑ If task requires specific knowledge: experts might be needed.

Correcting for errors in cost is difficult (universal bias): Yet, experienced managers may have better mental models = ***theory !!***

Kadous and Zhou (2016)

- Think in core concepts to define and motivate the theory
- Have a broad interest; econ/psych sociology → theory & design
- Be prepared for the tension question? Why should we care? What is new? Has this been done?
- Narrow down the question after staring broad
- Going from theory to design requires deep thought (Bloomfield et al. 2016)



external validity via strong internal validity

CSR as **preference** & matching
- Attract the right investors
- Bias of internal reporting

CSR and **social norms**:
- Positive spill-over on employees
- Signaling device

CSR is Important

"Formatting" & valuations
Integrated in business models?
Vs. stand alone reporting

Information processing story
Link investment too benefits ?
- Disclosure effectiveness
- Learning via accounting

INDIRECT RECIPROCITY (Balakrishnan et al. 2011)

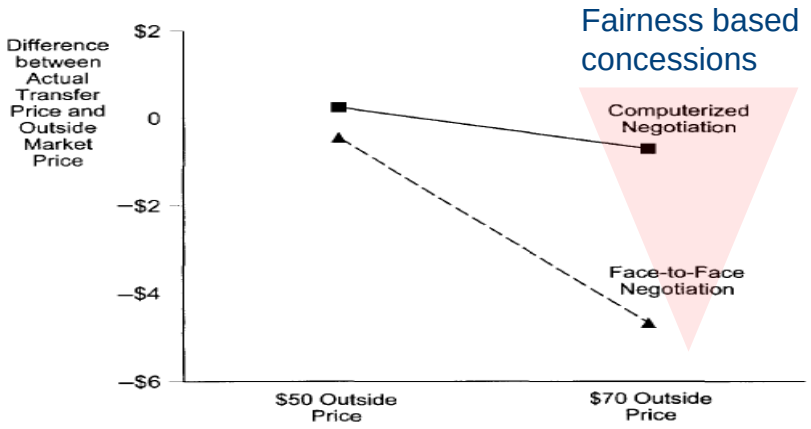
- Corporate giving to society: B invest in A
- Employees invest more effort: C invest in B
  "CSR" as relational contracting device (Gibbons 2005)
  instead of using costly incentives: You can trust your boss

Theory is key and less is more

EXPERIMENT = TREATMENT (present) vs. CONTROL (absent)

Moderation:

- When do we expect effect to be stronger (often 2 by 2 design)
- Or nested, Tafkov (2013): RPI absent, within present: public vs. private RPI



Mediation: Effect of treatment (IV) on Y (DV) is mediated by M
How do we get the effect = causal theoretical construct

- Too many effects, than consider if survey, archival study are suitable
- We focus on “extremes” to capture theoretical construct;
- MEDIATION: Think about this in your instrument. Do you have theory?

Something about introducing to many effects

Var 1

Var 2

Var 3

….

Dependent

This should not be too difficult, right

Var 1: Risk
- High risk
- Moderate Risk
- Low Risk

Var 2: measured Overconfidence (continuous)

Measure 1: risk taking
Measure 2: Decision time

Something about introducing too many effects

1) too many treatments: 2x2x2x2 designs

Main	2-way	3-way	4-way
A	AxB	AxBxC	AxBxCxD
B	AxC	AxBxD	
C	AxD	BxDxC	
D	BxC	AxCxD	
	BxD		
	CxD		

→ Even impossible to analyze
→ Often defined from practice

2) Levels: we don't explore all variations of practice

Towry and Kachelmeier (2002)
✓ 50$ TP (equal split of profit)
✓ 70$ TP (very unequal)

Why not?
50$ - 55$ - 60$ -65$ - 70$

INTERNAL ⇔ EXTERNAL validity

Common pitfall: too many factors are often deemed important !

KU LEUVEN

Conducting Experiments

"Even at execution, we take control"

Administer the manipulation

Check effect on Key dependent

Running experiments– within vs. between subjects

Experimental Studies in Accounting: **construct = important !**
E.g. often expressed as treatments e.g. incentive is either present or absent

- Most studies employ between subjects design: every participant is only assigned to one condition (one treatment)
  - E.g. most factors in Banker et al. (2004) are between subjects
  - Randomization is important

- Sometimes within-subjects design: subjects receive more than one treatment, he or she is exposed to the two levels of a factor:
  - Kachelmeier and Towry (2002): all participants receive outside selling options:
    price of 50 not favoring anyone (equal split)
    vs. 70 favoring seller (unequal split)

Running experiments– within vs. between subjects

Within-subjects advantages:

- ✓ Less participants required e.g. Kachelmeier and Towry (2002)
- ✓ Subject variables as a result of imperfect randomization do not influence the study.

Disadvantages (careful consideration):

- ✓ Not all variables are suitable: Arousal-studies in marketing, loud music
- ✓ Subjects can see through the manipulation → desirable responses
- ✓ Effect of order of presentation. If you first 50 dollar scenario first, before 70, people may split money more fair also under 70, because they are still in the 'equal' split mood.

Consideration

- ✓ ALWAYS COUNTER-BALANCE order if possible
- ✓ Maturation effect (learning across treatments), threats to internal validity !!
  A treatment needs to "wear off", before the other treatment can be given

randomization is key

In many studies you see the statement: Participants are randomly assigned to one
E.g. Banker et al. P14. Why would this be the case?

e.g. Why not run one condition day one and the other the next day?
(e.g. Monday morning: T1; Friday evening: T2)

Y = treatment +
holding constant (randomization)

Often we randomize within sessions

randomization is key – manipulation checks

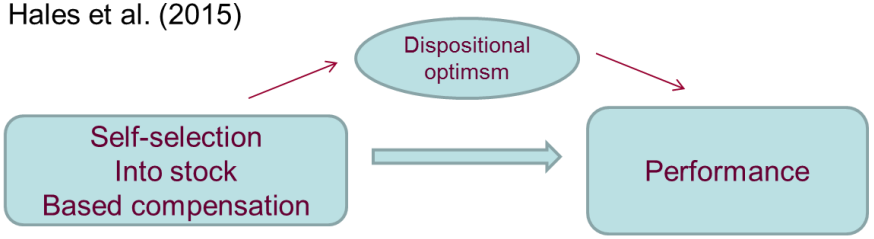
Randomization: rule out any other factors that may enter into the study

By randomization, you assume that differences in other factors
e.g. differences in age, motivation, work experience, knowledge, gender etc…
do not play a role, because participants are randomly distributed across cells.

Internal validity: We can attribute differences on DV to the treatment effects (IV).

MEDIATION: IV on DV explained by core theoretical construct

Hales et al. (2015)



randomization is key – manipulation checks

Post-questionnaire held after the task is crucial !!

- Collect demographics (e.g. Age, Gender, work experience, etc.) → You can use it as covariates to reduce the noise.
- Set of questions related to ruling out alternative explanations;
- Questions to see whether manipulation worked. Did people have greater strategy comprehension?
- MEDIATORS: often validated scales from psychology
- Potential covariates: sometimes moderator; effect depends on personal trait
  ⇔ Scales of innate preferences; people who trust, delegate more often)

Technology (black box): tape conversations, measure time, eye-tracking

Quasi experimentation and archival research

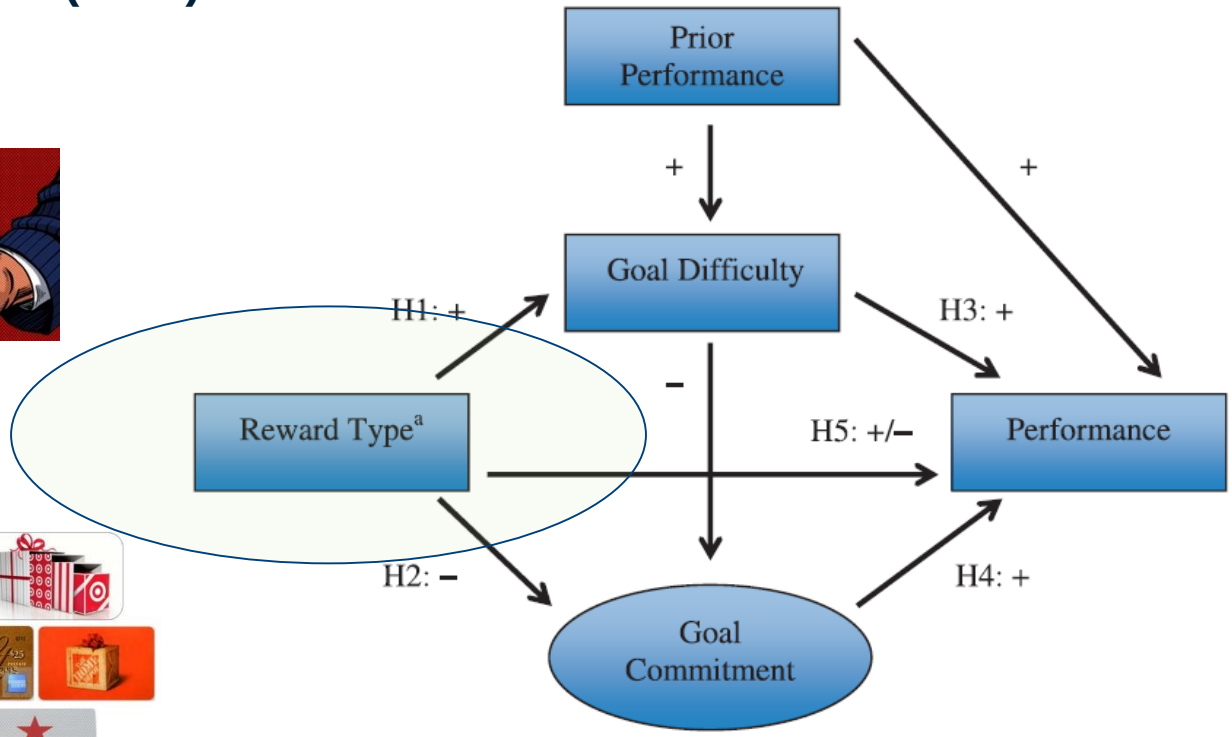
“We can give up control and make room for validation”

KU LEUVEN

Quasi experimentation

Presslee, Vance and Webb (2013)

FIGURE 1
Hypothesized Model



KU LEUVEN

Quasi experimentation

Presslee, Vance and Webb (2013)

BU 1 (NY)

30 people

BU 2 (Atlanta)

40 people

BU 3 (LA)

10 people

KU LEUVEN

Quasi experimentation

- Estimates the causal impact of an intervention on its target population, but they lack random assignment: one business cash; the other BU tangibles

- Quasi-experiments: subject to concerns regarding internal validity, because the treatment and control groups may not be comparable at baseline.

  - *Less control over the covariate, randomization process*
  - *End up with unequal amount of people across between subject factor*
  - *More room for "confounds": Control and treatment differ on many dimensions*

- **quasi-experiment allows interesting things to test !!**

  - Real ECON effects ⇔ more difficult in the lab (always depend on parameters)
  - Done to sample knowledgeable participants e.g. Auditors, Financial Analyst
  - Field validation of theories: From lab to actual practice

Validation via archival research

We do inform practice !!
Chen and Sandino (2012, JAR): Can Wages Buy Honesty? The Relationship Between Relative Wages and Employee Theft.

$$\begin{aligned} Theft_{it} = {} & \beta_0 + \beta_1 Relative\ Wages_{it} + \beta_2\ Coworker\ Presence_{it} \\ & + \beta_3\ Relative\ Wages_{it} \times Coworker\ Presence_{it} + \beta_4\ Employee\ Skills_{it} \\ & + \beta_5\ Store\ Manager\ Turnover_{it} + \beta_6\ Property\ Crimes\ Per\ Capita_{it} \\ & + \beta_7\ Unemployment_{it} + \beta_8\ Year\ 2004 + \varepsilon_{it} \end{aligned} \quad (2)$$

Archival data from several databases (store level data)

➔ RECIPROCITY: Relative wages relative to other employees in sector:
➔ NORM-based explanation: Co-workers presence