

Eksperimenter

Erik Gahner Larsen

Kausalanalyse i offentlig politik

Info: Talentprogram

- ▶ Talentprogram for kandidatstuderende i statskundskab og samfundsfag
- ▶ Giver bl.a. adgang til studiejobs, der udbydes i samarbejde med Odense Kommune, Socialstyrelsen og KMD
- ▶ Besøg sdu.dk/poltalent
- ▶ Deadline: 10. april

Info: Afleveringsdato

- ▶ Ingen afleveringsdato endnu
- ▶ Vil blive offentliggjort i begyndelsen af april

Info: Add ons

- ▶ Add on I: AGORA 16

- ▶ Mere info på facebook:

- <https://www.facebook.com/events/1539248969708294/>

- ▶ Add on II: Introduktion til R

- ▶ Format: Workshop
 - ▶ En eller to dage
 - ▶ Dato: TBA

Dagsorden

- ▶ Hvorfor eksperimenter?
- ▶ Typer af eksperimentelle designs
- ▶ Eksperimentelle tilgange
 - ▶ Laboratorieeksperimenter
 - ▶ Spørgeskemaeksperimenter
 - ▶ Felteksperimenter
- ▶ Styrker og svagheder

Hvorfor eksperimenter? Kausal inferens og randomisering

- ▶ Randomisering
 - ▶ Randomisering er let at forstå (tænk lotteri)
 - ▶ Ingen selvseleksion
 - ▶ Fair
 - ▶ Nemt at analysere
- ▶ Rubin: Realiseringen af potentielle udfald (eksperimenter er guldstandard ift. potentielle udfald og kontrafaktisk kausalitet)
- ▶ Campbell: Reducerer trusselsniveauet fra en række validitetstrusler (alternative forklaringer)

Hvorfor eksperimenter? Kausal inferens og randomisering

- ▶ Eksperimenter kan vi os en gennemsnitlig stimulieffekt, *average treatment effect*
 - ▶ $\tau_{ATE} = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$
 - ▶ Gennemsnit på udfaldsvariablen for stimuligruppen ($E[Y(1)]$)
 - ▶ Gennemsnit på udfaldsvariablen for kontrolgruppen ($E[Y(0)]$)
- ▶ Der er dog mange andre begreber
 - ▶ Forskellige effekter (SATE, PATE, LATE, ITT, ATT, ATC)
 - ▶ Heterogene effekter (altså ikke gennemsnitseffekter)

Hvorfor eksperimenter? Enkelthed

- ▶ Eksperimenter fordrer et eksplicit fokus på, hvad der er stimuli (*uafhængig variabel*) og hvad der er udfaldsvariabel (*afhængig variabel*)
- ▶ Tænk altid i eksperimenter
 - ▶ “If you can’t devise an experiment that answers your question in a world where anything goes, then the odds of generating useful results with a modest budget and nonexperimental survey data seem pretty slim. The description of an ideal experiment also helps you formulate causal questions precisely.” (Angrist og Pischke 2009, 5)

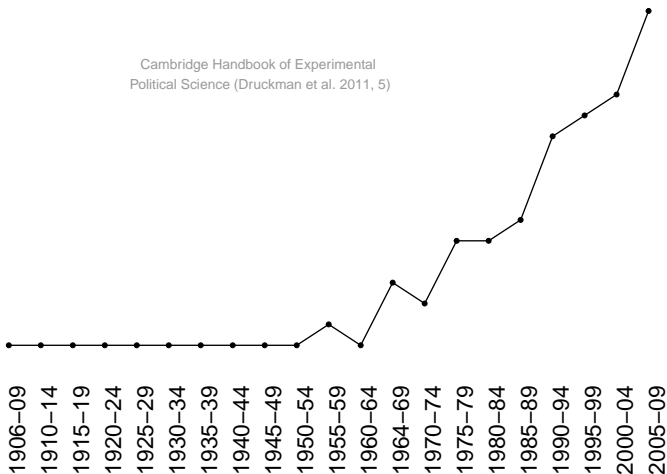
Hvorfor eksperimenter? Mere moderne end før

- ▶ Eksperimenter anvendes i hyppigere omfang til at besvare kausale spørgsmål
- ▶ Flere introduktionsartikler
 - ▶ Blom-Hansen og Serritzlew (2014): Endogenitet og eksperimenter - forskningsdesignet som løsning
 - ▶ Serritzlew (2007): Det politologiske eksperiment: hvorfor, hvornår og hvordan?
 - ▶ McDermott (2002): Experimental methods in political science
 - ▶ McDermott (2002): Experimental methodology in political science

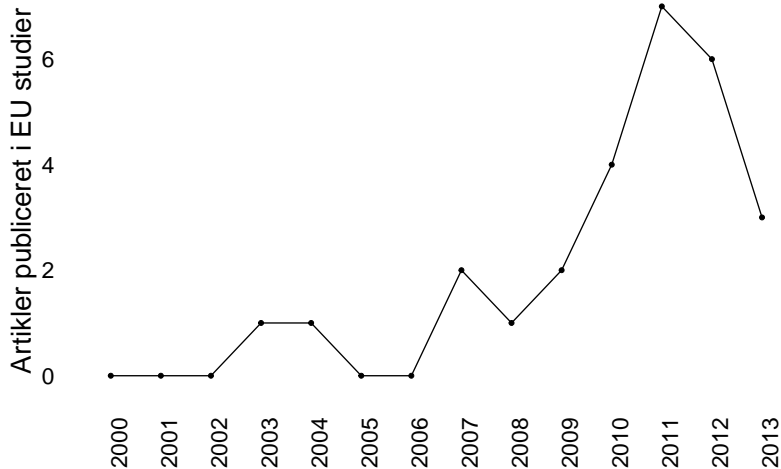
Hvorfor eksperimenter? Mere moderne end før

Artikler publiceret i APSR

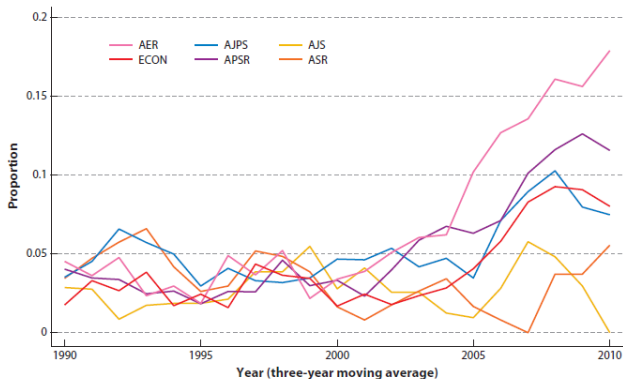
Cambridge Handbook of Experimental
Political Science (Druckman et al. 2011, 5)



Hvorfor eksperimenter? Mere moderne end før (Nielsen 2014)



Hvorfor eksperimenter? Mere moderne end før (Jackson og Cox 2013)



10 situationer hvor eksperimenter er anbefalelsesværdige

1. Når efterspørgsel overstiger udbud
 - ▶ Distribution af goder på en fair måde
2. Når et stimuli ikke kan gives til alle på samme tid
 - ▶ Økonomisk eller fysisk umuligt
3. Når enheder kan blive midlertidigt isoleret fra hinanden
 - ▶ Randomisering i tid
 - ▶ “Equiavalent Time Samples Design”

10 situationer hvor eksperimenter er anbefalelsesværdige

4. Når kontakten mellem enheder er lav
5. Når forandringer skal finde sted men løsningerne er ikke kendte
6. Når kriterier for tildeling af en service er tvetydig
7. Når enheder er indifferente i forhold til stimuli
8. Når du kan skabe din egen organisation
9. Når du har kontrol over enheder
10. Når lotterier forventes

Hvad er vores kontrafaktiske scenarie?

- ▶ Kausaleffekten tager udgangspunkt i et kontrafaktisk scenarie
- ▶ Kontrolgruppe (uden stimuli) eller flere stimuli der sammenlignes?
- ▶ Tre grupper: Stimuli, kontrol og placebo (Gerber et al. 2010)
- ▶ Forskellige stimuli
 - ▶ Kategorisk forskellige
 - ▶ Kvantitativt forskellige (parametrisk, doseringsrespons)
- ▶ Kontinuerlige stimuli er betydeligt mere komplicerede at analysere
 - ▶ Begynd i det små!

Eksempel: GOTV, stimuli, placebo og kontrol

- ▶ Nickerson (2008): “During a face-to-face canvassing experiment targeting households with two registered voters, residents who answered the door were exposed to either a Get Out the Vote message (treatment) or a recycling pitch (placebo).”

TABLE 3. Treatment Effect among Contacted Households

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in GOTV Group	47.7% (3.0)	42.4% (2.9)	27.1% (3.1)	23.6% (3.0)		
Percent Voting in Recycling Group	39.1% (2.9)	36.9% (2.9)	16.2% (2.7)	17.3% (2.7)		
Estimated Treatment Effect	8.6% (4.2)	5.5% (4.1)	10.9% (4.1)	6.4% (4.1)	9.8% (2.9)	6.0% (2.9)
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

Forskellige designs: sammenligning af stimuli og kontrol

- ▶ R : Randomiseret gruppe
- ▶ X : Stimuli
- ▶ O : Udfald

R	X	O
R		O

Forskellige designs: sammenligning af to stimuli

- ▶ R : Randomiseret gruppe
- ▶ X_A : Stimuli A
- ▶ X_B : Stimuli B
- ▶ O : Udfald

R	X_A	O
R	X_B	O

Forskellige designs: sammenligning af to stimuli og en kontrol

- ▶ R : Randomiseret gruppe
- ▶ X_A : Stimuli A
- ▶ X_B : Stimuli B
- ▶ O : Udfald

R	X_A	O
R	X_B	O
R		O

Forskellige designs: Pretest-posttest kontrolgruppe design

- ▶ R : Randomiseret gruppe
- ▶ X : Stimuli
- ▶ O_{pre} : Udfald *før* stimuli
- ▶ O_{post} : Udfald *efter* stimuli

R	O_{pre}	X	O_{post}
R	O_{pre}		O_{post}

Forskellige designs: sammenligning af to stimuli med pretests

- ▶ R : Randomiseret gruppe
- ▶ X_A : Stimuli A
- ▶ X_B : Stimuli B
- ▶ O_{pre} : Udfald *før* stimuli
- ▶ O_{post} : Udfald *efter* stimuli

R	O_{pre}	X_A	O_{post}
R	O_{pre}	X_B	O_{post}

Forskellige designs: sammenligning af to stimuli og en kontrol med pretests

- ▶ R : Randomiseret gruppe
- ▶ X_A : Stimuli A
- ▶ X_B : Stimuli B
- ▶ O_{pre} : Udfald *før* stimuli
- ▶ O_{post} : Udfald *efter* stimuli

R	O_{pre}	X_A	O_{post}
R	O_{pre}	X_B	O_{post}
R	O_{pre}		O_{post}

Forskellige designs: Faktorielt design

- ▶ Faktorforsøg
 - ▶ Flere stimuli der testes i *kombination* med hinanden
- ▶ Fordele
 - ▶ Kan kræve færre enheder (da én person bliver udsat for flere stimuli)
 - ▶ Gør det nemt at teste kombinationerne af stimuli
 - ▶ Muliggør interaktionseffekter
- ▶ Eksempel med et 2×2 faktorielt design:

	$B1$	$B2$
$A1$	X_{A1B1}	X_{A1B2}
$A2$	X_{A2B1}	X_{A2B2}

Forskellige designs: Faktorielt design

- ▶ R : Randomiseret gruppe
- ▶ $X_{A_1B_1}$: Stimuli A1B1
- ▶ $X_{A_1B_2}$: Stimuli A1B2
- ▶ $X_{A_2B_1}$: Stimuli A2B1
- ▶ $X_{A_2B_2}$: Stimuli A2B2
- ▶ O : Udfald

R	$X_{A_1B_1}$	O
R	$X_{A_1B_2}$	O
R	$X_{A_2B_1}$	O
R	$X_{A_2B_2}$	O

Forskellige designs: Tidsseriedesign

- ▶ R : Randomiseret gruppe
- ▶ X : Stimuli
- ▶ O : Udfald

R	$O...O$	X	O	$O...O$
R	$O...O$		O	$O...O$

Forskellige designs: Krydsdesign

- ▶ R : Randomiseret gruppe
- ▶ X_A : Stimuli A
- ▶ X_B : Stimuli B
- ▶ O : Udfald

R	O	X_A	O	X_B	O
R	O	X_B	O	X_A	O

Forskellige empiriske tilgange

- ▶ Eksperimenter (i dag)
 - ▶ Laboratorieeksperimenter
 - ▶ Spørgeskemaeksperimenter
 - ▶ Felteksperimenter
- ▶ Naturlige og kvasieksperimentelle tilgange (lektion 9)
- ▶ Observationsstudier (lektion 10 og 11)

Laboratorieeksperimenter

- ▶ Eksperiment der gennemføres i et kontrolleret miljø
- ▶ Ræsonnement: “Causal knowledge requires controlled variation.” (Falk og Heckman 2009, 537)
- ▶ Styrker
 - ▶ Kontrol med konteksten
 - ▶ Viden om konteksten
- ▶ Svaghed
 - ▶ Mindre realistisk kontekst

Eksempel: Toneleje og opfattelse af lederkompetence (Klofstad et al. 2012)

- ▶ Stimuli: Stemmeleje (tonehøjde)
 - ▶ “[. . .] we recorded men and women saying ‘I urge you to vote for me this November’. Each recording was manipulated digitally to yield a higher- and lower-pitched version of the original.”
- ▶ Udfald: Opfattelse af kompetence, styrke og troværdighed
- ▶ Resultater: “Our results show that both men and women select male and female leaders with lower voices. These findings suggest that men and women with lower-pitched voices may be more successful in obtaining positions of leadership.”

Eksempel: Politiske kandidaters (over)vægt (Miller og Lundgren 2010)

- ▶ Stimuli: “One-half of the participants were randomly assigned to read about female candidates and one-half were assigned to read about male candidates. These groups were then further divided into two groups, with one-half of participants reading about obese candidates and one-half of participants reading about nonobese candidates.”

Male, nonobese



Male, obese



Eksempel: Suppe i laboratoriet (Wansink et al. 2015)

- ▶ Bruger personer visuelle cues, når de spiser?
- ▶ “Fifty-four participants (BMI, 17.3 to 36.0 kg/m^2 ; 18 to 46 years of age) were recruited to participate in a study involving soup. The experiment was a between-subject design with two visibility levels: 1) an accurate visual cue of a food portion (normal bowl) vs. 2) a biased visual cue (self-refilling bowl). The soup apparatus was housed in a modified restaurant-style table in which two of four bowls slowly and imperceptibly refilled as their contents were consumed. Outcomes included intake volume, intake estimation, consumption monitoring, and satiety.”

Eksempel: Suppe i laboratoriet (Wansink et al. 2015)

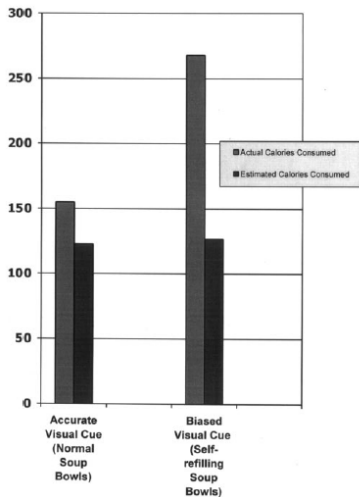


Figure 2: Biased visual cues provided by self-refilling soup bowls increased intake but not the perception of intake.

Spørgeskemaeksperimenter

- ▶ Forskellige versioner af et spørgeskema, tilfældigt hvilket spørgeskema respondenteren får
- ▶ Styrker og svagheder: Som ved andre spørgeskemaer
- ▶ Vigtigt at forholde sig til den eksterne validitet (Barabas og Jerit 2010; Hainmueller et al. 2015)
- ▶ Overvej spillover effekter (Gaines et al. 2007; Transue et al. 2009)

Spørgeskemaeksperimenter: Stimuliforstyrrelser

- ▶ Information (stimuli) kan korrelere med specifikke forhold, der kan drive en effekt
- ▶ Eksempel: Hvilken effekt har information om, at et land er et *demokrati*?
 - ▶ Hvad tænker du på, når du hører ordet demokrati?
 - ▶ Måske mere end demokrati. Vestligt land, rigt land osv.
- ▶ Vær bevidst omkring hvad stimuli er og hvad der kan forstyrre det

Spørgeskemaeksperimenter: Holdninger til immigranter (Hainmueller og Hiscox 2010)

- ▶ Spørgsmål: Påvirker immigranternes kvalifikationer vælgerens holdninger?
- ▶ Stimuli:

Version 1: Do you agree or disagree that the US should allow more *highly skilled* immigrants from other countries to come and live here? (emphasis added)

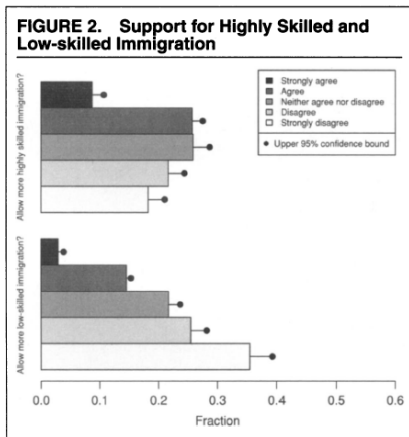
Version 2: Do you agree or disagree that the US should allow more *low-skilled* immigrants from other countries to come and live here? (emphasis added)

Answer options (both versions):

Strongly disagree	Somewhat disagree	Neither agree nor disagree
1	2	3
Somewhat agree	Strongly agree	
4	5	

Spørgekemaeksperimenter: Holdninger til immigranter (Hainmueller og Hiscox 2010)

► Resultater:

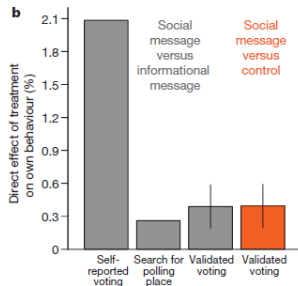
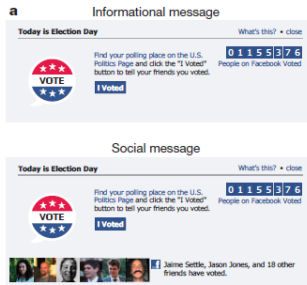


Felteksperimenter

- ▶ Eksperiment der foregår “i felten”
- ▶ Styrker
 - ▶ Realistisk kontekst
 - ▶ Kræver ikke at forsøgsparticipanterne skal ind i en ny kontekst
- ▶ Svagheder
 - ▶ Manglende viden omkring og kontrol med konteksten
 - ▶ Kan være yderst ressourcekrævende

Eksempel: Indhold på Facebook og politisk deltagelse

- ▶ A 61-million-person experiment in social influence and political mobilization (Bond et al. 2012)



FAQ: Er effekterne fra laboratorieeksperimenter forskellige fra felteksperimenter? (Mitchell 2012)

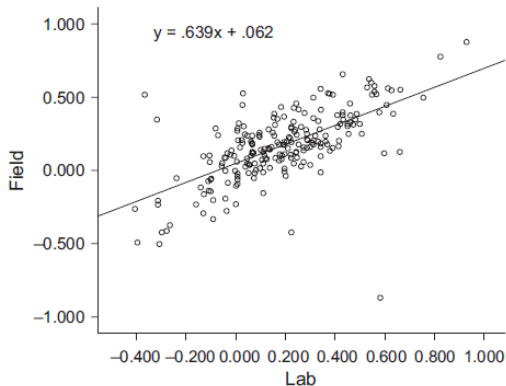
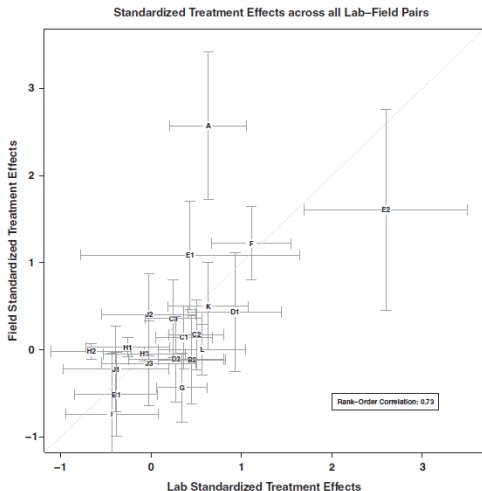


Fig. 1. Scatter plot of paired lab and field effects across all meta-analyses.

FAQ: Er effekterne fra laboratorieeksperimenter forskellige fra felteksperimenter? (Coppock og Green 2015)



FAQ: Hvilke eksperimenteffekter er der?

- ▶ Dem der gennemfører eksperimentet, kan påvirke forsøgsdeltagernes adfærd
- ▶ Tre effekter (der har et klart overlap)
 - ▶ Forventningseffekter
 - ▶ Når det kommunikeres - bevidst eller ubevidst - hvilke forventninger forsøget har til adfærden
 - ▶ Designeffekter
 - ▶ Når forskerens holdninger og forventninger påvirker den måde eksperimentet designes på
 - ▶ Efterspørgselseffekter
 - ▶ Når forsøgspersonen selv kan tolke, hvad der er formålet med forsøget
- ▶ Eksempel: Behavioral Priming: It's all in the Mind, but Whose Mind

FAQ: Hvad med sensitive spørgsmål?

- ▶ En stor udfordring ved mange undersøgelser er, at folk lyver i deres selvrappede holdninger
- ▶ Implicitte mål
 - ▶ Listeksperimenter (Blair og Imai 2012; Glynn 2013)
 - ▶ Implicit association test

FAQ: Er eksperimenter kun anvendelige inden for specifikke områder?

- ▶ Alle *videnskabelige* områder inden for samfundsvidenskaberne gennemfører eksperimenter
- ▶ Selv inden for international politik sker der metodiske fremskridt
 - ▶ Overblik:
 - ▶ Hudson (2010): The State of Experimental Research in IR: An Analytical Survey
 - ▶ Hyde (2010): The Future of Field Experiments in International Relations
 - ▶ Hyde (2015): Experiments in International Relations: Lab, Survey, and Field
 - ▶ McDermott (2011): New Directions for Experimental Work in International Relations
 - ▶ Mintz et al. (2011): Experimental Approaches to International Relations

FAQ: Hvad med kontrolvariable i eksperimenter?

- ▶ Hvorfor gennemfører vi eksperimenter? For at gøre kontrolvariable overflødige.
 - ▶ Derfor: Rapporter altid effekter, hvor der ikke er kontrolleret for noget
 - ▶ “If an estimated treatment effect is insignificant in the absence of controls, this should clearly shape our interpretation of the effect being estimated” (Dunning 2012, 268)
- ▶ Derefter: Variable målt før introduktionen af et stimuli er som regel gode kontrolvariable (Angrist og Pischke 2009, 64ff)
 - ▶ Kontrolvariable reducerer støj, større chance for at få statistisk signifikante resultater (Mutz 2011, 123f)

FAQ: Hvordan er forholdet mellem intern og ekstern validitet i eksperimenter?

- ▶ Forkert: “an inherent trade-off exists between the two.” (McDermott 2013, 610)
- ▶ Korrekt: “It makes no sense to say that some empirical research is low on internal validity but high on external validity.” (Morton og Williams 2010, 275)
- ▶ Ekstern validitet skal altid ses i relation til ens teori (Lucas 2003)

FAQ: Hvilke styrker og svagheder er der ved forskellige forsøgsdeltagere?

- ▶ Debat om styrker og svagheder ved at bruge studerende som forsøgsdeltagere (Sears 1986; Henry 2008)
- ▶ Debat omkring implikationerne af Amazon mTurk (Berinsky et al. 2012; Goodman et al. 2013; Rand 2012)
- ▶ Igen: ekstern validitet
 - ▶ WEIRD

FAQ: Hvad skal rapporteres fra et eksperiment?

- ▶ Gerber et al. (2014) har udformet en tjekliste
- ▶ Der skal være informationer relateret til
 - ▶ Hypoteser
 - ▶ Forsøgshederne og kontekst
 - ▶ Allokeringemetode
 - ▶ Information omkring randomisering eller fraværet af samme
 - ▶ Stimuli
 - ▶ Udfaldsmål og kontrolvariable
 - ▶ CONSORT (*Consolidated Standards of Reporting Trials*)
 - ▶ Statistiske analyser
 - ▶ m.v.
- ▶ Se også Boutron et al. (2010): Reporting Methodological Items in Randomized Experiments in Political Science

FAQ: Hvad med etiske overvejelser?

- ▶ Informeret samtykke
 - ▶ Information om, at man tager part i et eksperiment
- ▶ Risiko/gevinst overvejelser
 - ▶ *Do no/minimal harm*
- ▶ Debriefing
 - ▶ Information om eksperimentet
 - ▶ Feedback fra forsøgsdeltagere
- ▶ Misinformation (*deception*)
 - ▶ Vi kan ofte ikke give fuld information omkring eksperimentet
 - ▶ Påvirker forsøgsdeltagernes adfærd (Jamison et al. 2008)

Misinformation, forskellige metoder (Hertwig og Ortmann 2008)

Percentage (and Number) of Studies Published in the *Journal of Experimental Social Psychology* in 2002 Using a Classification of Methods of Deception Proposed by Sieber et al. (1995)

<i>Method of Deception</i>	<i>How Many of the Deception Studies Use a Given Method</i>
<i>False purpose.</i> Participants are given, or be caused to hold, false information about the main purpose of the study	87% (55)
<i>Bogus device.</i> Participants are given false information concerning stimulus material ^a	62% (39)
<i>Role deception.</i> Participants interact with participants about whose identify they have been given false information	24% (15)
<i>False feedback regarding self.</i> Participants are given false feedback about themselves	30% (19)
<i>False feedback regarding others.</i> Participants are given false feedback about another person	24% (15)
<i>Two related studies.</i> Two related studies are presented as unrelated	9.5% (6)
<i>Unaware of measure.</i> Participants are kept unaware that a study is in progress at the time of manipulation or measurement, or unaware of being measured (e.g., videotaped)	3% (2)
<i>Unaware of participation.</i> Participants are kept unaware of being subjects in research	0%

Note. There were 117 studies, of which 63 used deception.

Diskussion: Ulemper ved eksperimenter

- ▶ Ikke alle fænomener kan studeres eksperimentelt
 - ▶ Praktiske begrænsninger
 - ▶ Ethiske begrænsninger
- ▶ Lav økologisk validitet
- ▶ Ikke-repræsentative forsøgsparticipanter
- ▶ Eksperimentel bias
- ▶ Implementeringsproblemer
 - ▶ Hvem står for randomiseringen?
 - ▶ Hvem selekterer sig ud fra og ind i eksperimentet?
- ▶ Potentielt forsimplede statistiske analyser
- ▶ Se blandt andet Berk (2005), Deaton (2010) og West (2009)

Gennemførelsen af eksperimenter i offentlig politik

- ▶ Overordnede råd (fra Stoker 2010):
 - ▶ Dyrk menneskelige og sociale færdigheder og opbyg relationer/netværk
 - ▶ Forstå at politik er en væsentlig del af policyprocessen
 - ▶ “Policy programs are not neutral, antiseptic, laboratory type entities. They emerged from the rough and tumble of political support, opposition, and bargaining.”
 - ▶ Gennemfør eksperimenter og vær tålmodig
 - ▶ Vær selvkritisk (forskning er en politisk aktivitet)
- ▶ Specifikke råd (fra Hansen og Pontoppidan 2014)
 - ▶ Tænk evaluator ind tidligt i planlægningen
 - ▶ Hele organisationen skal være med
 - ▶ Vær opmærksom på at forberedelsesfasen kan være lang
 - ▶ Sørg for at samarbejde på tværs for at få nok deltagere
 - ▶ Inddrag evt. kvalitative interviews
 - ▶ Hav styr på økonomien

Dagens øvelse: Lab, spørgeskema og felt

- ▶ Find et kausalt argument (evt. fra din seminaropgave)
- ▶ Beskriv hvordan det kausale argument kan testes i forskellige eksperimenter
 - ▶ Laboratorieeksperiment
 - ▶ Spørgeskemaeksperiment
 - ▶ Felteksperiment

Næste gang

- ▶ Naturlige eksperimenter
- ▶ Onsdag, næste uge (30. marts), 10-12 i U25A.
- ▶ God påske