



Surveyeksperimentet: Et underutnyttet forskningsdesign for sosiologisk kausalanalyse

The survey experiment: An underutilized research design for sociological causal analysis

Christer Thrane

Professor (dr.polit. i sosiologi), Handelshøgskolen ved Høgskolen i Innlandet

Christer.Thrane@inn.no

Sammendrag

Sosiologer har typisk stått nokså fjernt fra å benytte eksperimentelle forskningsdesign til kausalanalyse. I denne artikkelen argumenterer jeg for at ett bestemt eksperimentelt design – surveyeksperimentet – utgjør et underutnyttet design for en kvantitativ sosiologi med kausale ambisjoner. Jeg presenterer også et surveyeksperiment for illustrasjonsformål; dvs. en case om naboeffekter på restaurantgjesters tipsbeløp. I lys av casestudien diskuteres noen momenter som ofte kommer opp ved forberedelse til, og gjennomføring av, et surveyeksperiment. Min konklusjon er at surveyeksperimentet i fremtiden bør få større plass i kvantitativ sosiologi med kausale ambisjoner, ikke minst fordi dette designet ofte vil være relativt enkelt og lite ressurskrevende å gjennomføre på sosiologiske temaer som til nå har vært lite utsatt for et kausalt perspektiv.

Nøkkelord

Survey, surveyeksperiment, sosiologisk kausalanalyse, eksperimentelt design

Abstract

Traditionally, sociologists have typically been reluctant in using experimental research designs for causal analysis. Against this background, I argue in this study that one type of experimental design – the survey experiment – is an underutilized research design for a quantitative sociology with causal ambitions. For illustration purposes, I also present a case study based on a survey experiment, focusing on the effect of peers on restaurant guests' tip amounts. In light of the case study, I also discuss some general features related to the planning and carrying out of survey experiments. My conclusion is that the survey experiment in the future should get a more prominent place in a causally oriented sociology, especially since this research design quite often is relatively easy and cheap to put into practice on sociological topics that so far have not been examined from a causal perspective.

Keywords

Survey, survey experiment, sociological causal analysis, experimental design

Introduksjon

Mye sosiologi har gjennom historien hatt kausale ambisjoner i oppsang og teorikapittel, for så å falle ned på tentative kausaltolkninger av hvordan x påvirker y basert på regresjonsanalyser av surveydata.¹ Løsningen på denne diskrepansen mellom teoretiske ambisjoner og empiriske begrensninger, har de siste 30 årene ofte vært bruken av paneldata og såkalt *fixed effects* regresjon.² Mye av den samme utviklingen har funnet sted i økonomi og statsvitenskap, men i senere tid taler mye for at økonomene og statsviterne har skilt lag med sosiologene. Det vil si, mens sosiologene som oftest fortsatt holder seg til paneler og statistisk kontrolltenkning i sin søken etter å avdekke årsak-virkning-forhold, så har økonomene og statsviterne i langt større grad begynt å ta eksperimentelle design i bruk for å avdekke slik kausalitet.³

Så kan man spekulere i hvorfor sosiologene ikke har gått i samme retning, og da mener jeg spekulere. Utover den åpenbare grunnen at mange sosiologiske spørsmål er lite egnet for et eksperimentelt design, ligger én grunn trolig i at sosiologer tenker på eksperimenter som kunstige, og at de derfor har liten overførbarhet til det virkelige liv (jf. Jackson & Cox, 2013; Thye, 2014). En annen grunn kan være at det historisk sett i liten grad har blitt undervist i eksperimentelle design ved sosiologiutdanningene (jf. Walker & Willer, 2007a), og at derfor har få sosiologer in spe lært særlig mye om disse designene. En tredje grunn følger i så fall av de to første. Siden eksperimentelle design med noen unntak som jeg kommer mer tilbake til, glimrer med sitt fravær i nyere sosiologisk forskning, er det få som tør å satse på disse når det søkes om midler til forskning.

Denne artikkelen handler om det som for mange sosiologer trolig er det minst kjente eksperimentelle designet i samfunnsvitenskapen: surveyeksperimentet.⁴ Walker og Willer (2007a) rangerer de fire eksperimentelle designene etter hvilken kontroll forskeren har på eksperimentbetingelsene, dvs. fra ingen til stor: (1) naturlige eksperiment, (2) felteksperiment, (3) surveyeksperiment og (4) laboratorieeksperiment. (Se også Jackson og Cox (2013).) Sosiologer har i nyere tid befattet seg hovedsakelig med (1) og (2);⁵ Ugreninov og

1. Med «kausale ambisjoner» menes i denne artikkelen at man søker å påvise at x er en årsak til at y varierer (eller endres). Konkret forfektes den kontrafaktiske kausalposisjonen som gjerne kalles Holland-Rubin-modellen, som igjen ligger bak gullstandarden for medisinske eksperimenter: *Randomized Controlled Trials* (RCTs). Omfanget på denne artikkelen samt dens hovedtema tillater ikke en bred innføring i kausalitet som fenomen i en sosiologisk og/eller kvantitativ kontekst; lesere henvises til Tuftes (2013) for dette. Se også Imbens og Rubin (2016).
2. Merk at Morgan og Winship (2015) hevder at *fixed effects* ikke er en vidunderkur for sikkert å kunne identifisere kausaleffekter. Populasjonspaneler basert på registerdata løser heller ikke noe identifikasjonsproblem, selv om de gjør signifikanstester mindre relevante.
3. Det er ikke enkelt å belegge denne utviklingen med tall, og vi taler uansett om en tendens, jf. Jackson og Cox (2013). Men innen økonomi er det neppe tilfeldig at fire nylige mottakere av Nobels minnepris – R. Thaler i 2017 og M. Kremer, A. Banjeree og E. Duflo i 2019 – har bygget sine karrierer på eksperimentelle design. Thalers pris var også en honnør til faget atferdsøkonomi, som eksplisitt bygger på psykologi og benytter eksperimentelle design i stor utstrekning. Her hjemme er FAIR-senteret ved NHH (tidligere *ChoiceLab*) et eksempel på denne vendingen i økonomifaget. I statsvitenskapen har spesielt surveyeksperimentet, temaet i denne artikkelen, blitt svært populært de senere årene (Barabas & Jerit, 2010; Mullinix et al., 2015; Mutz, 2011). Christensen og Aars (2017) er et eksempel på en norsk statsvitenskapelig studie med kausalambisjoner basert på et surveyeksperiment.
4. Her er det behov for å presisere: Sosiologiske surveyeksperimenter med kausalambisjoner virker til å være få i tallet totalt – og i praksis til å være fraværende i en norsk kontekst. I denne forstand er de trolig «minst kjente» for sosiologer. (Men det er som kjent vanskelig å dokumentere at noe ikke finnes eller glimrer med sitt fravær.) Når det er sagt, finnes det mange studier basert på surveyeksperimenter med andre formål. For det første har vi vignett-studiene. Her blir respondenter bedt om å vurdere ulike scenarier fremstilt via bilder, dvs. vignetter. Vignettene randomiseres som eksperimentelle stimuli, og y er i neste omgang de spurtes vurdering av disse vignettene (jf. Wallander, 2009). For det andre har vi surveyeksperimenter der formålet er å utbedre selve måleinstrumentet (*survey method experiments*); jf. Bakken et al. (2019).
5. Spesielt på 50- og 60-tallet gjorde også sosiologer en rekke smågruppestudier basert på laboratorieeksperimenter.

Birkelund (2013) gir en innføring i naturlige eksperimenter, mens Birkelund et al. (2014) er et eksempel på et felteksperiment. Psykologene har historisk befattet seg med (4), mens økonomene av i dag gjør (1), (2) og (4). Men det er grunn til å merke seg at når psykologer og økonomer gjør (4), så har de ulike tilnærminger. Mens psykologene ofte forleder sine subjekter om hensikten bak en studie, er økonomene alltid åpne om hva eksperimentet handler om (Helland, 2016). Videre foretrekker økonomene normalt å gi incentiver, dvs. belønning, til subjektene for å delta i eksperimenter.

Denne artikkelen handler altså om surveyeksperimentet. Dette er det tre grunner til. For det første er naturlige eksperimenter og felteksperimenter velkjent for sosiologer. For det andre finnes det en omfangsrik litteratur om laboratorieeksperimenter enkelt tilgjengelig. Motsatt taler som nevnt mye for at surveyeksperimentet er ukjent for mange sosiologer (jf. fotnote 4), trolig på grunn av at det sjeldent omtales i metodebøker, og at man sjeldent finner kausalorienterte studier basert på dette designet i tidsskrifter. Det siste er hovedmotivasjonen for denne artikkelen, dvs. at surveyeksperimentet er en underutnyttet mulighet for en kvantitativ sosiologi med kausale ambisjoner. Den korte begrunnelsen for dette, er at surveyen basert på et sannsynlighetsutvalg ofte har høy ekstern validitet; at den tilfeldige trekningen sikrer at man kan uttale seg presist om x og y sin sentraltendens i en populasjon basert på studiet av et tilfeldig utvalg fra denne. Motsatt har surveyen lav intern validitet; den er uegnet til å si noe om den kausale effekten som x eventuelt har på y . Et eksperiment har ofte lav ekstern validitet; det er sjelden rett-frem å overføre dets resultat til en annen setting eller gruppe mennesker. Derimot har eksperimenter høy intern validitet, i betydningen at de er egnede til å identifisere den eventuelle kausaleffekten x har på y blant de som studeres. Slik sett kombinerer surveyeksperimentet potensielt det beste fra to verdener: både høy intern og høy ekstern validitet (jf. Nock & Guterbock, 2010).

Resten av artikkelen er bygd opp som følger: Først plasseres surveyeksperimentet i en større kontekst; dernest presenterer jeg en casestudie basert på dette designet fra A til Å. Noen momenter ved gjennomføring av surveyeksperimenter diskuteres så med referanse til den forutgående casen. Til slutt sammenfatter jeg og trekker opp noen implikasjoner for surveyeksperimentet i fremtidig sosiologisk forskning med kausale ambisjoner. Artikkelen retter seg primært mot sosiologer og andre med noe erfaring i bruk av kvantitative metoder, men kan trolig også leses med utbytte av dem som er nokså ferske i en slik sammenheng. Lesere på jakt etter en rigorøs behandling av hvordan og hvorfor eksperimentelle design gir grunnlag for en kausal slutning, samt hva som kan gå galt i denne sammenhengen, henvises til for eksempel Imbens og Rubin (2016).

Kausalitet, eksperimenter og surveyeksperimenter

Til hverdags tenker vi gjerne forenklet på kausalitet som at en endring i ett fenomen, årsaken x , skaper en endring i et annet fenomen, virkningen y . Et tidsaktuelt eksempel fra våren 2020 kan være at et politisk tiltak om å «stenge ned» offentlige rom i Norge (x) fører til en økning i antall konkurser i restaurantbransjen (y). Det praktiske problemet i all kvantitativ forskning består i å identifisere så sikkert som mulig at det er endringen i x , og ikke i en annen variabel, z , som er den reelle årsaken til y . Og med en henvisning til for eksempel Angrist og Pischke (2015), vil mange si at den sikreste måten å avdekke slik kausalitet på, er ved å gjennomføre et eksperiment der dette er mulig.

Walker og Willer (2007a, 2007b) hevder at det konseptuelt finnes to typer eksperimenter: (A) det oppdagende og (B) det teoritestende. Hensikten med (A), med røtter hos J. S. Mill og senere raffinert av R. A. Fisher, er å finne ut hva som skjer med y når *ett sett* kontrollerte

eksperimentbetingelser sammenlignes med, eller endres til, et annet sett slike betingelser. Det typiske for vår tid er en *Randomized Controlled Trial*-studie (RCT): Variabelen x manipuleres etter randomisering (les: myntkast): medisin eller placebo. Andre betingelser holdes dermed konstante i gjennomsnitt. Etter en tid sammenlignes de som fikk medisin med de som fikk placebo, dvs. kontrollgruppen, for y – typisk en diagnostisk test for sykdomsplager. Forskjellen i snitt for y i de to eksperimentgruppene blir nå kausaleffekten. Grunnen er at en vellykket randomisering sikrer at det ikke er noen systematisk korrelasjon mellom eksperimentbetingelsene og andre uavhengige variabler som påvirker y (jf. Angrist & Pischke, 2015). Dermed må medisineringsen, og ikke noe annet, være årsaken til at pasientene som fikk ekte medisin er blitt friskere (i snitt) enn pasientene som fikk placebo. For (B) er det teoretiske ideer og hypotesene avledet fra disse som utgjør utgangspunktet for eksperimentbetingelsene. Medisin versus placebo for x erstattes av to teoretisk definerte og (som regel) mer jevnbyrdige eksperimentbetingelser. Om (A) eller (B) velges i praksis, er avhengig av problemstillingens art og tilstanden for teoriutviklingen på feltet. Når det er sagt, er det ofte glidende overganger mellom de to typene eksperimenter i praksis ifølge Jackson og Cox (2013).

Surveyeksperimentet kan også klassifiseres langs linjene over, med den forskjellen at det kan foregå hvor som helst. Formelt er surveyeksperimentet en survey som lages i n versjoner, der n viser til antallet eksperimentbetingelser. I det enkleste tilfellet, $n = 2$, skjer tildelingen av surveyversjon i eller ii etter samme randomiserte prosedyre som gir medisin eller placebo i en RCT. Survey i kan være at enkelte betingelser er til stede (medisin), mens survey ii kan være at de ikke er til stede (placebo); i andre tilfeller kan i og ii vise til to ulike teoretiske posisjoner. Etter tildelingen av i eller ii følger så spørsmål/variabel y . I praksis viser i og ii ofte til lav eller høy verdi på en uavhengig variabel, x , gjerne kalt en faktor. Den uavhengige variabelen sees opp mot y , og tilsvarende utgjør snittforskjellen mellom de to surveygruppene for y kausaleffekten. Man kan også manipulere to uavhengige variabler med verdiene lav eller høy, noe som gir et 2×2 design og i alt fire surveyversjoner. (Ulike eksperimentbetingelser kan også tenkes langs en mer kontinuerlig skala, uten at dette får andre konsekvenser enn at det kan tvinge frem et spørsmål om korrekt funksjonell form mellom x og y .)

Surveyeksperimentet kan tenkes benyttet til å studere mange kausale problemstillinger innen sosiologi, men peker seg i dag kanskje spesielt ut som egnet på områder der det *ikke* finnes registerdata. Videre er problemstillinger rundt normers og andre sosiale krefters påvirkning på (individuell) atferd svært aktuelle – her er det bare den sosiologiske fantasien som setter grensene. Under presenteres en illustrativ casestudie basert på et surveyeksperiment og en kommenterende analyse av dette.

Naboeffekter på restaurantgjesters tipsbeløp: En casestudie basert på et surveyeksperiment

Bakgrunn og forskningsspørsmål

At våre individuelle valg formes av våre omgivelser, er opplagt. Mye av denne effekten har å gjøre med hva mennesker i vår nærhet gjør når de er i situasjoner som ligner på situasjonene vi selv er i. Nærhet vil ofte være fysisk, men den kan også være i overført betydning: kulturell, verdimeisig eller tankemessig. På fagspråket kalles dette *peer effects*, *peer pressure* eller *social influence*. (Likemannseffekt lyder ikke bra.) Men selv om slike naboeffekter, som jeg kaller det her, er åpenbare og godt dokumenterte på mange felt (jf. rusmidler, skoleprestasjoner, risikovurderinger, forbrukeratferd), er de sjelden enkle å studere med et kausalt blikk. Grunnen er at andre årsaksvariabler enn «hva naboen gjør» oftest er relevante, men at det å

kontrollere for alle disse er vanskelig i praksis. Dermed blir kausaleffekten usikker. Å gjennomføre et surveyeksperiment kan da være en bedre egnet strategi for å få identifisert en slik naboeffekt.

Under ser jeg på naboeffekter i forbindelse med betaling for restaurantbesøk. I mange land i verden, men først og fremst i USA, er det tilnærmet obligatorisk å gi et tips på ca. 10 til 20 % av totalbeløpet når man betaler for et måltid på restaurant (Lynn, 2015). I Norge er det ingen slik utbredt tipsnorm, men noe tipsing foregår likevel. Motiver for å tipse kan være et signal om at man setter stor pris på god service (og liten pris på dårlig), at man ønsker god service neste gang man besøker restauranten eller for å bidra til servitørers relativt lave fastlønn (Saunders & Lynn, 2010). Men spørsmålet om tipsing har nå altså en annen teoretisk inngang: Påvirkes restaurantgjesters tipsbeløp av hvor mye sidemannen eller sidekvinnen ved bordet – altså naboen – gir i tips?

Elektroniske spørreskjemaer og data

Spørreskjemaene startet med et felles spørsmål til alle respondentene, som kartla om de kunne se for seg restaurantscenariet (settingen) som surveyeksperimentet handlet om.⁶ Blant dem som svarte ja, fikk ca. 50 % følgende spørsmål etter en randomisert prosedyre: «Måltidet smakte bra, servicen var god og din personlige regning er på 510 kr. Personen ved siden av deg, som spiste/drakk ca. det samme som deg, gir 0 kr i tips. Hvor mye gir du i tips? (Skriv inn tipsbeløp, 0 kr er et mulig beløp.)» Den andre halvparten av respondentene fikk det samme spørsmålet med *én* forskjell, dvs. at scenariet ble avsluttet med «... gir 60 kr i tips».⁷ Til slutt fulgte noen spørsmål (dvs. kontrollvariabler) om besøksfrekvens i restauranter, arbeidserfaring i restauranter samt et fåtall sosio-demografiske spørsmål.

Vi laget en webside for undersøkelsen med en lenke til de elektroniske spørreskjemaene.⁸ Skjult bak lenken lå en *random number generator* som tilfeldig valgte ut én versjon av de to skjematypene med det felles åpningsspørsmålet. Informasjonen om websiden, samt lenken til denne, ble fremvist hver dag i én uke i en posting på forfatterens Facebook-side, supplert med oppfordringer om å svare på undersøkelsen/dele lenken videre til andre FB-venner. Den empiriske analysen gjelder for 736 personer over 18 år som svarte på undersøkelsen i løpet av uke 5 og 6 i 2020.

Noen kjennetegn ved utvalget går frem i Tabell 1. Å spise på restaurant gjøres 20 ganger i snitt per år (median = 15; ikke vist), dvs. én til to ganger i måneden for de fleste. Den typiske respondenten har ingen erfaring med arbeid i restaurantbransjen (62 %) og er kvinne (56 %). Gjennomsnittsalderen er 48 år (median = 50; ikke vist). Vestfold, Telemark og Innlandet er de mest typiske bostedsfylkene. Utvalget er selvsagt verken tilfeldig trukket eller representativt for en populasjon, noe jeg kommer tilbake til når jeg tar opp momenter det er relevant å tenke igjennom når man skal gjennomføre et surveyeksperiment i praksis.

6. Spørsmålet lød: «Du spiser middag på restaurant sammen med venner en lørdag kveld, det går mot slutten av måltidet, og du skal kun betale for din egen mat/drikke. Kan du se for deg dette?»

7. Dette er en forenkling av presentasjonsmessige grunner. Det var i alt fire versjoner av spørreskjemaet, dvs. et 2 × 2 design, der «Personen» alternerte mellom «Mannen» og «Kvinnen». Men siden jeg ikke tar for meg denne kjønn/nabo-variabelen nå, gjør jeg det enklest mulig.

8. Erik Haugom laget websiden, og takk til ham for dette. Spørreskjemaet ble laget via Nettskjema (<https://nettskjema.no/>).

Tabell 1. Kjennetegn ved utvalget. N = 736.

	Gjennomsnitt	Standardavvik	Andel (%)
Antall restaurantbesøk siste år ^a	19,75	19,99	
Arbeidserfaring i restaurant:			
Nei			62
Ja, litt erfaring			26
Ja, mye erfaring			12
Kjønn (kvinne = 0; mann = 1)	0,44	0,50	
Alder	48,33	10,51	
Bosted:			
Landlig			6
Tettsted			10
By ≤ 50 000 innbyggere			41
By > 50 000 innbyggere			43
Fylke:			
Oslo			9
Viken			21
Vestfold og Telemark			32
Innlandet			33
Annet			14

^a Utenom besøk på *fast-food* restauranter.

Deskriptive analyser og kommentarer

Tipsbeløpene det er tale om varierer mellom 0 og 100 kr. Det er med andre ord *liten* variasjon i tipsbeløpene, noe som dels er en konsekvens av at prisen på måltidet var den samme for alle respondentene, dvs. 510 kr. Panel A i Tabell 2 viser frem tre aspekter ved tipsbeløpene: snitt, median og andel som har gitt 60 kr eller mer i tips. I snitt gir utvalget drøyt 46 kr i tips, eller om lag 9 % av totalbeløpet på 510 kr. Medianen er lavere, dvs. 40 kr. Det er rimelig å tolke medianen på 40 kr som en avrundningseffekt, dvs. at folk legger på 40 kr for å komme opp til nærmeste runde tall, nemlig 550 kr. (Det er en tilsvarende opphopning, om enn noe mindre, av tips på 90 kr.) Til slutt ser vi at nesten 23 % av de spurte gir 60 kr eller mer i tips.

Tabell 2 Tipsbeløp gitt på restaurant, totalt (Panel A) og brutt ned på to eksperimentgrupper (Panel B). N = 736.

Panel A:	Tipsbeløp: Snitt	Tipsbeløp: Median	Tips ≥ 60 kr: Andel (%)
Hele utvalget	46,44 (21,91)	40	22,6 %
Panel B:			
Brutt ned på eksperimentgrupper			
Nabotips: 0 kr	43,98 (21,85)	40	15,4 %
Nabotips: 60 kr	49,10 (21,69)	50	30,3 %
Differanse	5,12 kr	10 kr	14,9 prosentpoeng

Note: Standardavvik i parentes.

I Panel B brytes de tre aspektene ved tipsene ned på de to eksperimentbetingelsene; altså om bordnaboen gir 0 eller 60 kr i tips for sitt tilsvarende måltid. Vi ser at forskjellen i snitt er på ca. 5 kr og at medianforskjellen er på 10 kr. Videre er forskjellen i sannsynligheten for å gi 60 kr eller mer i tips på ca. 15 prosentpoeng. Kort fortalt peker alle tre forskjeller i samme retning: restaurantgjester som opplever at bordnaboen deres gir 60 kr i tips, gir *selv* mer i tips enn de som opplever at bordnaboen deres gir 0 kr i tips. Og førstnevnte gir også oftere et tips på over 60 kr.

Forskjellen i snitt på 5 kr kan tolkes som en kausaleffekt, selv om den ikke er stor. Hvorfor? Svaret ligger i randomisering av eksperimentbetingelsene. Siden denne randomiseringen, gitt at den var vellykket, sørget for at andre mulige årsaker (observerte og uobserverte) til tipsvariasjon kan utelukkes, så må forskjellen på 5 kr skyldes variasjonen i bordnaboens tipsbeløp. I prinsippet vil heller ikke innføring av kontrollvariabler endre denne snittforskjellen noe særlig, gitt en vellykket randomisering av eksperimentbetingelsene (Angrist & Pischke, 2009; Imbens & Rubin, 2016). Det siste kan også undersøkes ved å teste for *balanse*; dvs. om kontrollvariablene er ukorrelerte med eksperimentvariabelen (Angrist & Pischke, 2015). Via *t*-tester og kjikvadrat-tester har jeg analysert eksperimentvariabelen opp mot alle variablene i Tabell 1, og i ingen av disse er *p*-verdien lavere enn 0,13. Eksperimenttildelingen synes derfor å være balansert. En vellykket randomisering sikrer også at andre, uobserverte kontrollvariabler er tilfeldig fordelt i eksperimentgruppene. Samme resonnement som for størrelsen på tipsene gjelder også for *andelen av tips på 60 kr* eller mer. For tipsforskjellen i median er det statistisk sett litt mer komplisert, men det samme prinsippet gjelder.

Analyser med kontrollvariabler

Til tross for at de svært enkle og deskriptive analysene over utgjør *the main event*, er det vanlig i forbindelse med (survey)eksperimenter å presentere supplerende analyser der kontrollvariabler inngår.⁹ I Tabell 3 er dette gjort for tipsbeløp og andel av tips på 60 kr eller mer. Vi ser at forskjellene mellom eksperimentgruppene på 5 kr og 15 prosentpoeng fra Tabell 2 som forventet består. Ellers går det frem at de med inngående arbeidserfaring fra restaurantbransjen gir mer i tips enn de uten slik erfaring (dvs. nesten 7 kr mer i snitt), og at menn gir noe mer i tips enn kvinner. Videre tipser de som bor landlig mer enn de som bor mer urbant. For det å gi et tips på 60 kr eller mer, er færre av kontrollvariablene relevante.¹⁰ Men vi ser at menn også har 8 prosentpoeng høyere sannsynlighet enn kvinner for å gi dette.

9. Det kan for eksempel skje at randomiseringen *ikke* blir helt vellykket, og da vil innføringen av kontrollvariabler kunne bidra til en sikrere kausalslutning. Se Imbens og Rubin (2016).

10. Jeg rapporterer OLS-regresjon for andelen av tips på 60 kr eller mer versus mindre enn dette av hensyn til transparens, i tråd med Angrist og Pischke (2009). Logistisk regresjon gir som forventet samme resultater i kvalitativ forstand.

Tabell 3 Tipsbeløp og tipsandel over 60 kr etter eksperimentvariabel og kontrollvariabler. OLS-regresjon.

Uavhengige variabler:	Tipsbeløp	Tips ≥ 60 kr
Eksperimenteffekt:		
Nabotips: 0 kr = 0; 60 kr = 1	4,900*** (1,580)	0,150*** (0,031)
Antall restaurantbesøk siste år	0,078** (0,033)	0,000 (0,000)
Arbeidserfaring i restaurant: ^a		
Ja, litt erfaring = 1	0,047 (1,789)	-0,008 (0,036)
Ja, mye erfaring = 1	6,783** (2,806)	0,073 (0,055)
Kjønn:		
Kvinne = 0; mann = 1	3,932** (1,726)	0,081** (0,032)
Alder	0,050 (0,085)	-0,003 (0,001)
Bosted: ^b		
Tettsted = 1	-9,399** (4,088)	-0,073 (0,079)
By ≤ 50 000 innbyggere = 1	-6,636** (3,197)	-0,072 (0,065)
By > 50 000 innbyggere = 1	-3,423 (3,444)	-0,063 (0,071)
Konstant	40,382	0,360
R ²	0,082	0,073
N	736	736

Note: Robuste standardfeil i parentes. Analysene kontrollerer også for fem bostedsfylker (fire dummies).

^a = Referansen er ingen arbeidserfaring i restaurant.

^b = Referansen er å bo landlig.

* $p < 0,07$; ** $p < 0,05$; *** $p < 0,01$ (to-halet test).

Effektheterogenitet

Det bør understrekes at naboeffekten på tipsene over kun har en kausal tolkning for utvalget sett under ett (Imbens & Rubin, 2016). Effekten kan med andre ord variere i størrelse blant ulike subgrupper i dataene. Slik sett *kan* det foreligge effektheterogenitet. Som i annen kvantitativ dataanalyse, er det to måter å avdekke dette på: subgruppeanalyse eller bruk av interaksjonsmodeller. For enkelhetens skyld ser jeg i Tabell 4 på det første. For Panel A er tendensen at naboeffekten på tipsbeløpene er sterkere for kvinner enn for menn, mens det er kjønnsparitet for det å gi tips på 60 kr eller mer. I Panel B går det frem at begge naboeffektene først og fremst gjelder for dem som er 50 år eller eldre. Panel C viser at naboeffekten for det å gi tips på 60 kr eller mer er sterkere blant de oftest restaurantfrekventerende, og Panel D forteller at begge naboeffektene er sterkest for dem uten jobberfaring fra restauranter. I sum er det derfor *tendenser til* effektheterogenitet i utvalget. Men jeg gjør oppmerksom på at en slik konklusjon for en reell studie bør følges opp med formelle signifikanstester og hensyntaken til kontrollvariabler – uten at dette er nødvendig i denne introduksjonssammenhengen.

Tabell 4 Eksperimenteffekter for subgrupper i utvalget. OLS-regresjon.

	Tipsbeløp	Tips \geq 60 kr
Panel A: Eksperimenteffekt av nabotips: 0 kr = 0; 60 kr = 1		
Kjønn = kvinne (n = 412)	6,133***	0,152***
Kjønn = mann (n = 324)	3,95*	0,148***
Panel B: Eksperimenteffekt av nabotips: 0 kr = 0; 60 kr = 1		
Alder \geq 50 år (n = 394)	7,289***	0,220***
Alder $<$ 50 år (n = 342)	2,771	0,067
Panel C: Eksperimenteffekt av nabotips: 0 kr = 0; 60 kr = 1		
Antall restaurantbesøk \geq 15 (n = 378)	5,657***	0,200***
Antall restaurantbesøk $<$ 15 (n = 358)	4,05*	0,088**
Panel D: Eksperimenteffekt av nabotips: 0 kr = 0; 60 kr = 1		
Arbeidserfaring i restaurantbransjen = nei (n = 456)	6,733***	0,187***
Arbeidserfaring i restaurantbransjen = ja (n = 456)	2,342	0,87**

* $p < 0,07$; ** $p < 0,05$; *** $p < 0,01$ (en-halet test).

Noen momenter ved gjennomføring av et surveyeksperiment

Under tas det opp noen momenter som ofte kommer opp i forbindelse med forberedelse til og gjennomføring av et surveyeksperiment – blant annet sett i lys av casestudien presentert over.

Surveyeksperiment versus tradisjonell survey

Når bør man ta i bruk et surveyeksperiment (for et representativt utvalg) fremfor en tradisjonell survey for det samme utvalget? Det første svaret på dette er som alltid at problemstillingen avgjør; i dette tilfellet om problemstillingen lar seg inkorporere i et eksperimentelt design. Det andre svaret er at surveyeksperimentet foretrekkes når det er viktigere å finne den kausale effekten av x på y i populasjonen enn det er å finne det korrekte nivået på x og y i populasjonen. I min studie foran er gjennomsnittet på 46 kr i tips kanskje interessant i seg selv, men likevel underordnet. Det viktige er om, og eventuelt i hvilken grad, de to eksperimentgruppene varierer i forhold til dette snittet.

Representative utvalg versus bekvemmelighetsutvalg

Et utvalg generert ved en tilfeldighetsmekanisme ligger normalt til grunn for et representativt utvalg. Når et surveyeksperiment er basert på et slikt utvalg,¹¹ sikrer sentralgrenseteoremet og signifikanstester at sammenhengen mellom x og y i utvalget gjelder for populasjonen nærmest automatisk. Dette medfører muligens at Mutz (2011) avgrensner surveyeksperimentet til en undersøkelse med (1) tilfeldig trekning fra en klart definert populasjon og (2) randomiserte eksperimentbetingelser. Andre er mer liberale, og peker på at randomiseringen er det mest essensielle (Nock & Guterbock, 2010). I denne forbindelse er det interessant at forskning som har sammenlignet effekter fra surveyeksperimentet basert på bekvemmelig-

11. Trengs det 2 000 tilfeldige respondenter til en tradisjonell survey, kan man for et surveyeksperiment med to eksperimentbetingelser for eksempel trekke 1 000 tilfeldige til å motta survey *i* og 1 000 andre tilfeldige til å motta survey *ii*. For fire eksperimentbetingelser blir det tilsvarende 500 tilfeldige respondenter for survey *i*, *ii*, *iii* og *iv*, gitt et tak på 2 000 respondenter totalt.

hetsutvalg med tilsvarende effekter fra surveyeksperimenter basert på representative utvalg finner små forskjeller, spesielt når det kontrolleres for sosio-demografiske variabler (Jeong et al., 2019; Mullinix et al., 2015; Weinberg et al., 2014). Sagt enklere, taler en del for at en kausal sammenheng for et surveyeksperiment basert på et gjennomtenkt bekvemmelighetsutvalg ofte kan generaliseres til en større populasjon. Dette knytter direkte an til spørsmålet om ekstern validitet, som jeg også skal si mer om under. Nå umiddelbart blir spørsmålet om den lokale årsakssammenhengen avdekket for et selvselektert bekvemmelighetsutvalg fra Facebook kan antyde noe om en mer generell årsakssammenheng blant voksne nordmenn. Et slikt spørsmål er ikke enkelt å besvare, og dette er heller ikke anledningen for å gjøre et helhetlig forsøk. Men de amerikanske forskerne nevnt over antyder som sagt et forsiktig og generelt ja på dette spørsmålet, uten at dette kan sees på som et *carte blanche*. Aaberge og Laake (1984) kan vel også tas til inntekt for et slikt synspunkt, gitt at ikke utvalget er *for* spesielt. Dette er også gode nyheter på et pragmatisk plan, siden et bekvemmelighetsutvalg med stor spredning på kontrollvariabler jevnt over er mindre ressurskrevende å fremskaffe enn et tilfeldig trukket og derfor representativt utvalg.

Oppdagende versus teoritestende surveyeksperimenter

Som nevnt kan et surveyeksperiment ha oppdagende eller teoritestende ambisjoner, mens det i praksis trolig er mer treffende med en skala fra det mer oppdagende til det rent teoritestende. Tas en slik skala i bruk for casestudien over, heller den kanskje mot det teoritestende. Men den har også et oppdagende RCT-innslag i seg, med 60 kr (medisin) versus 0 kr (placebo) som eksperimentbetingelser, noe som igjen understreker at en enten/eller-klassifisering er lite egnet. Hvorvidt man ser på et surveyeksperiment som primært oppdagende eller teoritestende, kan også ha noen konsekvenser for spørsmålet om ekstern validitet. Mer om dette under.

Ekstern og økologisk validitet

Med Shadish, Cook og Campbell (2002, s. 83) ønskes alltid høy ekstern validitet; at det man finner i et surveyeksperiment holder på tvers av personer, settinger og eksperimentbetingelser. Vi har sett at et bekvemmelighetsutvalg, gitt noen forutsetninger, åpner opp for å trekke en kausalsammenheng videre til andre populasjoner, dvs. at dette ikke må forutsette et representativt utvalg. Vi har også sett at et surveyeksperiment med subgruppeanalyse (eller interaksjonsledd) gir mulighet for å avdekke at kausalsammenhenger kan være av varierende styrke for ulike subgrupper i et utvalg. Men ekstern validitet handler også om et surveyeksperiments funn lar seg overføre til andre settinger. Thye (2014) peker på at ekstern validitet spiller ulik rolle for oppdagende og teoritestende eksperimenter. For oppdagende eksperimenter står statistikken i sentrum med hensyn til å generalisere eksperimentet til andre settinger. For teoritestende eksperimenter er det derimot nettopp *teorien* som bærer denne generaliseringen. For surveyeksperimentet over, kan for eksempel et spørsmål bli om naboeffekten lar seg generalisere til andre typer (restaurant)atferd og til andre settinger der folk observerer andre folks atferd, samtidig som de selv blir observert.

Økologisk validitet (*mundane realism*) handler om et eksperiment klarer å gjenskape noe som skjer i det virkelige liv i tilstrekkelig grad. I forhold til vårt surveyeksperiment, er det flere spørsmål som melder seg her. Påvirker det konklusjonene at de spurte ikke er i restaurantsituasjonen, men heller svarer fra en PC, Pad eller en mobiltelefon? Påvirker det konklusjonene at de spurte bruker «papirpenger»? Slike spørsmål har heller ikke enkle svar, og det er i det helt tatt mange kompliserende faktorer fra eksperimentsituasjonen til virkeligheten (jf. Levitt & List, 2007). I forhold til min casestudie over, kan det muligens hevdes at

situasjonen respondentene ble bedt om å se for seg er velkjent for de alle fleste. Videre fikk disse et inngangsspørsmål som beskrev settingen for å sikre at kun de som kjente seg igjen gikk videre med undersøkelsen. For det tredje var de skisserte eksperimentbetingelsene ikke satt på spissen på noen måte. Mer generelt taler mye for at det å skape en troverdig eksperimentsituasjon alltid er fordelaktig, i betydningen at scenariene som folk utsettes for ikke bør oppleves som (for) kunstige (Jackson & Cox, 2013).

Spørsmålet om definering av eksperimentbetingelsene er kanskje like viktig som en realistisk setting. Hammer versus slegge er her en nyttig metafor (Walker & Willer, 2007b), og jeg mener at 0 versus 60 kr i nabotips som kontraster, dvs. hammer, gir høy hverdagsrealisme. En kontrast på 0 versus 100 kr, dvs. slegge, ville gitt lavere hverdagsrealisme, men trolig større naboeffekt. Kontrastene man setter opp i surveyeksperimenter må med andre ord balanse-res mellom ønsket effektstørrelse og krav til hverdagsrealisme. Morales et al. (2017) gir flere råd om hvordan oppnå eksperimentell realisme.

Utvalgsstørrelse og statistisk power

Spørsmålene om hvor stort utvalg som trengs, og hvor mange observasjoner som trengs i hver eksperimentgruppe, er relevante for alle surveyeksperimenter. Slike spørsmål handler om statistisk *power*, dvs. om sannsynligheten for å oppdage en effekt i et utvalg når det faktisk er en effekt i populasjonen (Land & Zheng, 2010). I min casestudie var det 383 respondenter i gruppen som fikk vite at naboen tipset 0 kr, og 353 respondenter i gruppen som fikk vite at naboen tipset 60 kr. Førstnevnte gruppe ga ca. 44 kr i tips i snitt, mens sistnevnte ga ca. 49 kr i snitt; en forskjell (effekt) på 5 kr med en *p*-verdi på 0,0008 (en-halet test). I den grad en så lav *p*-verdi er unødvendig, er det naturlig å spørre om hvor stort utvalg man måtte hatt for å oppnå en effekt med en *p*-verdi på 0,05 – altså et 5 % signifikansnivå. Noen enkle instruksjoner til et statistikkprogram forteller at det da trengs et utvalg på $n = 450$, med 225 respondenter i hver eksperimentgruppe.¹² Slik sett har min studie 286 respondenter «for mye», gitt et ønske om en *p*-verdi på 0,05 (en-halet test). Planlegges flere kontraster, f.eks. et 2×2 design, vokser kravet til antall respondenter. Land og Zheng (2010) gir flere råd og tips for slike *power*-kalkulasjoner.

Sammenfatning og implikasjoner

Jakten på kausaleffekter er sentral i samfunnsforskningen generelt og i sosiologen spesielt. Sosiologene har til nå, med noen unntak, valgt regresjon og beslektede teknikker anvendt på paneldata for å identifisere kausaleffekter. Dette er det ikke noe galt med; tvert imot ville det ha vært en unnlåtelsessynd om for eksempel norske sosiologer ikke hadde tatt i bruk de mulighetene som ligger i norske registerdata. Men noe taler også for at denne utviklingen har hatt som utilsiktet konsekvens at kausalanalyser begrenses til temaer der det finnes registerdata i tidsserier, typisk utdanning, arbeidsliv og helse. Dette er etter mitt syn uheldig av to grunner. For det første er det prinsipielt uheldig at data- og variabeltilgang styrer hva som skal bli utsatt for et kausalt perspektiv. Den andre grunnen er at det finnes mange sosio-

12. For å foreta beregningen, trenger man også standardavvikene i de to eksperimentgruppene. Disse er som kjent 21,85 (for 0 kr i tips) og 21,69 (for 60 kr i tips). I f.eks. Stata skrives da følgende for å beregne utvalgsstørrelsene med 80 % såkalt *power*:

```
power twomeans 43.98 49.10, sd1(21.85) sd2(21.69) onesid
```

Merk at jo større forskjell det er i snitt (eller i andel) mellom to eksperimentgrupper, desto færre observasjoner trengs i de to gruppene for å finne en signifikant forskjell – alt annet likt.

logiske temaer utover produktivitetssfæren som kunne trenge et mer kausalt blikk. Nok å nevne her er fritid, forbruk, sport, politikk, rusmiddelfeltet og det sivile samfunnet.

Mens sosiologene i hovedsak har holdt seg til sine panelregresjoner, har det i søsterdisiplinene økonomi og statsvitenskap pågått en eksperimentell revolusjon de senere årene. Her synes oppfatningen å være at bare et eksperimentelt design/RCT er det som *sikkert* kan identifisere en kausaleffekt. Som Angrist og Pischke (2015, s. 47) skriver lettere lakonisk: «When the path to random assignment is blocked, we look for alternative routes to causal knowledge.» Sagt på en annen måte: Eksperimentelle løsninger bør prøves ut først hvis mulig! I denne artikkelen har jeg på denne bakgrunnen rettet søkelyset mot et eksperimentelt design som kan sies å være en underutnyttet kilde til sosiologisk kausalkunnskap: surveyeksperimentet.

Det er ofte mer effektivt å vise noe frem enn det er å fortelle om det; *show, don't tell!* I tråd med dette, har jeg i denne artikkelen gjennomført og rapportert fra et reelt surveyeksperiment. Denne studien eksemplifiserte hvordan man kan avdekke en kausal naboeffekt for en setting der mennesker blir observert, samtidig som de selv observerer andre. Det skal neppe mye sosiologisk kløkt til å komme opp med lignende sosiale settinger der det eksisterer et slags sosialt press, og da kan et surveyeksperiment være et velegnet design for et mer kausalt blikk. Men det finnes også andre og mer dagsaktuelle muligheter. I denne korona-tiden ønsker for eksempel norske myndigheter å få større anledning til å overvåke folks individuelle mobiltrafikk. For å finne ut om årsakene som eventuelt fører til at folk er mer (eller mindre) positive til en slik tillatelse, kan man tenke seg at man tilfeldig manipulerer ulike frykt-scenarioer i forkant av et spørsmål som kartlegger folks syn på denne tillatelsen.¹³

Gitt premisset om at sosiologisk kausalkunnskap er et aktverdig forskningsformål, er det på bakgrunn av det forannevnte i denne artikkelen i hvert fall fire grunner til at vi bør få se flere sosiologiske surveyeksperimentet fremover:

1. Et eksperimentelt design/RCT er det eneste som tilnærmet sikkert gir kausal kunnskap. Det betyr ikke at RCT er en *magic bullet* (jf. Deaton & Cartwright, 2018), og det betyr heller ikke at RCT og surveyeksperimentet alltid er gjennomførbare og hensiktsmessige i en sosiologisk setting. Men fra det å trekke den motsatte konklusjonen at de bare unntaksvis kan benyttes, noe som trolig er en mer presis beskrivelse av nåsituasjonen, blir fort å skylle barnet ut med badevannet.
2. I et surveyeksperiment er det alltid (den kausale) problemstillingen og teorien som styrer data- og variabelvalg. I forskning som ikke innebærer ny datainnsamling, men som altså benytter foreliggende data, er det neppe til å komme helt utenom i mange situasjoner at det er dataene og variabeltilfanget som i siste instans styrer problemstillingen.
3. Surveyeksperimentet er av to grunner ofte relativt lite ressurskrevende å gjennomføre. For det første trengs ikke nødvendigvis et representativt utvalg, i betydningen et tilfeldig trukket utvalg. For det andre har man sjelden behov for de 1 000 til 1 500 respondentene som typisk intervjues i vanlige surveyer. Ja, i de fleste tilfeller vil man klare seg med et klart lavere antall respondenter enn dette. For mindre (og derfor underfinansierte?) forskningsprosjekter samt masteroppgaver blir dermed surveyeksperimentet kanskje spesielt aktuelt.
4. Legges de tre grunnene over til at sosiologer har tradisjon i å utvikle og analysere vanlige surveyer, er det derfor kun et lite sprang til å gjennomføre surveyeksperimentet. Ofte handler det bare om å lage to nesten identiske versjoner av spørreskjemaet, der x settes til en lav verdi i den ene spørreskjema-versjonen og til en høy verdi i den andre.

13. Jf. Christensen og Aars' studie (2017), nevnt i fotnote 3.

Mot dette bør kanskje også eksperimentelle design få større plass i metodeundervisningen på sosiologiutdanningene enn det som har vært typisk til nå. Slik sett er det muligens et paradoks at mange metodebøker i samfunnsfag nærmest lanserer den statistiske kontrolltankegangen som en erstatning for at man «dessverre» ikke har hatt anledning til – eller later som man ikke har hatt anledning til – å benytte et eksperimentelt design. Kanskje er det på tide å snu på flisa og la eksperimentet først vise veien til kausalslutningen, og så la den statistiske kontrolltenkningen overta i tilfellene der eksperimenter ikke lar seg gjennomføre i praksis. I denne forbindelsen tillater jeg meg den personlige spekuleringen at «der eksperimenter ikke lar seg gjennomføre», i for stor grad er basert på sosiologisk vanetenkning innenfor fagtradisjonen man er opplært i. Håpet med denne artikkelen er at sosiologer fremover oftere stiller seg følgende spørsmål når de, eller deres masterstudent, ønsker å finne ut av om og eventuelt hvordan x påvirker y i en sosiologisk kontekst uten tilgang til registerdata: Hvordan kan jeg/vi best designe et surveyeksperiment som besvarer dette spørsmålet? En slik inngang til et forskningsprosjekt eller masteroppgave vil også understøtte fyndordet om at teori og problemstilling bør legge føringer på forskningsdesign samt data- og variabeltilgang – og ikke omvendt.

Om artikkelen

Takk til Erik Haugom, Martin Rønningen og to anonyme konsulenter for verdifulle kommentarer til tidligere utkast av denne artikkelen.

Referanser

- Aaberge, R., & Laake, P. (1984). Om statistiske teoriar for tolking av data. *Tidsskrift for samfunnsforskning*, 25(2), 165–186.
- Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Angrist, J. D. & Pischke, J.-S. (2015). *Mastering 'Metrics. The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.
- Bakken, A., Andersen, P. L., Frøyland, L. R. & Abebe, D. S. (2019). Rekkefølgeeffekter i spørreundersøkelser blant ungdom. *Norsk sosiologisk tidsskrift*, 3(1), 66–82. <https://doi.org/10.18261/issn.2535-2512-2019-01-05>
- Barabas, J. & Jerit, J. (2010). Are Survey Experiments Externally Valid? *American Political Science Review*, 104(2), 226–242. <https://doi.org/10.1017/s0003055410000092>
- Birkelund, G. E., Rogstad, J., Heggebø, K., Aspøy, T. M. & Bjelland, H. F. (2014). Diskriminering i arbeidslivet. Resultater fra randomiserte felteksperiment i Oslo, Stavanger, Bergen og Trondheim. *Sosiologisk tidsskrift*, 22(4), 352–382.
- Christensen, D. A. & Aars, J. (2017). Nordmenns holdninger til telefonavlytting: Resultater fra et surveyeksperiment. *Tidsskrift for Samfunnsforskning*, 58(2), 191–209. <https://doi.org/10.18261/issn.1504-291x-2017-02-03>
- Deaton, A. & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2–21. <https://doi.org/10.3386/w22595>
- Helland, L. (2016). *Årsak og tilfeldighet. Modeller, eksperimenter og atferd i samfunnsvitenskapene*. Oslo: Cappelen Damm Akademisk.
- Imbens, G. W. & Rubin, D. B. (2016). *Causal Inference for Statistics, Social, and Biomedical Sciences*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/cbo9781139025751>

- Jackson, M. & Cox, D. R. (2013). The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology*, 39, 27–49. <https://doi.org/10.1146/annurev-soc-071811-145443>
- Jeong, M., Zhang, D., Morgan, J. C., Cornacchione Ross, J., Osman, A., Boynton, M. H., Mendel, J. R. & Brewer, N. T. (2019). Similarities and Differences in Tobacco Control Research Findings from Convenience and Probability Samples. *Annals of Behavioral Medicine*, 53(5), 476–485. <https://doi.org/10.1093/abm/kay059>
- Land, K. C. & Zheng, H. (2010). Sample Size, Optimum Allocation, and Power Analysis. I P.V. Marsden & J. D. Wright. (Red.), *Handbook of Survey Research. Second Edition* (s. 199–219). UK: Emerald.
- Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174. <https://doi.org/10.1257/jep.21.2.153>
- Lynn, M. (2015). Service gratuities and tipping: A motivational framework. *Journal of Economic Psychology*, 46, 74–88. <https://doi.org/10.1016/j.joep.2014.12.002>
- Morales, A. C., Amir, O. & Lee, L. (2017). Keeping It Real in Experimental Research – Understanding When, Where, and How to Enhance Realism and Measure Consumer Behavior. *Journal of Consumer Research*, 44(2), 465–476. <https://doi.org/10.1093/jcr/ucx048>
- Morgan, S. L. & Winship, C. (2015). *Counterfactuals and Causal Inference. Methods and Principles for Social Research* (2. utg.). New York, NY: Cambridge University Press. <https://doi.org/10.1017/cbo9781107587991>
- Mullinix, K. J., Leeper, T. J., Druckman, J. N. & Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science*, 2(2), 109–138. <https://doi.org/10.1017/xps.2015.19>
- Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400840489>
- Nock, S. L. & Guterbock, T. M. (2010). Survey Experiments. I P. V. Marsden & J. D. Wright (Red.), *Handbook of Survey Research*. (2. utg., s. 837–864). UK: Emerald.
- Saunders, S. G. & Lynn, M. (2010). Why tip? An empirical test of motivations for tipping car guards. *Journal of Economic Psychology*, 31(1), 106–113. <https://doi.org/10.1016/j.joep.2009.11.007>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Thye, S. R. (2014). Logical and Philosophical Foundations of Experimental Research in the Social Sciences. I M. Webster & J. Sell (Red.), *Laboratory Experiments in the Social Sciences*. (2. utg., s. 53–82). Saint Louis, US: Academic Press.
- Tufte, P. A. (2013). Å studere sosiale årsakssammenhenger. *Sosiologi i dag*, 43(3), 91–109.
- Ugreninov, E. & Birkelund, G. E. (2013). Naturlige eksperiment. *Sosiologi i dag*, 43(3), 65–89.
- Walker, H. A. & Willer, D. (2007a). Experimental methods. I G. Ritzer (Red.), *The Blackwell Encyclopedia of Sociology*, Vol. 3 (s. 1537–1541). Oxford, UK: Blackwell.
- Walker, H. A. & Willer, D. (2007b). Experiments and the Science of Sociology. I M. Webster & J. Sell (Red.), *Laboratory Experiments in the Social Sciences* (s. 25–54). Burlington, MA: Elsevier.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A Review. *Social Science Research*, 38(3), 505–520. <https://doi.org/10.1016/j.ssresearch.2009.03.004>
- Weinberg, J. D., Freese, J. & McElhattan, D. (2014). Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsome-Recruited Sample. *Sociological Science*, 1, 292–310. <https://doi.org/10.15195/v1.a19>