

Søren Serritzlew

Det politologiske eksperiment: hvorfor, hvornår og hvordan?¹

Eksperimentet bygger på en simpel logik, som genfindes mange steder i statskundskabens arsenal af metoder, nemlig at effekten af en uafhængig variabel bedst kan forstås, hvis man kan variere den systematisk, samtidig med at andre variabler holdes konstante. Det særlige ved eksperimentet er, at forskeren selv kan styre variablerne og dermed sikre sig, at de rette variabler varierer og holdes konstante, som de skal. Det betyder, at eksperimentet kan bruges til at løse metodiske problemer med kontrol, måling og endogenitet, som ofte er yderst vanskelige at håndtere med andre metoder. Artiklen giver et bud på, hvornår eksperimenter især er velegnede og endda uomgængelige. Samtidig gives der en række råd til, hvordan et politologisk eksperiment kan gennemføres i praksis.

Eksperimentet er på mange måder den mest intuitive og grundlæggende kilde til videnskabelige data. Den er intuitiv, fordi fremgangsmåden i eksperimentet er, at man selv tilrettelægger et forløb på en sådan måde, at svaret kan ses så tydeligt som muligt. Den er grundlæggende, fordi andre metoder i vidt omfang bygger på dens indbyggede logik, nemlig at variere én variabel mens alle andre holdes konstante. Vil man eksempelvis undersøge betydningen af anonymitet i fangernes dilemma, handler det om at sammenligne resultaterne af to varianter af spillet, hvor den eneste forskel er, om deltagerne kan se hinanden. Hvis der er forskel i resultaterne, har man stor sikkerhed for, at det skyldes den uafhængige variabel. I traditionelle spørgeskemaundersøgelser og case-studier er man henvist til at søge svar på sine spørgsmål ved at studere begivenheder, som man ikke selv har kontrol over. Det giver anledning til en række svære metodiske problemer, som eksperimentet er velegnet til at løse.

Et eksperiment er karakteriseret ved, at forskeren for at studere en kausal-sammenhæng bevidst griber ind i eller påvirker verden og dermed kontrollerer den måde, hvorpå begivenhederne udfolder sig.² I sin mest rendyrkede form finder eksperimentet sted i et laboratorium, hvor alle påvirkninger af eksperimentets deltagere, subjekterne, kan kontrolleres. Hvor de fysiske rammer er helt afgørende i naturvidenskabelige eksperimenter, vil et undervisnings-lokale på et universitet – eller et hvilket som helst andet rum af passende størrelse og med nogenlunde ro – være fuldt tilstrækkeligt som laboratorium for de fleste politologiske eksperimenter.

En nær slægtning til disse eksperimenter er surveyeksperimentet (se Petersen et al., 2007), hvor det kun er nødvendigt at kontrollere den information, som svarpersonen præsenteres for, og hvor undersøgelsen derfor kan ske skrift-

ligt via et spørgeskema eller mundtligt via et telefoninterview. Påvirkningen af subjektet er af relativt lav intensitet og kan ske ved, at formuleringen af spørgsmål varieres, således at forskellene i svar kan henføres hertil (Sniderman og Grob, 1996: 393).

Denne artikel fokuserer på det klassiske eksperiment, hvor subjekterne fysisk er til stede, og hvor alt (herunder typisk både information, omgivelser og interaktion) i princippet kontrolleres af forskeren.³ Formålet er dels at diskutere, hvornår eksperimentet er velegnet og endda nødvendigt, dels – på basis af især rational choice-inspirerede eksperimenter – at give nogle praktiske råd om, hvordan eksperimenter gennemføres.

Eksperimenter behøver hverken være komplicerede eller besværlige. Thomas C. Schelling er en sand mester i at opstille yderst simple, men instruktive eksperimenter. Tag fx. hans undersøgelse af konflikter i *The Strategy of Conflict*.⁴ Her beskæftiger Schelling sig med spørgsmålet om, hvorfor og hvordan konflikter kan begrænses, herunder hvorfor interessekonflikter mellem lande kun sjældent eskalerer til krig. Han argumenterer for, at en del af forklaringen er, at interessekonflikter ikke ender med endeløs uenighed om, hvem der skal have hvad, men oftere ved at de involverede parter enes om et *fokalpunkt*, dvs. en løsning, der forekommer „oplagt“, også selvom den stiller den ene part bedre end den anden. Spørgsmålet er, hvordan man kan undersøge, om ideen har noget for sig. I en traditionel empirisk undersøgelse ville det være nødvendigt at finde et eksempel på en konflikt mellem to eller flere parter, nøje undersøge deres mål og interesser, om de truer eller påvirker hinanden og derefter finde ud af, hvilken løsning parterne vælger. Det vil i praksis være meget vanskeligt at afgøre, om den valgte løsning skyldes, at den var „oplagt“ i Schellings forstand og ikke et andet forhold. Schelling afklarar det over fire sider i sin bog ved hjælp af otte usædvanligt simple eksperimenter. Et af dem lyder således:

„You and your two partners (or rivals) each have one of the letters A, B, and C. Each of you is to write these three letters, A, B, and C, in any order. If the order is the same on all three of your lists, you get prizes totaling \$6, of which \$3 go to the one whose letter is first on all three lists, \$2 to the one whose letter is second, and \$1 to the person whose letter is third. If the letters are not in identical order on all three lists, none of you gets anything. Your letter is A (or B, or C); write here the three letters in the order you choose: __, __, __“ (Schelling, 1960: 60f.).

C har naturligvis en interesse i en rækkefølge, hvor bogstavet C kommer først, ligesom både A og B ønsker at begynde med deres bogstav. Der findes med andre ord ikke noget kompromis, som stiller de tre rivaler lige godt. Der er tale om en situation med en klar interessekonflikt. Hvis Schelling har ret i, at der alligevel kan etableres samarbejde, fordi parterne vil vælge et fokuspunkt –

altså en oplagt løsning – så vil de tre rivaler ende med at kunne koordinere deres adfærd, således at de sikrer sig en gevinst. Hvis ikke må man forudse, at de tre rivaler vil stå på hver sit, at rækkefølgerne ikke bliver ens, og at de dermed taber gevinsten på gulvet. Schellings eksperiment viste meget klart, at samarbejdet faktisk bliver etableret: Ni af de 12 A'ere, der deltog i eksperimentet, valgte rækkefølgen A, B, C; det samme gjorde ti ud af 12 B'ere og 14 af 16 C'ere. En benhård fordelingskonflikt blev altså uden videre løst, fordi der var en oplagt løsning; alle kan se, at rækkefølgen A, B, C er et naturligt valg, også selvom denne løsning er til fordel for A og til ugunst for C.

Schelling har ikke forsøgt at gøre eksperimentet virkelighedsnært. Han benyttede i stedet et simpelt spil til at undersøge dele af en problemstilling, der i sidste ende handler om konflikter mellem lande. Det gælder helt generelt, at eksperimenter ikke behøver være – og sjældent er – realistiske. Det er snarere en styrke ved eksperimentet, at det er muligt at udelukke alt det, der hører med til begivenheder i virkeligheden uden for laboratoriet, og som uden tvivl vil få stor betydning for resultatet. I fordelingskonflikter vil det blandt andet dreje sig om magtforhold, forskelligt informationsniveau, tradition, aftaler og trusler. Eksemplet viser også, at eksperimenter er lette at håndtere, netop fordi de er designede til at belyse et bestemt spørgsmål på den bedst mulige måde. I eksemplet er der hverken behov for at foretage statistisk kontrol for tredjevariabler eller for at bekymre sig om måleusikkerhed, for der er ingen.

I næste afsnit diskuterer jeg nærmere, hvornår eksperimenter især er velegnede. Det sker ved først at se nærmere på tre metodiske problemer, som er yderst vanskelige at løse i traditionelle empiriske undersøgelser. Dernæst diskuteres eksperimentets interne og eksterne validitet, og sidst i artiklen gives der en række råd om, hvordan eksperimenter kan gennemføres i praksis.

Først vil jeg introducere det eksperiment, som især vil blive brugt som eksempel undervejs. Det blev designet til at teste, om rammebudgettering virker efter hensigten, altså om politisk vedtagelse af en ramme for udgifterne i løbet af budgetforhandlinger medvirker til at nedbringe udgifterne. Hovedhypotesen er, at effekten af rammebudgettering afhænger af beslutningstagernes præferencer for budgettet og udgifternes størrelse. Afhængig af præferencerne vil rammerne virke forskelligt: I nogle situationer (når præferencesammensætningen er *kontraktiv*) vil rammerne virke dæmpende på udgifterne, men i andre situationer (når præferencesammensætningen er *ekspansiv*) vil de virke modsat hensigten og altså øge udgifterne (se Serritzlew, 2003, 2005). Det giver, som det fremgår af tabel 1, fire kombinationsmuligheder, som blev afprøvet i hver sin runde (kaldet *trials*) af eksperimentet. Hvis hypotesen er rigtig, vil udgifterne være mindre i *trial* I end i III og større i *trial* II end i IV.

Table 1. Typer af beslutninger i rammebudgetteringeksperimentet

	Kontraktive præferencer	Ekspansive præferencer
Brug af ramme	<i>Trial I</i>	<i>Trial II</i>
Ingen brug af ramme	<i>Trial III</i>	<i>Trial IV</i>

Eksperimentet som løsning på tre vanskelige metodiske problemer
Enhver undersøgelse af en hypotese om et kausalforhold vil i større eller mindre grad støde på tre potentielt alvorlige problemer, nemlig kontrolproblemet, måleproblemet og endogenitetsproblemet.

Kontrolproblemet opstår, når tredjevariabler, der har betydning for den afhængige variabel, ikke inkluderes i analysen. Konsekvensen er, at effekten af de uafhængige variabler, som er med i analysen, vil kunne blive fejlestimeret. Et velkendt eksempel er sammenhængen mellem antallet af storke og fødselsraten i et område. Undersøger man denne sammenhæng, viser det sig, at der er en stærk, positiv sammenhæng, således at antallet af fødsler ser ud til at vokse med antallet af storke. Problemet er imidlertid, at en relevant tredjevariabel er udeladt, nemlig urbanisering. I landområder er fødselsraten højere end i byer, og samtidig er det også her, storke trives bedst. Når der kontrolleres for urbanisering, forsvinder den tilsyneladende sammenhæng mellem fødselsraten og antallet af storke. I dette eksempel er kontrolproblemet selvindlysende.

I praksis kan det være langt mere subtilt. For det første er der altid en risiko for at overse en væsentlig tredjevariabel. Jo mere komplekst et fænomen, man studerer, jo større er risikoen for at overse faktorer af væsentlig betydning. Dette problem kan ikke udelukkes i ikke-eksperimentelle studier. For det andet kan det være svært at måle de tredjevariabler, som man på forhånd finder relevante at inddrage. I rammebudgetteringseksemplet var kontrolproblemet stort. I en traditionel test af effekten af rammer er det relevant at kontrollere for, om der samtidig med indførelsen af rammebudgettering er sket andre organisatoriske eller institutionelle reformer. En ændring kommer sjældent alene, og derfor vil det være nyttigt at kontrollere for, om de observerede effekter kunne skyldes nogle af disse snarere end rammerne. Problemet er, at det er yderst vanskeligt at få overblik over, hvilke øvrige organisatoriske ændringer der måtte have fundet sted.

I eksperimentet kan problemet løses ved hjælp af tre strategier. Den første strategi er randomisering. Denne strategi bør benyttes under alle omstændigheder. Ved at lade tilfældigheden råde i eksempelvis inddelingen af subjekter i grupper og i rækkefølgen af de *trials*, som eksperimentet består af, vil indvirkningen af en række potentielt forstyrrende kontrolvariabler automatisk

minimeres. Hvis en kontrolvariabel helt tilfældigt er knyttet til de forskellige observationer i et eksperiment, vil den ikke systematisk påvirke resultaterne. Forestiller man sig, at subjekternes alder har betydning for de beslutninger, der træffes i rammebudgetteringseksperimentet, og sikrer man, at subjekterne fordeles tilfældigt, således at der ikke er nogen systematisk forskel på deltagerne alder i de forskellige *trials*, som senere skal sammenlignes, vil alderen ikke systematisk påvirke resultaterne. Den tilfældige variation vil naturligvis give anledning til usikkerhed, men da den ikke er systematisk, vil almindelige statistiske tests helt som i traditionelle kvantitative studier kunne afsløre, om resultaterne er så robuste, at man kan udelukke, at de skyldes tilfældigheder. Fordelen ved denne strategi er, at den også kan håndtere ukendte og uobserverbare kontrolvariabler. Det vil stort set altid være sådan, at adfærd og svar fra personer, der deltager som subjekter i et eksperiment, svarpersoner i en spørgeskemaundersøgelse eller interviewpersoner i et casestudie, vil være påvirket af en række personlige forhold, holdninger og egenskaber, som det vil være umuligt at få fuld indsigt i. Hvis subjekterne i eksperimentet fordeles tilfældigt, vil dette problem være minimeret.

Den anden strategi til løsning af kontrolproblemet er at variere kontrolvariablen. Hvis man har en forventning om, at kønssammensætningen blandt beslutningstagerne har betydning for budgetforhandlingernes forløb, kan der tages højde for dette ved at gennemføre eksperimentet for grupper med forskellig kønssammensætning. Ved at sammenligne resultaterne for grupper med blandet sammensætning med grupper bestående eksempelvis alene af kvinder, vil man kunne se, om kønssammensætningen faktisk har betydning. Denne strategi indebærer altså, at kontrolvariablen behandles som en uafhængig variabel. Fordelen herved er, at man får meget detaljeret information om variabelens indvirkning på sammenhængen. Skulle der vise sig at være interaktion, således at rammebudgetteringens virkning afhænger af, hvordan gruppen er sammensat, så vil eksperimentet vise det. Ulempen er, at antallet af uafhængige variabler har betydning for, hvor mange *trials* eksperimentet vil bestå af. Hvis hver uafhængig variabel kan antage to værdier, vil antallet af *trials* i eksperimentet være 2^k , hvor k er antallet af uafhængige variabler. I rammebudgetteringseksperimentet er der som udgangspunkt to uafhængige variabler, der hver kan antage to værdier. Det betyder, at $2^2 = 4$ typer af *trials* skal sammenlignes, sådan som det fremgik af tabel 1. Med køn som endnu en uafhængig variabel når vi op på otte *trials*. Eftersom antallet af *trials* forøger eksperimentets kompleksitet og varighed, er der meget klare grænser for, hvor mange uafhængige variabler der kan medtages.

Den tredje strategi består i at holde kontrolvariablen konstant. I et veldesignet eksperiment sker det stort set af sig selv. Tager man problemet med kontrol for øvrige organisatoriske forandringer, er alt, hvad der kræves i et experi-

ment, at man kun foretager de ændringer i reglerne, som man er interesseret i at måle effekten af. Derved sikres, at alle forskelle, der måtte være mellem gruppernes beslutninger i de forskellige *trials*, kan henføres til de uafhængige variabler (i eksemplet præferencer og brugen af rammer) og ikke andre organisatoriske eller institutionelle forskelle. Styrken ved eksperimentet er, at denne strategi kan bruges i forbindelse med næsten alle de kontrolvariabler, som i naturligt forekommende begivenheder ikke kan styres. Ulempen er, at strategien overhovedet ikke giver nogen information om kontrolvariablen. Efter som formålet jo ikke er at undersøge disse kontrolvariabler, er dette normalt kun en mindre omkostning. Undtagelsen er, når kontrolvariablen interagerer med en af de uafhængige variabler. I det tilfælde vil eksperimentets resultater blive påvirket af den måde, hvorpå variablen holdes konstant. Var politikernes præferencer i rammebudgetteringseksperimentet blevet behandlet som en konstant kontrolvariabel, ville eksperimentet således have givet et misvisende resultat. Hvis præferencerne var holdt konstant som kontraktive, ville rammerne se ud til altid at virke efter hensigten, altså besparende, og hvis præferencerne omvendt blev holdt konstant som ekspansive, ville rammerne se ud til altid at virke udgiftsdrivende. Da præferencerne betinger rammernes effekt er begge konklusioner forkerte.

Måleproblemet opstår, når der er knyttet stor usikkerhed til observation af de centrale uafhængige variabler. En sådan usikkerhed medfører nemlig, at effekterne underestimeres (King et al., 1994: 158ff.). I en statistisk undersøgelse vil risikoen være stor for at begå en Type II-fejl (ikke at forkaste H_0 -hypotesen, selvom den faktisk er forkert, dvs. selvom der i virkeligheden findes en sammenhæng). Det sker, fordi den store måleusikkerhed i sig selv reducerer koefficienterne og deres signifikans. Problemets omfang afhænger naturligvis af, hvordan data indsamles. I registerstudier vil måleproblemet typisk være relativt mindre end i survey- og interviewbaseret dataindsamling.

I studiet af rammebudgettering optrådte måleproblemet også. For at teste hypotesen om, at effekten af rammebudgettering er betinget af beslutningstageres præferencer, er det nødvendigt at måle dem. En mulig datakilde kunne være at spørge politikerne selv ved hjælp af en survey. Den almindelige fremgangsmåde er her at lade politikerne angive på en fempunktsskala, om de i forhold til sidste år ønsker at bruge flere eller færre ressourcer på de forskellige udgiftsområder. Målingen af præferencer er altså behæftet med betydelig usikkerhed. I studiet af rammebudgettering var usikkerheden så stor, at hypotesen hverken kunne af- eller bekræftes ved hjælp af spørgeskemadata (Serritzlew, 2004: 28, 94).

I et eksperiment kan måleproblemet reduceres eller helt elimineres. I rammebudgetteringseksemplet er måleproblemet som nævnt knyttet til politikernes præferencer. Fordelen ved eksperimentet er, at præferencerne kan fastlægges

præcist. I rammebudgetteringseksperimentet var det nødvendigt at sammenligne beslutninger truffet af grupper med ekspansive og kontraktive præferencer, med og uden brug af rammer, som det fremgår af tabel 1 ovenfor. Tabellen viser de to uafhængige variabler, som hver har to værdier (kaldet *treatments*). Dermed er det, som det vil fremgå nedenfor, oplagt at benytte et faktorielt design. Det giver fire kombinationer, som forventes at føre til forskellige resultater i eksperimentet. De fire celler i tabellen udgør eksperimentets *trials*, og som tidligere nævnt forventes grupperne at ende med at beslutte sig for højere udgifter i *trial III* end i *trial I*, for her forventes rammerne at virke dæmpende på udgifterne, og omvendt lavere udgifter i *trial IV* end II, idet rammerne her forventes at øge udgifterne.

Præferencerne fastlægges ved, at subjekterne instrueres i, at de skal arbejde for nogle bestemte mål (her for udgifterne), der svarer præcist til den præferencesammensætning, man ønsker at undersøge. Det sker ved at sikre, at subjekterne har helt bestemte, på forhånd definerede mål. Subjekterne instrueres i, at de vil modtage en belønning for at deltage i eksperimentet, og at den vil afhænge af, hvor tæt resultatet af beslutningsprocessen kommer på det mål, de skal arbejde for. Hvis belønningen er tilstrækkeligt attraktiv, vil målet – eller præferencen – simpelt hen blive overtaget af subjektet. Dermed kendes subjektets præferencer helt præcist. Samtidig kan præferencerne styres og designes, således at testen bliver så stærk som mulig, fx. ved som i rammeeksemplet at svare til de fire *trials* i tabel 1. Eksperimentet er altså særlig velegnet, når måleusikkerheden er stor. Hvis det er helt umuligt at måle centrale uafhængige variabler, vil eksperimentet kunne være den eneste løsning.

*Endogenitetsproblemet*⁵ opstår, når den afhængige variabel påvirker den uafhængige (King et al., 1994: 185f.). I så fald vil korrelationen mellem den uafhængige og den afhængige variabel ikke kun skyldes, at den uafhængige variabel forårsager den afhængige. Undtagen når det klart kan fastslås, at den uafhængige variabel må være før i tid end den afhængige (det gælder eksempelvis køn), er endogenitet altid et potentielt problem. Den ideelle løsning er selv at styre den uafhængige variabel. Det er imidlertid oftest udelukket, når man undersøger naturligt forekommende begivenheder. Spørgsmålet er derfor sjældent, om der er et endogenitetsproblem, men derimod om problemet er så stort, at analyserne giver misvisende resultater.

Dette problem er særligt stort, når man beskæftiger sig med en problemstilling, hvor aktørerne selv kan påvirke den uafhængige variabel. Det er meget ofte tilfældet i politologiske problemstillinger: Hvis man vil undersøge, om den for tiden meget populære *Lean*-metode til effektiviseringer i den offentlige sektor faktisk virker, vil det kunne være misvisende blot at se på sammenhængen mellem brugen af *Lean* og effektiviteten. Det kunne jo være, at aktørerne vælger at indføre *Lean* pga. lav effektivitet. Det vil trække fra i den estimerede

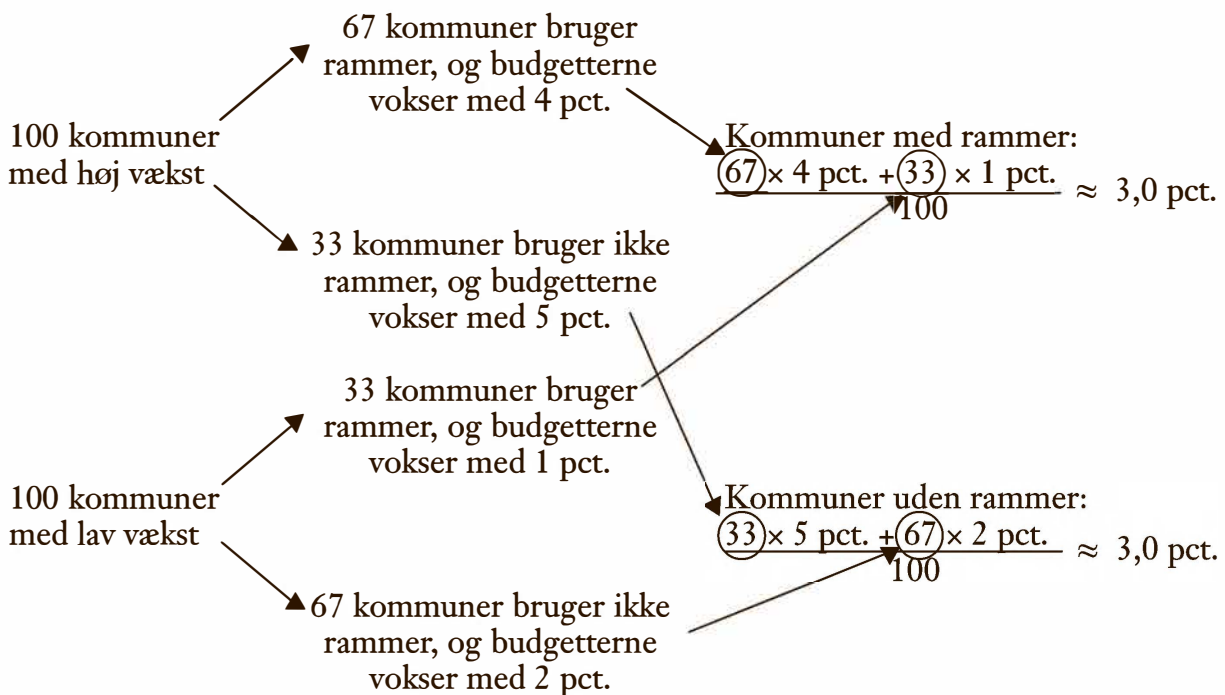
effekt. Hvis man kan konstatere, at demokratiske lande sjældnere er i krig, må man overveje, om sammenhængen kunne være omvendt, altså at demokrati lettere opstår og opretholdes i lande, der ikke er i krig. Hvis man vil undersøge, om friskoler er bedre til at uddanne elever, må man tage højde for, at det ikke er tilfældigt, hvilke børn der går i friskole. Måske er det sådan, at de dygtigste børn lidt oftere sendes i friskole. Det vil kunne resultere i en markant, positiv estimeret effekt, selv uden at friskolerne leverer bedre undervisning.

Også i rammebudgetteringseksemplet er endogenitetsproblemet stort. Rammer for de samlede udgifter indføres typisk for at begrænse udgifterne (Budgetdepartementet, 1983; Kommunernes Landsforening, 1991, 2001). Man må derfor forvente, at rammebudgettering hyppigere indføres, hvis væksten i budgetterne i en periode har været relativt høj. Når effekten af rammebudgettering herefter undersøges, vil resultaterne være uklare. Et eksempel illustrerer problemet. Antag at vi har 200 kommuner, hvoraf den ene halvdel har en høj vækst i udgifterne på 5 pct., og den anden en lav vækst på 2 pct. Det antages videre, at rammebudgettering er et effektivt middel og altid nedbringer væksten i udgifterne med 1 procentpoint. Hvis halvdelen af kommunerne bruger rammebudgettering, og hvis det er helt tilfældigt, hvilke der gør det, må vi forvente, at de kommuner, der benytter rammer, vil have en vækst på 2,5 pct. i gennemsnit (50 kommuner har en vækst på 4 pct., fordi de har høj vækst og bruger rammer, og 50 en vækst på 1 pct., fordi de har lav vækst og bruger rammer). De kommuner, der ikke bruger rammer, vil i gennemsnit have en vækst på 3,5 pct. (nemlig 50 kommuner med en vækst på 5 pct. og 50 kommuner med en på 2 pct.). Her vil man ganske korrekt konkludere, at brugen af rammer medfører, at udgifterne vokser 1 procentpoint mindre.

Endogenitetsproblemet opstår, hvis den afhængige variabel påvirker den uafhængige, altså hvis væksten i udgifterne har indflydelse på valget af, om der skal bruges rammer eller ej. Lad os antage, at sammenhængen er således, at to ud af tre af de kommuner, der oplever høj vækst, vil vælge at bruge rammer, mens dette kun vil gælde for én ud af tre af de kommuner, der oplever lav vækst. Nu vil regnestykket se anderledes ud, jf. figur 1.

Ved at sammenligne de kommuner, der benytter rammebudgettering med dem, der ikke gør, kommer man frem til, at rammebudgettering ikke har nogen effekt, selvom rammerne i det tænkte eksempel faktisk nedbringer væksten med 1 procentpoint. Hvis den afhængige variabel har større effekt på den uafhængige, vil det endda komme til at se ud til, at rammerne har den modsatte effekt. Hvis 80 pct. af højvækstkommunerne og kun 20 pct. af lavvækstkommunerne vælger at bruge rammer, vil en tilsvarende beregning vise, at rammebudgettering i gennemsnit fører til en vækst på 3,4 pct., mens de kommuner, der ikke benytter rammer, har en gennemsnitlig vækst på 2,6 pct. Det

Figur 1. Illustration af endogenitetsproblemet. Eksemplet er konstrueret



vil i dette tilfælde være lige så fristende som forkert at konkludere, at rammer fører til højere udgifter.

Hvis den uafhængige variabel kan styres eksogent, vil endogenitetsproblemet forsvinde helt, for hvis det er tilfældet, vil den afhængige variabel ikke kunne påvirke den uafhængige. Et af de definerende træk ved eksperimentel metode er netop, at den tillader fuld kontrol med den uafhængige variabel (Friedman og Sunder, 1994: 22; Kinder og Palfrey, 1993a: 6). Forskeren bestemmer selv værdierne på den uafhængige variabel, og dermed kan den – modsat i studier af naturligt forekommende begivenheder – umuligt være påvirket af den afhængige variabel.

Intern og ekstern validitet

Muligheden for at manipulere de uafhængige variabler og kontrollere for en lang række forskellige tredjevariabler er både en styrke og en svaghed ved eksperimentel metode. Styrken er, at den interne validitet er høj. Cook og Campbell definerer intern validitet som „the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of a cause“ (1979: 37). Når de centrale uafhængige variabler kendes helt præcist og varieres på en kendt måde, og når alle relevante tredjevariabler samtidig holdes konstante, kan man med stor sikkerhed konkludere, at eventuelle systematiske ændringer af den afhængige variabel skyldes variation i de uafhængige variabler.

Til gengæld er den eksterne validitet ofte relativt lav (Kinder og Palfrey, 1993a: 26-31). Cook og Campbell (1979: 37) definerer ekstern validitet som „the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings and times“. Ekstern validitet vedrører altså, om resultaterne af en undersøgelse kan generaliseres til at gælde ud over den kontekst, hvori undersøgelsen er foretaget. Det vil være tilfældet, hvis undersøgelsen har inddraget mange forskelligartede cases, dels fordi resultatet dermed i sig selv gælder et relativt bredt empirisk felt, dels fordi sammenhængen har vist sig at være robust over for stor variation i, hvad der undersøges. I sådanne tilfælde vil det være rimeligt at forvente, at resultatet også vil kunne findes uden for undersøgelsespopulationen. Derfor er den eksterne validitet ofte høj i kvantitative studier. Hvis resultatet derimod er opnået i særlige, kontrollerede omgivelser med bestemte typer af deltagere og uden forstyrrende faktorer, vil den eksterne validitet typisk være lav. Her vil man ikke uden videre kunne slutte sig til, at tilsvarende resultater vil kunne findes i andre sammenhænge.

Hvis sigtet først og fremmest er empirisk, således at man ønsker at kunne forklare et bestemt fænomen så godt som muligt, er den eksterne validitet vigtigst. Når man med traditionelle kvantitative data undersøger, om en teoretisk forudsigelse rammer plet, testes i virkeligheden både argumentet og dets antagelser. Hvis forudsigelsen viser sig at være forkert, kan det være svært at se, om det skyldes problemer med selve det teoretiske argument eller med antagelserne, dvs. argumentets forudsætninger. Hvis formålet snarere er at undersøge det teoretiske argument, er den interne validitet vigtigere. Her vil det være uhensigtsmæssigt, hvis argumentet afvises, blot fordi antagelserne ikke passer i den bestemte empiriske kontekst, hvori testen blev foretaget.

Eksperimentel metode minder om komparative casestudier ved, at den interne validitet kan være ganske høj. I eksperimentel metode kan den høje interne validitet opnås ved at kontrollere for alle relevante tredjevariabler. I casestudier kan man til gengæld grundigt undersøge og sætte sig ind i hele konteksten og derved tage højde for potentielt forstyrrende faktorer. Til gengæld er den eksterne validitet oftest tilsvarende lav. Forskellen mellem de to tilgange er, at eksperimenter kun involverer få, på forhånd definerede variabler, mens casestudier er velegnede til at håndtere et stort antal variabler, der ikke nødvendigvis kendes i forvejen. Eksperimenter har derfor normalt et konfirmativt sigte, mens komparative casestudier ofte har et større eksplorativt element. Dermed er der også ligheder mellem eksperimentet og kvantitative metoder, hvor sigtet også typisk er konfirmativt, og hvor variablerne må defineres før data indsamles. Forskellene og lighederne er skitseret i tabel 2.

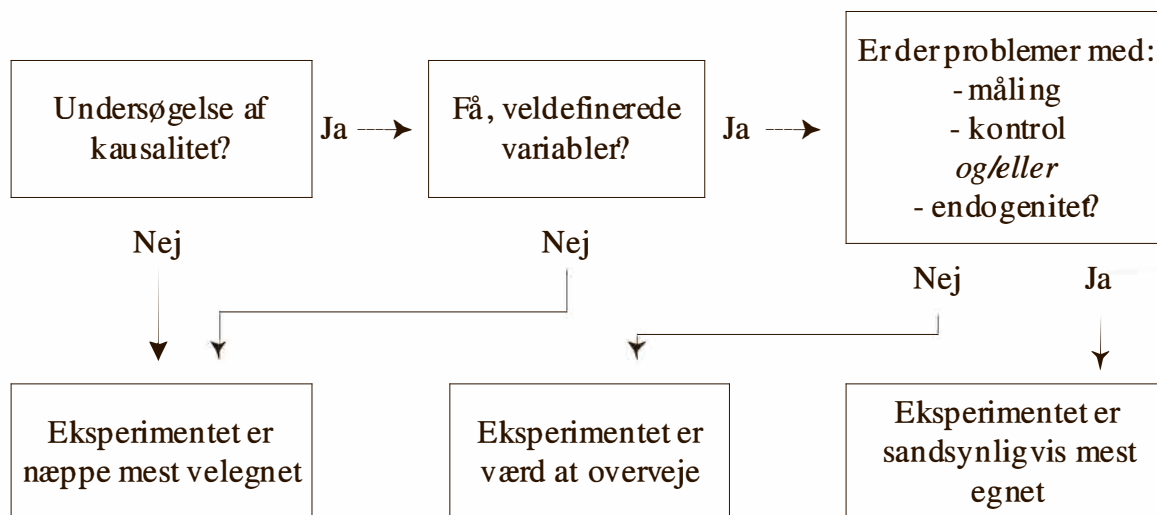
Tabel 2. Eksperimentet sammenlignet med kvantitativ metode og komparative casestudier

	Få, veldefinerede variabler	Mange, komplekse variabler
Intern validitet relativt højere end ekstern validitet	Eksperimenter	Komparative casestudier
Ekstern validitet relativt højere end intern validitet	Kvantitativ metode	

Eksperimentet indtager altså en vigtig plads blandt de metodiske tilgange. Det adskiller sig fra kvantitative metoder ved at sikre høj intern validitet, selv når der er tale om et komplekst argument, eller når en eller flere vigtige uafhængige variabler eller kontrolvariabler ikke kan måles med tilstrækkelig præcision. Kombinationen af den høje interne validitet og det konfirmative sigte vil sjældent kunne opnås med en traditionel kvantitativ eller en kvalitativ tilgang.

Dermed kan der som en sammenfatning af dette og det foregående afsnit i figur 2 gives et bud på, hvornår eksperimentet er en særlig velegnet metode.

Figur 2. Hvornår er eksperimentet velegnet?



Eksperimentet er sjældent velegnet, hvis man ikke ønsker at teste en kausal-sammenhæng. Uden en helt klar opfattelse af, hvad der er uafhængige og afhængige variabler, og af hvordan de påvirker hinanden, vil det være vanskeligt overhovedet at designe et eksperiment. Hvis kausalsammenhængen involverer et stort antal variabler, vil en eksperimentel test hurtigt kræve alt for mange forskellige sammenligninger. Men hvis begge disse betingelser er opfyldt, er

eksperimentet altid værd at overveje – og det kan vise sig at være den eneste farbare vej, hvis der er betydelige måle-, kontrol- eller endogenitetsproblemer. Spørgsmålet er så, hvad man gør, hvis man står med en sådan problemstilling. Det vil næste afsnit give et bud på.

Eksperimenter i praksis

Når eksperimentet skal udformes, skal der tages stilling til en række praktiske spørgsmål og dilemmaer. Det er ikke muligt at gennemgå dem alle her. Jeg vil dog sige et par ord om nogle af de vigtigste overvejelser om først instruktioner, dernæst design og antallet af deltagere og endelig om motivation og belønninger.

Når instruktionerne skal udformes, er den bedste strategi at tage udgangspunkt i et eksperiment, som tidligere er blevet gennemført med succes.⁶ Det er der to grunde til. For det første vil man dermed kunne sikre sammenligneligheden med andre studier. Det er en fordel, når resultaterne skal fortolkes og sammenlignes med, hvad litteraturen i øvrigt siger om emnet. For det andet vil det spare en masse kostbare fejltagelser. Skulle man være i den sjældne situation, at man beskæftiger sig med et område, som slet ikke er undersøgt eksperimentelt før, er det alligevel en god ide at lade sig inspirere af andres erfaringer. En række standardløsninger vil ofte kunne overføres. Det gælder især selve instruktionerne, som ofte vil kunne formuleres ved at lave passende ændringer i, hvad andre tidligere har brugt.⁷ Det siger sig selv, at reglerne skal være klare, præcise og letforståelige, samtidig med at det ikke på nogen måde antydes, hvilken adfærd der teoretisk forventes. Et hyppigt praktisk problem er, at subjekterne bruger forskellig tid på at læse og forstå reglerne for eksperimentet. Derfor er det en dårlig ide blot at omdele et sæt regler til subjekterne og derefter lade dem gå i gang, efterhånden som de er klar. Det fører til, at eksperimentet begynder, før man har sikret sig, at alle har forstået, hvad det hele går ud på. Disse problemer kan undgås ved at afsætte rigelig tid til gennemgang af eksperimentet og til efterfølgende spørgsmål fra deltagerne. Det er også en god ide at oplæse instruktionerne.

Hvad angår designet, er det første spørgsmål, hvilke og hvor mange uafhængige variabler der skal inkluderes. Den første overvejelse vedrører dermed også, hvilke variabler der skal behandles som kontrolvariabler og hvordan. Det skal altså afklares, om kontrolvariablerne skal varieres (og dermed behandles som uafhængige variabler), holdes konstante eller randomiseres. Som det fremgik ovenfor, er det en forudsætning, at antallet af uafhængige variabler er begrænset. I praksis vil det være vanskeligt at håndtere mere end tre i samme eksperiment.

Det fremgik også, at variablerne skal være veldefinerede. Det skyldes, at de uafhængige variabler skal operationaliseres, dvs. omsættes til *treatments*. I

rammebudgetteringseksemplet vedrørte den ene uafhængige variabel brugen af rammebudgettering. Det er et eksempel på en variabel, der er så veldefineret, at den uden videre lader sig omsætte til to *treatments*, nemlig (1) brug af rammebudgettering og (2) ingen brug af rammebudgettering. For mange variabler vil der i operationaliseringen kunne vælges mellem adskillige *treatments*. Ønsker man eksempelvis at undersøge gruppestørrelsens betydning for de beslutninger, der træffes, vil den uafhængige variabel kunne antage alle mulige forskellige værdier. Eftersom det er ret omkostningstungt at anvende mange *treatments* for hver uafhængig variabel, er det normalt fornuftigt at nøjes med to *treatments*.⁸ For gruppestørrelsens vedkommende altså en lille og en stor gruppe, hvor man sikrer sig, at forskellen er så tilpas stor, at man på forhånd kan være ret sikker på tydeligt at kunne spore en eventuel effekt i resultaterne.

Den næste serie af overvejelser vedrører, hvordan subjekterne bedst kan udsættes for de forskellige *treatments*. Det er her nyttigt at sondre mellem på den ene side de tilfælde, hvor man tester én uafhængig variabel og de tilfælde, hvor man tester flere. Hvis kun én variabel med to *treatments* skal testes, vil eksperimentet bestå af to *trials*, en for hver *treatment*. Det er uden betydning, om *treatments* rettes mod enkeltsubjekter (som i Schellings eksperiment fra indledningen, hvor det erindres, at *treatment* var, om et enkeltsubjekt var af type A, B eller C) eller mod grupper af subjekter (som i rammebudgetterings-eksemplet om grupperes beslutninger om udgifter). Det har altså ikke betydning for designet, når der i det følgende både tales om, at *treatments* rettes mod enkeltsubjekter og mod grupper.

Når eksperimentet skal belyse mere end en uafhængig variabel, vokser – som det fremgik ovenfor – antallet af *trials*. I et fuldt faktorielt design undersøges samtlige kombinationer af de uafhængige variabler (Carlsmith et al., 1976: 243). Tabel 1 ovenfor viser et 2 x 2 faktorielt design. Her giver to variabler med hver to *treatments* fire forskellige kombinationer. Effekten af hver af de uafhængige variabler kan ses på to måder. Hvis vi kalder resultatet af *trial* I for I, *trial* II for II osv., vil effekten af rammer, givet at præferencerne er kontraktive, være I minus III. Effekten af rammer, givet ekspansive præferencer, vil være II minus IV. Omvendt kan effekten af præferencer, givet en bestemt værdi på rammevariablen, aflæses ved at sammenligne resultaterne i henholdsvis *trial* I med II (effekten er altså II minus I) og III med IV (altså IV minus III). Hvis de to variabler interagerer, vil effekten af den ene uafhængige variabel afhænge af værdien af den anden. Det var netop forventningen i eksemplet: Effekten af rammer forventedes at afhænge af præferencerne. Interaktionseffekten vil kunne beregnes som (II-IV)-(I-III). Hvis værdien er 0, har variablen samme effekt, uanset værdien af den anden uafhængige variabel, og der er dermed ikke tale om interaktion. Er værdien positiv, betyder det, at effekten af den ene uafhæn-

gige variabel (rammerne) afhænger positivt af den anden uafhængige variabel (præferencerne) – og omvendt hvis værdien er negativ.

I det faktorielle design findes der et trick (kaldet *fractional factorial design*), som kan reducere behovet for *trials*. Hvis man på forhånd kan udelukke interaktion mellem de uafhængige variabler (eller hvis man ved, at interaktionseffekten er så ubetydelig, at den ikke er af interesse), ville effekten af to variabler kunne undersøges ved blot tre *trials*. Effekten af den første variabel (i eksemplet brug af rammer) vil, jf. tabel 1, kunne undersøges med enten *trial* I og III eller *trial* II og IV. Vælger man at gennemføre *trial* I og III, vil den anden variabel (præferencerne) kunne undersøges ved sammenligning af *trial* I og II. *Trial* IV kan altså undværes. I designs med flere uafhængige variabler eller med mere end to *treatments* pr. variabel vil samme trick kunne bruges til at begrænse antallet af *trials* – samtidig med at man kan få indsigt i udvalgte interaktionseffekter (se Friedman og Sunder, 1994: 28-29).

Den sidste grundlæggende overvejelse om designet er, om effekten af de forskellige *treatments* skal undersøges ved at sammenligne, hvordan det samme subjekt reagerer på forskellige *treatments* (*within subjects*-designet), eller hvordan forskellige subjekter reagerer på samme *treatment* (*between subjects*-designet).

I *within subjects*-designet udsættes samme subjekt eller samme gruppe af subjekter for alle *trials* (Carlsmith et al., 1976: 266). I Schelling-eksperimentet ville det indebære, at alle subjekter prøver at spille rollen som A, B og C. Fordelen ved dette design er, at forskelle imellem de forskellige *trials* for det samme subjekt ikke kan skyldes alder, køn, motiver osv. Ulempen er, at subjektets adfærd i de forskellige *trials* i et *within subjects*-design vil kunne være gensidigt afhængig. I Schellings eksempel vil det således være ret oplagt, at det samme subjekt igen og igen vil vælge den samme bogstavrækkefølge. *Within subjects*-designet vil altså være uegnet. Men overvej i stedet følgende eksempel: I det såkaldte skoeksperiment er formålet at undersøge holdbarheden af forskellige typer skosåler:

„The focus is the sole material, a treatment variable with two levels: old and new. Measured wear varies considerably, mostly from subjects' different activities and habits: Some boys are couch potatoes, others ride scooters using a shoe for a brake. Clever experimental design prevents these nuisances from obscuring the focus variable's effects: Each boy gets a pair of shoes with one sole of new material and the other sole of old“ (Friedman og Sunder, 1994: 25).

Her vil subjektets adfærd i forhold til de to *treatments* ikke være gensidigt afhængig, samtidig med at det kan udelukkes, at forskellene i slid skyldes forskelle mellem subjekter.

I *between subjects*-designet (Friedman og Sunder, 1994: 25) varierer *treatments*

mellem forskellige subjekter eller grupper. Her vil den samme gruppe altså kun opleve en *trial*. Fordele og ulemper er her præcis modsat. På den ene side kan det udelukkes, at adfærden i forskellige *trials* er indbyrdes afhængig, men på den anden side er der en risiko for, at forskelle i adfærd skyldes irrelevante forskelle mellem grupperne. Hertil kommer en ulempe af praktisk art, som nok er hovedårsagen til, at *within subjects*-designet er mere udbredt end *between subjects*-designet. *Between subjects*-designet kræver nemlig væsentligt flere subjekter end *within subjects*-designet, hvor alle kan udsættes for de forskellige *trials*. Der kan læses mere om disse to grundlæggende designs og varianter heraf i fx. Carlsmith et al., (1976: 268-273), Kidder og Judd (1986: 84ff.) og Friedman og Sunder (1994: 26ff.).

Der skal også træffes en række beslutninger om deltagerne i eksperimentet, herunder hvor mange og hvilke subjekter der skal rekrutteres. Antallet afhænger helt af, hvordan eksperimentet er designet. I et 2 x 2 faktorielt *within subjects*-design, hvor hver *trial* kræver fem subjekter (altså som i rammebudgetterings-eksperimentet), vil fire observationer, dvs. en for hver af de fire *trials*, kræve fem subjekter. Hvor mange observationer, der er nødvendige for at kunne konkludere med en vis vægt, afhænger af, hvor klare effekterne er, og hvor stor sikkerhed der kræves. Hvis effekterne er meget tydelige, vil enhver tvivl forsvinde selv med relativt få observationer. I rammebudgetterings-eksperimentet var det tilstrækkeligt med 45 subjekter (Serritzlew, 2003: 266). Hvis effekterne er svage, er det nødvendigt med væsentligt flere for med rimelig sikkerhed at kunne konkludere, om de uafhængige variabler har den forventede effekt eller ej.

Når man overvejer, hvem der skal rekrutteres, er det værd at erindre, at et eksperiment skal afdække effekten af de uafhængige variabler. Det er altså ikke afgørende, om subjekterne har en særlig relevant baggrund. Rammebudgetterings-eksperimentet handler om, hvordan politiske beslutningstagere reagerer på bestemte regler for budgetlægningen. Men det betyder ikke, at det er en fordel, at subjekterne i eksperimentet er politikere, beslutningstagere eller på anden måde i deres daglige liv er knyttet til det fænomen, som studeres. Hvis den uafhængige variabel har den forventede effekt, må den kunne ses, uanset om subjekterne er politikere eller ej. Tværtimod vil det kunne være en ulempe, hvis subjekterne på forhånd har grundig viden om eller erfaringer med det fænomen, der studeres. Hvis det er tilfældet, vil resultaterne kunne blive påvirket af subjekternes eventuelle forudindtagethed. Man kunne således let forestille sig, at politikere konfronteret med et eksperiment om betydningen af rammer i højere grad kender rammers formål og – bevidst eller ubevidst – lader dette kendskab påvirke deres adfærd.

Det er ikke kun subjekternes forhåndskendskab, der kan have betydning. Det kan også alle andre ikke-manipulerbare egenskaber ved subjekterne, altså eksempelvis alder, selvtillid og uddannelse. Sådanne variabler kaldes *subjekt-*

variabler (Carlsmith et al., 1976: 13). I praksis gennemføres mange eksperimenter med studerende som subjekter. De fleste forhandlingseksperimenter er således gennemført med økonomistuderende som deltagere. En serie eksperimenter har vist, at økonomistuderende er mindre tilbøjelige til at samarbejde (Marwell og Ames, 1981). Dette mønster er ikke overraskende. Økonomistuderende lærer, at samarbejde ikke altid er den økonomisk rationelle adfærd. Konsekvensen af denne subjektvariabel er, at man i mange forhandlingseksperimenter risikerer at undervurdere omfanget af samarbejde. Problemet er dog mindre, end det lyder, for påvirkningen må forventes at være konstant for forskellige *treatments*. Skævheden vil altså forsvinde, når effekterne af disse sammenlignes. Man skal blot undgå, at subjektvariablerne er korreleret med bestemte *treatments*. Det ville ske, hvis økonomistuderende kun deltager i én bestemt *trial* i eksperimentet. De mulige systematiske fejl kan konverteres til tilfældige fejl ved at randomisere, dvs. at fordele deltagerne tilfældigt i de grupper, der indgår i eksperimentet (Kidder og Judd, 1986: 72).

Det er selvfølgelig ikke tilfældigt, at subjekter ofte rekrutteres blandt studerende. Eftersom det gælder om at finde en større gruppe mennesker, der let kan samles i et lokale i en god times tid, er det simpelthen oplagt. Uanset om subjekterne er studerende eller ej, er det vigtigt at være opmærksom på, om subjekterne kender hinanden i forvejen. I et eksperiment, hvor interaktionen mellem subjekterne er afgørende, vil indbyrdes venskaber eller bekendtskaber kunne påvirke resultaterne. Det understreger igen fordelene ved at randomisere.

Deltagerne skal motiveres til at deltage og levere en seriøs indsats. Hvis der er rigeligt med ressourcer, kunne en tommelfingerregel være, at deltagerens belønning skal svare til, hvad de ville kunne tjene ved at arbejde i stedet for.⁹ Der er dog gennemført mange vellykkede eksperimenter for færre midler. Belønningen tjener to formål. For det første letter den arbejdet med at rekruttere. For det andet bruges den til at styre subjekternes præferencer eller mål. I de fleste eksperimenter indgår deltagerens målrettede adfærd som et centralt element. I forhandlingseksperimenter vil deltagerne således stort set altid skulle arbejde for et bestemt resultat eller beslutning, og det interessante er så at se, hvilken kollektiv beslutning der træffes. Udfordringen er at sikre, at alle subjekter virkelig bestræber sig på at opnå det på forhånd definerede mål eller resultat. Schellings resultater om koordination på rækkefølgen A, B, C er kun interessant, fordi subjekterne har forskellig interesse i rækkefølgerne. Belønningen bruges til at inducere målene i subjekterne. Ved at knytte belønningen til om målet opnås, sikres det, at subjekterne *faktisk* har de mål, som eksperimentet forudsætter. I Schellings eksempel skete det simpelthen ved at belønne deltagerne med \$3, hvis deres eget bogstav kom først, \$2 hvis det kom som nummer to, og kun \$1 hvis det kom sidst.

Belønninger sikrer imidlertid ikke uden videre, at mål og præferencer indu-

ceres i subjekterne. Der er tre forudsætninger: monotonicitet, saliens og dominans (Friedman og Sunder, 1994: 13). Monotonicitet indebærer, at subjekterne altid foretrækker at få mere af belønningen. Hvis man, måske i et forsøg på at undgå problemer med beskatning af gevinsten, belønner deltagerne med præmier i form af bøger, mad og drikke eller lignende, vil de på et tidspunkt ikke længere værdsætte flere gevinster. I det øjeblik vil målene i eksperimentet ophøre med at være relevante for subjektet. Den bedste løsning er derfor at bruge kontante præmier. Salienskravet betyder, at der skal være en relevant sammenhæng mellem subjektets adfærd og muligheden for at opnå belønningen. Hvis belønningen udbetales uafhængigt af eksperimentets udfald, vil mål eller præferencer ikke blive induceret i subjekterne. I forhandlingseksperimenter vil salienskravet typisk kunne opfyldes ved at lade det enkelte subjekts belønning afhænge af, hvor stor forskel der er på hans ellers hendes mål og gruppens beslutninger i de enkelte *trials*, helt som i Schellingeksemplet. Kravet om dominans indebærer, at belønningen skal være så stor, at den forekommer vigtigere for subjekterne end andre hensyn. Et potentielt problem er, at subjekterne vil kunne opfatte eksperimentet som et slags spil, hvor der kan være hensyn til bekendte, muligheder for at drille andre deltagere og en fristelse til at se, hvad der sker, hvis man gør noget helt ulogisk. Subjekterne skal ikke blot opfatte eksperimentet som noget, de gør for sjov. Det er afgørende, at belønningen er stor nok til at sikre deres interesse i at deltage koncentreret i eksperimentet.

Eksperimentet som normal metode

Eksperimentet er ikke særlig udbredt i dansk politologi. Det er en skam, for metoden har, ligesom andre og mere udbredte metoder, sine særlige fordele. Eksperimentet er velegnet, når man vil undersøge en kausalsammenhæng og er interesseret i få, veldefinerede variabler. Hvis man tilmed støder på store kontrol-, måle- eller endogenitetsproblemer, er eksperimentet næsten uomgængeligt. Samtidig er eksperimenter hverken særlig vanskelige eller dyre at gennemføre. Eksperimenter indgår derfor også som en del af standardrepertoiret i særligt amerikansk politologisk forskning, og de publiceres stadig oftere i de bedste tidsskrifter (Petersen et al., 2007: 7f.). Samtidig kendes og bruges disse studier også hyppigt i dansk politologi. Mange vil kunne genkende eksempelvis Miller og Whitford (2002), Iyengar og Kinder (1987) og Milgram (1974) som centrale eksperimentelle studier med generel relevans inden for tre forskellige grene af statskundskaben. Eksperimentet er med andre ord ikke specielt eksotisk, men en normal, velafprøvet metode, som ikke kan undværes, hvis man vil have mulighed for at få klare svar på en lang række problemstillinger.

Noter

1. Forfatteren takker Rune Slothuus, Mads Leth Jakobsen, Pia Vedel Ankensen og Lotte Bøgh Andersen for gode og grundige kommentarer til tidligere versioner af denne artikel.
2. Der findes mange definitioner af, hvad et eksperiment er. Kinder og Palfreys definition lyder således: „In a fully realized experiment, the investigator seizes control over the production of settings, the creation of treatments, and the scheduling of observations“ (1993a: 6f.). McGraw definerer eksperimentet på denne måde: „Structurally, experiments are marked by a deliberate intervention in the natural, ongoing state of affairs. (...) Functionally, experiments test propositions about cause and effect relationships – whether the ‘intervention’ or ‘treatment’ is causally responsible for some observed outcome ‘in the restricted sense that the change would not have occurred without the intervention’“ (1996: 770).
3. En god introduktion til surveyeksperimentet findes i Sniderman og Grob (1996) og Petersen et al., (2007). Eksempler på danske surveyeksperimenter kan findes i et temanummer af *Politica* (nr. 1, marts 2007) redigeret af samme forfattere.
4. Eksemplet er også forklaret i Kagel og Roth (1995: 12).
5. Dette og det følgende afsnit bygger på Serritzlew (2004: 31ff.).
6. Gode eksempler inden for studiet af forhandlings- og beslutningsprocesser er Fiorina og Plott (1978), Miller og Oppenheimer (1982), McKelvey og Ordeshook (1984) og Miller og Whitford (2002).
7. Der er optrykt eksempler på instruktioner i appendix II i Friedman og Sunder (1994). Instruktionerne vil også ofte fremgå (desværre typisk i forkortet form) af de konkrete studier. Der kan findes mange referencer til gode eksperimenter i fx Kinder og Palfrey (1993b), Friedman og Sunder (1994: 145ff.) og i Kagel og Roth (1995).
8. Antallet af *treatments* har betydning for, hvor mange grupper der skal sammenlignes. Antallet af grupper er mere præcist m^k , hvor m er antallet af *treatments* på hver uafhængig variabel, og k er antallet af uafhængige variabler. Med fire *treatments* og tre variabler vil hele $4^3 = 64$ forskellige kombinationer skulle sammenlignes. Med to *treatments* ville antallet blot være $2^3 = 8$.
9. Friedman og Sunder (1994: 50) anbefaler endda, at den gennemsnitlige gevinst for studerende skal ligge 50-100 pct. over, hvad man ville kunne tjene i et tilsvarende tidsrum som studentermedhjælper.

Litteratur

- Budgetdepartementet (1983). *BRU-rapporten. Rapport fra arbejdsgruppen vdr. budget- og bevillingssystemet*, København: Budgetdepartementet, MF-centralen.
- Carlsmith, J. Merrill, Phoebe C. Ellsworth og Elliot Aronson (1976). *Methods of Research in Social Psychology*, Reading: Addison-Wesley.
- Cook, Thomas D. og Donald T. Campbell (1979). *Quasi-Experimentation. Design & Analysis Issues for Field Settings*, Chicago: Rand McNally College Publishing Company.
- Fiorina, Morris og Charles Plott (1978). „Committee Decisions under Majority Rule: An Experimental Study“, *American Political Science Review*, Vol. 72, pp. 575-598.
- Friedman, Daniel og Shynam Sunder (1994). *Experimental Economics*, Cambridge: Cambridge University Press.

- Iyengar, Shanto og Donald R. Kinder (1987). *News That Matters. Television and American Opinion*, Chicago: University of Chicago Press.
- Kagel, John H. og Alvin E. Roth (1995). *Handbook of Experimental Economics*, Princeton: Princeton University Press.
- Kidder, Louise H. og Charles M. Judd (1986). *Research Methods in Social Relations*, New York: Holt, Rinehart and Winston.
- Kinder, Donald R. og Thomas R. Palfrey (1993a). „On Behalf of Experimental Political Science“, pp. 1-39 in Donald R. Kinder og Thomas R. Palfrey (eds.), *Experimental Foundations of Political Science*, Ann Arbor: The University of Michigan University Press.
- Kinder, Donald R. og Thomas R. Palfrey (eds.) (1993b). *Experimental Foundations of Political Science*, Ann Arbor: The University of Michigan University Press.
- King, Gary, Robert O. Keohane og Sidney Verba (1994). *Designing Social Inquiry*, Princeton: Princeton University Press.
- Kommunernes Landsforening (1991). *Styring i udvikling*, København: Forlaget Kommuneinformation.
- Kommunernes Landsforening (2001). *Budgetlægningsmetoder – Appendiks til „Nye Vinde“*, udgivet på <http://www.kl.dk>.
- Marwell, Gerald og Ruth E. Ames (1981) „Economists free ride, does anyone else?“, *Journal of Public Economics*, Vol. 15, pp. 295-310.
- McGraw, Kathleen M. (1996). „Political Methodology: Research Design and Experimental Methods“, pp. 769-786 in Robert E. Goodin og Hans-Dieter Klingemann (eds.), *A New Handbook of Political Science*, Oxford: Oxford University Press.
- McKelvey, Richard D. og Peter C. Ordeshook (1984). „An Experimental Study of the Effects of Procedural Rules on Committee Behavior“, *Journal of Politics*, Vol. 46, pp. 182-205.
- Milgram, Stanley (1974). *Obedience to Authority*, New York: Harper & Row.
- Miller, Gary J. og Joe A. Oppenheimer (1982). „Universalism in Experimental Committees“, *American Political Science Review*, Vol. 76, No. 3, pp. 561-574.
- Miller, Gary J. og Andrew Whitford (2002). „Trust and Incentives in Principal-Agent Negotiations: The ‘Insurance-Incentive’ Tradeoff“, *Journal of Theoretical Politics*, Vol. 14, No. 2, pp. 231-267.
- Petersen, Michael B., Rune Slothuus, Rune Stubager og Lise Togeby (2007). „Eksperimenter: Et redskab i politologens værktøjskasse?“, *Politica*, 39. årgang, nr. 1, pp. 5-13.
- Schelling, Thomas C. (1960). *The Strategy of Conflict*, Cambridge: Cambridge University Press.
- Serritzlew, Søren (2003). „Kan udgiftsrammer begrænse væksten i budgetterne?“, *Politica*, 35. årgang, nr. 3, pp. 255-273.
- Serritzlew, Søren (2004). *Offentlig budgetlægning i et institutionelt perspektiv*, Århus: Politica.
- Serritzlew, Søren (2005). „The Perverse Effect of Spending Caps“, *Journal of Theoretical Politics*, Vol. 17, No. 1, pp. 75-105.
- Sniderman, Paul M. og Douglas B. Grob (1996). „Innovations in Experimental Design in Attitude Surveys“, *Annual Review of Sociology*, Vol. 22, pp. 377-399.