

TILSVAR TIL REPLIKK OM RETTSBEVISSTHETSUNDERSØKELSENE

AV FØRSTEAMANUENSIS LEIF PETTER OLAUSSEN

Replikken fra mine kolleger, Flemming Balvig, Helgi Gunnlaugsson og Henrik Tham (BGT), til min artikkel innbyr til avklaring og utdyping på enkelte punkter.

Å gjøre vold på datamateriale

Når alle avkryssinger for reaksjoner i det norske materialet blir tatt hensyn til, ser det ut til å være rimelig godt samsvar mellom dommerpanelenes straff og medianverdien for de reaksjoner respondentene seg ville gitt, målt med en skjønnsmessig straffeskala. Hovedproblemet er om denne konklusjonen skyldes *bias* som er bygd inn i straffeskalaen, eller om den gir et rimelig korrekt (*unbiased*) resultat. BGT anfører ingen argumenter for at skalaen gir et biased resultat, og metodestudien jeg redegjør for i min artikkel, viser at respondentene tenker *additivt* om fengsel, behandling og økonomisk kompensasjon, slik jeg har forutsatt i straffeskalaen. Det berøres ikke av BGT, og det overrasker at de ikke omtaler at også Flemming Balvig (2006 og 2010) har forsøkt å løse problemet med to avkryssinger ved å lage en straffeskala som han bruker i analyser som danner grunnlag for vidtrekkende konklusjoner. Betyr BGTs replikk at Balvigs analyser og konklusjoner nå ses bort fra fordi han i likhet med meg ”gjør vold på det innsamlede datamateriale”? Hvorfor angir ikke BGT et eneste argument for at resultatene blir mer pålitelige dersom vi – i strid med at respondenter tenker additivt – velger å *se bort fra* (evt. gi verdien *null* til) alle andre avkryssede reaksjoner enn ubetinget fengsel, slik det blir gjort av Jerre og Tham (2010)? Hvis jeg fulgte denne framgangsmåten for det norske materialet med over 75 % svar med to avkryssinger, ville jeg sett bort fra mer enn 42 % av reaksjonene som respondentene har krysset av for. Jeg savner BGTs *argumenter* som grunngir at dette i mindre grad gjør vold mot innsamlet datamateriale enn bruk av min straffeskala.

Jeg hadde to muligheter til å vurdere om det nevnte norske resultatet synes å være biased. For det første er det samstemmighet mellom dette resultat og det jeg får om jeg bare legger fordelingen av respondentenes avkryssinger for ubetinget fengselsstraff til grunn, og bruker *modus* som mål for hvor tyngdepunktet i fordelingen ligger. Jeg forstår ikke hvorfor BGT avstår fra å fortelle åpent om det er samsvar eller ikke mellom *modus* for ubetinget fengselsstraff i respondentenes svar og dommerpanelenes straff i det danske, svenske og islandske materialet. Jeg er ganske sikker på at det har allmenn interesse. For det andre stemmer resultatet med at det er utbredt *enighet* mellom fagdommere og lekdommere når det gjelder straffutmåling i domstolene. Resultatet fra en fersk norsk spørreskjemaundersø-

kelse blant våre lekdommere tyder ikke på at enigheten skyldes urimelig press fra fagdommernes side (NOU 2011:13, s. 72-74). Dette øker min tillit til at den skjønsmessige straffeskalaen som mine analyser bygger på, ikke gir et fortegnnet bilde.

Datamaterialet for alle land viser at respondentene våre ville gitt ubetinget fengsel i mindre grad enn våre domstoler. Etter metodestudien vinteren 2011 vet jeg at en viktig grunn til dette er at respondenter reduserer fengselsstraffen når behandling og økonomisk kompensasjon blir valgt i tillegg. Vi kunne sikkert redusert respondentenes tilbøyelighet til å krysse av for ubetinget fengsel enda mer ved å gjennomføre BGTs forslag om å tillate avkryssing for alle svaralternativer, ved å føre opp dødsstraff som mulig svaralternativ, eller ved bare å tillate avkryssing for ubetinget fengsel hvis respondenten ikke kunne finne på andre former for straff. Jeg er kanskje veldig gammeldags som forsker, men å lage et forskningsdesign som systematisk vil øke sannsynligheten for et bestemt resultat (lavest mulig oppslutning om fengselsstraff) var verken min hensikt da jeg i 2006 tok initiativ til vårt første møte for å lage en felles nordisk undersøkelse eller senere.

Nærhetshypotesen og testingen av den

Vi er alle enige i at det er noe virkelighetsfjernt å pådømme en sak ut fra lesing av en kortfattet fremstilling på et stykke papir, slik våre respondenter gjorde både i postsurveyen og i fokusgruppene, sammenlignet med situasjonen i en rettssal. Derfor ble fokusgruppedeltakerne vist en filmatisert versjon av en rettssak om en av de sakene de allerede hadde avsagt sin dom i, basert på papirfremstilling. Filmen skulle skape en ”nærhet” som ligner den man finner i en domstol. BGT hevder uriktig at jeg ikke definerer hva jeg mener med ”nærhet”. Det fremgår av min artikkel foran at det dreier seg om en avgrenset ”hypotese om at et møte mellom dommere og tiltalte – det medmennesket som dommeren skal gi en straff – vil påvirke dommeren i retning av å gi litt mildere straff enn når straff gis uten at et slikt møte finner sted, som ved vår bruk av spørreskjema. Vi ville forsøke å belyse dette spørsmålet i fokusgruppedelen av undersøkelsen.” Dette er identisk med Balvig (2010, s. 130). Og videre skriver jeg: I vårt design ”er det lagt til grunn at spørsmålet ... vil bli besvart ved å sammenligne fokusgruppedeltakernes forslag til straff *før* og *etter* at de hadde sett filmen. Men er dette en holdbar forutsetning?”, spør jeg.

I de foreløpige rapportene vi publiserte i august 2010 er det ingen vurderinger av denne forutsetningen. Balvig (2010) oppsummerer sine helt ukritiske sammenligninger av fokusgruppedeltakernes reaksjonsvalg basert på papirfremstilling av saken med deres reaksjonsvalg etter å ha sett filmen, slik: Når ”danskerne... med utgangspunkt i en filmet retssag og derved et ”personlig” møte med tiltalte,

forurettede og andre vidner – har skullet ta stilling til, hvordan de syntes, der burde straffes, har [de] en tendens til at være mildere i deres holdninger, end når de skal ta stilling til en skriftlig resumeret sak.” (Balvig 2010, s.130.) Balvig er så sikker på at det er fullstendig uproblematisk å sammenligne avsaatte ”dommer” på basis av et skriftlig resymé med de som ble gitt på basis av film, at han med stor nøyaktighet har beregnet størrelsen på effekten av å vise film (Balvig 2010, Oversikt 6.3.13, s. 138), og han finner at alle utregnede effekter er statistisk signifikante. Han går ikke dypt inn i å forklare hvorfor fokusgruppedeltakerne i følge hans effektberegninger ”gjennomgående blir mildere, når man skal dømme under rettsagslignende forhold, end når man skal dømme på baggrund af et skriftlig resymé. Vi formoder, at en væsentlig del af forklaringen ligger i det personlige møde. Jo større nærhed, desto større tilbakeholdenhet med straf – og omvendt, jo større fjernhet, desto mindre tilbakeholdende med straf og andre lignende negative reaksjoner.” (Balvig 2010, s. 140).

Designet i den nordiske undersøkelsen forutsetter at den skriftlige versjonen og filmversjonen av hver sak er nøytrale og objektive ikke i seg selv, men *i forhold til hverandre*. Vi kan imidlertid etter min oppfatning, ikke blindt legge til grunn at den *regisserte* filmversjonen av den komprimerte handlingsbeskrivelsen i spørreskjemaet, på alle relevante punkter er lik de fullstendig ukjente, *uregisserte* og individuelle forestillingene som respondentene dannet seg ved lesingen av den komprimerte handlingsbeskrivelsen i spørreskjemaet. Den komprimerte skriftlige versjonen og filmsversjonen av samme sak kan være *skjeve* (biased) *i forhold til hverandre*, og da vil forskjeller i ”dommer” som blir avsaatt på grunnlag av papirversjonen og filmversjonen kunne skyldes (bare) denne skjevheten og ikke den ”nærhet” som er forsøkt skapt med filmversjonen. Dette er et metodeproblem. BGT anfører at dommer og advokater som medvirket i filmene, mener at de gir et gjenkjennelig bilde av hverdagen i retten. Men metodeproblemet er ikke om disse dommere og advokater kjenner seg igjen når de ser filmene. Spissformulert er problemet om papirsak X som fokusgruppedeltakerne først tok stilling til, er identisk (i alle relevante henseender) med filmatisert sak X, bortsett fra den ”nærhet” som filmen (eventuelt) skaper. Det er heller ikke til mye hjelp om vi vet at fokusgruppedeltakere synes at tiltalte i filmene virker mer eller mindre sympatisk eller troverdig, så lenge vi ikke har tilsvarende kunnskaper om hvordan de forestilte seg ham da de hadde lest papirfremstillingen av saken.

Fordi papirversjonen av sakene er svært lite informative om personene i saken, om deres opplevelse av handlingene, og om tiltaltes holdning til det han har gjort, er det mulig for folk med filmfaglig utdanning/erfaring å få øye på hva slags ”tilleggsinformasjon” (som ikke finnes i papirversjonen) regissør og skuespillere har lagt inn i filmene, særlig i form av mimikk, kroppsspråk og ved bruk av film-

tekniske virkemidler. Jeg kan bare notere at BGT ikke har tiltro til at fagfolk i film ikke kan avdekke dette.

Det finnes også eksempler på *verbal* tilleggsinformasjon i filmene, og to kan nevnes. I filmene sier tiltalte (via sin advokat) tidlig i rettsmøtet at han er skyldig i det han er tiltalt for. Papirversjonene forteller derimot ikke at tiltalte sa seg skyldig i retten. Kan ikke denne forskjellen bidra til å gjøre fokusgruppedeltakerne litt ”mildere stemt” etter filmversjonen enn de var etter lesing av papirversjonen? Det står ingen ting i papirversjonen om hva mannen som begikk voldtekten tenkte om sin handling senere. I filmversjonen sier han at han innser at han handlet galt overfor kvinnen. Er det sannsynlig at et slikt utsagn er helt betydningsløst for hvilke forslag til straff som blir foreslått? Eksemplene illustrerer at vi etter min oppfatning ikke stilltiende kan legge til grunn at papirversjon og filmversjon er identiske når det gjelder momenter som kan ha betydning for hvilke straffereaksjoner våre respondenter har foreslått, og vi kan dessverre ikke utelukke at Balvigs effektfunnskyldes at filmversjon og papirversjon er systematisk *biased* i forhold til hverandre.

Sammenligning av foreslått straff før og etter film (effektberegning) medfører at man må ta stilling til hvordan man skal håndtere svar med to avkryssinger. Skal man utelate den ene avkryssingen, eller skal alle tas med? Balvig (2010) forteller ikke hva han gjorde, men i en E-post (januar 2011) om et annet spørsmål opplyste han at han ”vedrørende fokusgrupperne alene har fået kodet den strengeste straf ind.” Men å stryke en av to avkryssede reaksjoner kan imidlertid lett produsere et resultat som det Balvig observerte hvis det var flere avkryssinger for to reaksjoner etter filmversjonen enn etter papirversjonen, slik tilfellet er i det norske materialet. Metodevalg med en slik mulig virkning blir normalt ikke ansett for å være tilrådelig.

Faktisk rettspraksis og fokusgruppens endelige dom

Kanskje BGTs replikk kan oppfattes slik at testing av nærhetshypotesens holdbarhet ikke er så viktig lenger, siden de foreslår å sammenligne faktisk rettspraksis med den siste (tredje) dom i fokusgruppene? Det er meget mulig at dette er et godt forslag, men også dette forutsetter åpenhet om metodeproblemer. For det første blir vi ikke kvitt metodeproblemer knyttet til avkryssing for to reaksjoner. For det andre vil fokusgruppedeltakerne også i sin siste dom være påvirket av den filmversjonen de har sett, og hvis hele budskap vi ikke kjenner godt nok. For det tredje vil fokusgruppedeltakerne være påvirket av *tilleggsinformasjon om reaksjonene* som ble gitt i fokusgruppene. BGT er i sin replikk nokså ensidig opptatt av betydningen av tilleggsinformasjon om fengselsstraff. Men hva med tilleggsinformasjon de ga om behandling og økonomisk kompensasjon og andre reaksjoner? Jeg fant det så

vanskelig å lage tilleggsinformasjon som var *balansert*, at jeg avsto fra å gi de norske fokusgruppedeltakerne tilleggsinformasjon. Mitt avvik fra de øvrige land på dette punkt åpner imidlertid en mulighet for å undersøke om denne forskjellen mellom landene – å gi tilleggsinformasjon av en bestemt type, eller ikke gi det – har ført til ulike resultater. For det fjerde: Når fokusgruppedeltakernes endelige dom skal sammenlignes med rettspraksis, bør vi ikke ensidig bruke dommerpanelenes ”dom” som sammenligningsgrunnlag (slik vi hittil har gjort). Rettspraksis i alle nordiske land består også i at *prøveløslatelse* regelmessig blir gitt, selv om det ikke er noen lovbestemt rettighet. For så lange dommer som i våre seks saker, er det svært vanlig i Norge å gi prøveløslatelse når 2/3 av dommen er sonet. Dermed kan det med rette hevdes at *faktisk straffenivå er 2/3 av dommerpanelenes dom*, og at fokusgruppedeltakernes siste dom helst bør sammenlignes med dette.

Hvorfor åpenhet om metodeproblemer?

Tilsynelatende uforstående til mine motiver, skriver mine kolleger: ”Af en eller anden grund fokuserer LPO i sin artikkel udelukkende på metodekritiske problemer, som efter hans mening kan *svække* hovedkonklusjonen.”

I januar i år forslø jeg for mine kolleger at vi sammen skulle skrive en artikkel om metodevalg og – problemer i de nordiske undersøkelsene, men de ønsket ikke å ta del i det. Derfor skrev jeg artikkelen alene, og mine grunner for å skrive var enkle. Som forskere skal vi selvsagt dokumentere våre forsøk på å vurdere konklusjoners holdbarhet, herunder om metodisk fremgangsmåte og analyseopplegg kan ha hatt (systematiske) virkninger slik at konklusjoner blir skjeve (biased) eller framstår som sikrere enn de er. Selv om vi også må overlate slike vurderinger til fremtidige eller andre forskere som kan komme til å avdekke at våre analyser og konklusjoner er uholdbare, gir ikke denne fremtidsmuligheten legitimitet til at vi kan unnlate å være kritiske til eget arbeid *nå*.

Forskningens innflytelse er avhengig av troverdighet, som igjen er betinget av forskernes evne og vilje til kritikk og åpenhet om mulige svakheter ved valg av teori, metode og analyser, både i egne og andres arbeider. Hvis jeg kan bidra til at det blir klarere enn i våre publiseringer hittil hvilke konklusjoner vi med rimelig grad av sikkerhet kan trekke, og hvilke usikkerhetsmomenter som er til stede, vil mye være vunnet fordi viktige deler av grunnlaget for konklusjoner blir kjent for alle.

Referanser

Balvig, Flemming (2006): *Danskernes syn på straf*. Advokatsamfundet. København.

- Balvig, Flemming (2010): *Danskernes retsfølelse og retsfornuft – et forspil*. Det Juridiske fakultet, Københavns Universitet, København.
- Jerre, Kristina og Henrik Tham (2010): *Svenskarnas syn på straff*. Rapport 2010:1. Stockholms Universitet, Kriminologiska Institutionen. Stockholm.
- NOU 2011: 13 *Juryutvalget. Når sant skal skrives*. Utredning avgitt til Justis- og politidepartementet 17. juni 2011.