

Replikasjonskrisen



Når vi forsøker å gjøre tidligere studier om igjen, får vi ikke alltid de samme resultatene. Replikasjonskrisen har kastet skygger over psykologifaget. Hva kan gjøres?

TEKST

Bjørn Sætrevik

PUBLISERT 4. juli 2017

I henhold til etiske prinsipper for nordiske psykologer (Norsk psykologforening, 1998) skal vår praksis bygge på vitenskapelige prinsipper og vitenskapelig kunnskap. Dette har vi gjerne tolket som at vi skal basere arbeidet vårt på effekter som er vist i vitenskapelige studier publisert i fagfellelvurderte tidsskrift. Vi har stolt på den vitenskapelige metoden, forskernes praksis, fagdisiplinens tradisjoner for publisering og formidling, og at det er mekanismer for selv-korreksjon som gjør at psykologifaget stadig kommer nærmere en riktig beskrivelse av hvordan sinnet fungerer.

Det har imidlertid de siste årene blitt tydelig for oss at det kan være brister i fundamentet som psykologien bygger på. Når forskere forsøker å gjenta tidligere forskningsstudier, altså «å replikere» studiene, klarer de ofte ikke å finne de samme resultatene. Den økende oppmerksomheten rundt dette kalles ofte for «replikasjonskrisen», og kan få konsekvenser for psykologifagets forskning og praksis.

Skuffende replikasjoner

Noen av studiene som har vist seg å ikke få samme resultat når forskere forsøker å gjøre dem på nytt, er studier som har etablert grunnleggende antagelser om hvordan vi fungerer. Dette gjelder studier som har vist at detaljer i omgivelsene som vi ikke legger

merke til, kan ha en stor innvirkning på hvordan vi oppfører oss («goal priming»-forskning), at motivasjon er en begrenset ressurs («ego-depletion» effekten), at våre ansiktsuttrykk påvirker humøret («facial-feedback» hypotesen), at menn blir mer konkurranserettede når det er kvinner i nærheten («romantic priming»), eller at selvbildet vårt blir viktigere for oss av å tenke på døden («terror management theory»). Andre studier som har vært vanskelige å replikere, har gått på mer perifere effekter, som at bestemte kroppsholdninger gjør oss mer selvsikre («power posing»), at vi får bedre forståelse for andre av å lese skjønnlitteratur, eller føler et behov for å vaske oss etter å ha gjort noe umoralsk («Macbeth-effekten»). Noen av disse studiene må nå sies å være «falsifisert» av nyere forskning, mens andre har blitt trukket i tvil, og vi venter på mer endelige studier for å avgjøre kontroversen.

Det har vært mest oppmerksomhet om slike problemer innenfor grunnforskning i sosialpsykologi (Kahneman, 2012), men det får nå også mer oppmerksomhet innenfor andre eksperimentbaserte retninger, som kognitiv psykologi og nevrovitenskapelig forskning. For eksempel var det et lovende forskningsfelt som så på hvordan hormonet oxytocin spilte en rolle i menneskelig tilknytning, mens vi i dag er mer usikre på dette (Lane et al., 2015). Videre ble det nylig (Poldrack et al., 2017; Szucs & Ioannidis, 2017; Yarkoni, 2009) pekt på at fMRI-studier oftest har for lav statistisk kraft til å trekke konklusjoner, at effektene ofte er små, at det er en rekke urapporterte valg som tas i analysen, og at noen av de statistiske metodene har svakheter (Eklund, Nichols, & Knutsson, 2016).

Så langt har høydepunktet for replikasjonskrisen kommet fra forskningen knyttet til «Reproducibility initiative project» (Open Science Collaboration, 2015). Der samarbeidet 270 forskere fra flere land om å replikere 100 studier som ble publisert i tre av psykologiens ledende tidsskrift mellom 2010 og 2012. Selv med hjelp fra de opprinnelige forskerne klarte de bare å få samme resultat i mindre enn halvparten av replikasjonene. Resultatene ser ut til å ha overrasket og skuffet en del forskere og fagfolk som antok at flertallet av publiserte studier ville la seg replikere.

Inntrykket vi sitter igjen med etter replikasjonskrisen, er en mistanke om at mye av den publiserte forskningen ikke er reliabel, noe som tvinger oss til å være kritiske til en del av teoriene og praksisene som er utviklet på feltet. Vi kan ikke regne effekter som etablerte selv om de ser ut til å teste troverdige hypoteser, resultatene er kompatible med tidligere forskning, metoden virker solid, forskerne er kompetente, og artikkelen er fagfellevurdert i velrennomerte tidsskrift. Det er ikke noe spesielt som kjennetegner de studiene som ikke har latt seg replikere – faktisk ser det ut til at replikasjonsraten kan være lavere for hyppig siterte studier publisert i topp-tidsskrift (Ioannidis, 2005).

Hovedproblemet er kanskje ikke at noen studier har vist seg å være vanskelige å replikere, men at vi svært sjelden forsøker å replikere studier. Det meste av forskningen tar for gitt at det som er publisert er riktig, og forsøker å bygge videre på det, heller enn å se behovet for å replikere tidligere studier.

«Selv med hjelp fra de opprinnelige forskerne klarte de bare å få samme resultat i mindre enn halvparten av replikasjonene»

For mange av oss har replikasjonskrisen ført til en skepsis både til tidligere funn og ny forskning, og en endring i hva vi tenker på som robust psykologisk forskning.

Årsaker til replikasjonskrisen

Årsakene til replikasjonskrisen kan føres tilbake til svakheter ved måten psykologisk forskning vanligvis gjøres på, og til mekanismer som påvirker hvordan forskerne jobber, og hva som publiseres.

Forskningspraksiser som årsak

Det meste av publisert psykologisk forskning baserer seg på «null-hypotese-signifikanstesting.» For å bli publisert med en slik tilnærming er det avgjørende at du kan vise en statistisk test med en signifikansverdi lavere enn et gitt kriterium (ofte satt til $p = 0,05$). Det fører til at forskeren er sterkt motivert til å finne signifikante resultater i analysene sine. Selv om det har vært flere høyt profilerte tilfeller av juks i psykologisk forskning, regner vi med at dette bare i mindre grad bidrar til replikasjonskrisen. Derimot har det blitt vist (John, Loewenstein, & Prelec, 2012) at det er relativt utbredt med forskningspraksiser som ligger i gråsonen for det akseptable. Det å forsøke en rekke statistiske tilnærminger for å finne et signifikant resultat kalles noen ganger «p-hacking». Dette kan være ulike metodiske grep som ligger i spennet mellom «allmenn» og «uetisk» forskningspraksis, som å forsøke ut flere utfallsmål og interaksjoner, fjerne avvikende datapunkt, eller å samle mer data, til en finner en effekt som er statistisk signifikant. En annen tilnærming er at hypotesene kan være temmelig vage og uspesifiserte i begynnelsen av studien, men at de justeres, tilpasses og konkretiseres etter at en har sett på dataene (såkalt HARKing – «hypothesizing after the results are known»). I den publiserte forskningsartikkelen ser det da ut som at en har klart å predikere nøyaktig den sammenhengen som viste seg i data, heller enn at dette bare er en av mange mulige sammenhenger som ble testet. Den tilnærmingen som faktisk er brukt, ville utgjort en mindre overbevisende historie, og ville krevd en annen statistisk analyse.

Publikasjonspraksis som årsak

I tillegg til forskerens sterke motivasjon til å finne signifikante funn er det store mørketall for hvor mye av forskningen som *ikke* blir publisert. Forskere starter en rekke studier (eller analyser av innsamlede data), og dersom de ikke finner det forventede resultatet (eller finner uklare resultater), blir de ikke publisert. Det kan være fordi forskeren kan mistenke at noe har gått galt et sted i forskningsprosessen, eller tenker på

studien som «bare en pilotstudie», eller ønsker å gjøre en ny og forbedret studie. Det kan være at forskere er motvillige til å publisere funn som ikke støtter den teorien de tidligere har jobbet innenfor, eller ikke støtter den terapiformen eller intervensjonen de bruker. Det kan også være at forskeren forsøker å publisere, men at tidsskriftene er motvillige til å publisere artikler som viser uklare funn eller nullfunn. At mange studier aldri blir publisert, kalles ofte «skrivebordsskuff-effekten» eller «publikasjonsbias». Dette fører til at forskningslitteraturen bare består av en liten del av de studiene som faktisk er gjennomført, og at studier med uklare funn eller funn som strider mot det etablerte synet, er underrepresentert.

Strukturelle årsaker

Forskere er i kontinuerlig konkurranse for å få stillinger, forskningsmidler og status, og i denne konkurransen måles de på antall publikasjoner, hvor oppsiktsvekkende og betydningsfulle funnene er, hvem som er medforfattere, og status til tidsskriftet der forskningen er publisert. I de senere år har dette tilspisset seg på grunn av en økning i antall stipendiater (som alle forventes å publisere tidlig i karrieren) og internasjonal konkurranse om stillinger. I tillegg har den totale mengden artikler som publiseres, økt voldsomt de siste tiårene (Van Noorden, 2014), samtidig som hver publikasjon får mindre oppmerksomhet (Larsen & von Ins, 2010). Akademiske institusjoner har ofte insentivmidler som belønner forskere som publiserer mye, mens det er få insentiver som oppfordrer forskere til å være nøkterne, møysommelige og gjennomsluktig i forskningen (Nosek, Spies & Motyl, 2012).

Teknologisk utvikling har gjort det mye lettere å hurtig samle inn og analysere store datasett. Mens en tidligere følte seg forpliktet til å få publisert de dataene en hadde samlet inn, selv om resultatene var uklare eller svake, kan det i dag synes mer fristende å forkaste data og gjøre en ny datainnsamling. I tillegg måtte en tidligere investere flere timer på å regne ut analysen, mens dagens statistikkpakker lar oss gjøre raske «pek og klikk»-analyser for å teste ut forskjellige varianter av hypotesen.

Hva kan gjøres?

Det har blitt foreslått en rekke tiltak og retningslinjer for endret praksis som skal gjøre oss sikrere på at vår forskning vil kunne replikeres av andre, og dermed hjelpe psykologien til å reetablere seg som en robust vitenskap. Her vil jeg ta for meg noen av de viktigste tiltakene som diskuteres:

Pre-registrerte studier

Med pre-registrering menes at vi på forhånd skriver ned hvilke hypoteser vi vil teste, hvordan vi skal samle inn data, og hvordan analysene skal gjøres. Denne planen offentliggjøres på nettsider som osf.io/ eller AsPredicted.org/ med et tidsstempel. Dette gjør det enklere for oss å holde orden på at alt er planlagt på forhånd, og at vi faktisk følger den planen. Forskningen blir mer gjennomsluktig og pålitelig, og leseren kan ha tiltro til at det har vært en redelig forskningsprosess. Leseren kan sjekke at vi faktisk har

fulgt planen, og ikke har drevet med p-hacking eller HARKing. Pre-registrering lar oss også bruke mer sensitive statistiske tester enn vi ellers kunne gjort, og kan gjøre det lettere å få publisert studien også når resultatene er uklare eller viser null-funn.

Åpen forskning

Et annet tiltak mot replikasjonskrisen er en bevegelse mot mer åpen forskning. Det går blant annet ut på at vi gjør datasettene offentlig tilgjengelig, slik at andre forskere kan etterprøve den analysen vi har gjort, og se om de vil finne andre resultater dersom de bruker andre analyser. Dette kan også bidra til en effektivisering, ved at en større andel av datasettet blir utnyttet, og hever kvaliteten på metaanalysene. Vi kan også gjøre materialene som er utviklet for studien tilgjengelig, slik at de kan etterprøves eller brukes om igjen av andre. I noen tidsskrift blir det i dag forventet at data er åpne, og forskerne må ellers rettferdiggjøre hvorfor dette ikke er gjort.

Nye metoder

Som nevnt bruker det meste av psykologisk forskning signifikanstesting, og overdreven tillit til denne tilnærmingen har fått deler av skylden for replikasjonskrisen. En mulig løsning kan derfor være å rapportere effektstørrelser eller bruke alternative statistiske tilnærminger. Tidsskriftene forventer i dag at artikkelen har en lengre metodedel som inneholder alle nødvendige detaljer som skal til for å replikere studien, og at vi i forkant har regnet ut at studien vil ha tilstrekkelig statistisk kraft til å avvise null-hypotesen. For å unngå at analysen påvirkes av hva vi ønsker å finne, er det mulig å gjøre en «blindet analyse», der de statistiske operasjonene gjennomføres uten at vi vet hvilke data som hører til hvilken eksperimentell betingelse.

Nye publiseringsmåter

Det er også en viss utvikling i måten forskning publiseres på. Det er utviklet retningslinjer for å fremme gjennomsiktighet og åpenhet (Nosek et al., 20. april, 2017), som stadig flere tidsskrift slutter seg til. Om du utfører gjennomsiktig og åpen forskning, kan du vise det ved å knytte egne «merker» til artikkelen (Kidwell et al., 2016). Vi ser også at tidsskrift i større grad åpner for å publisere mer nøkterne og realistiske versjoner av studien, selv om den kan ha mindre imponerende funn (Inzlicht, 2015). Noen tidsskrift åpner også for «registrerte rapporter», noe som innebærer at de først vurderer en pre-registrering av forskningen, og forplikter seg til å publisere den ferdige artikkelen så lenge planen følges, uavhengig av hva resultatene viser. Mens det for få år siden var svært vanskelig å få publisert replikasjonsstudier eller funn som viser nullfunn, er det nå flere tidsskrift som åpner for dette. Noen nye tidsskrift er dedikert til dette, og det er også i større grad mulig å selv offentliggjøre sin forskning på en enkel måte i et repositorium (som www.psychfiledrawer.org eller osf.io/preprints/psyarxiv/). En siste endring er at mye av den vitenskapelige debatten er flyttet ut av tidsskriftene og over i sosiale medier og blogger. På bloggen www.pubpeer.com kan anonyme forskere kritisk diskutere publiserte artikler, og berettiget kritikk har ført til at artikler har blitt trukket tilbake.

Forskningskulturen

Til slutt er det også økt oppmerksomhet om svakhetene i forskningspraksisen vår, og på behovet for å endre kulturen blant forskere. Det *har* blitt mer interesse og aksept for å gjøre replikasjoner. Der forskerne tidligere heget om egne studier og data, er det i dag et større innslag av delingskultur. Eksempler på dette er at flere laboratorier kan samarbeide om å analysere de samme dataene uavhengig av hverandre, eller at de gjør den samme datainnsamlingen parallelt, og så ser om de finner den samme effekten. Mange forskere er i dag klar over de kognitive tankefellene vi er utsatt for i forskningsprosessen (Nuzzo, 2015), som subtilt kan gjøre oss overbevist om vår egen forsknings fortreffelighet.

Status nå

Dagens lesere av forskningsartikler vil i større grad enn før være skeptiske til spektakulære funn i enkeltstudier, spesielt når effektene er like innenfor signifikansgrensen, med få deltagere eller flere interaksjoner i analysen. Men også verdien av metaanalyser trekkes i dag i tvil, siden vi vet at det er systematiske effekter som både vrir på resultatene av enkeltstudiene som metaanalysen består av, og som avgjør hvilke enkeltstudier som blir publisert og dermed kan være med i metaanalysen. En del forskere vil i dag være kritiske til resultatene fra alle studier som ikke er pre-registrert, da vi ikke kan vite hvor mye forskningsprosessen har blitt justert underveis. Som et eksempel ble det tidligere i år publiserte en replikasjon av Milgrams lydighetsstudie (Doliński et al., 2017), som støttet opp om de opprinnelige funnene. Selv om mange ønsker slike replikasjoner velkommen, ble det uttrykt skuffelse over at den ikke var gjort «skikkelig» når det først skulle gjøres, med pre-registrering og tilstrekkelig antall deltagere.

De mottiltakene vi har nevnt over, får stadig mer oppmerksomhet og medvind, og blir i større grad integrert i psykologifagets praksis. Tidligere i år ble det publisert et manifest for reproducerbar psykologisk forskning (Munafò et al., 2017), og et spesifikt for nevrovitenskap (Poldrack et al., 2017). Årsakene til og løsninger på replikasjonskrisen blir i større grad inkludert i psykologiutdanningene, blant annet ved masterprogrammet i psykologi ved Universitetet i Bergen, der vi også trener på pre-registrering i så stor grad som mulig.

Samtidig bør vi være klar over at mange av våre kolleger nok vil være ganske ukjente med omfanget av replikasjonskrisen og hvilke konsekvenser det har for hvordan vi bør forske på og utøve faget. I tillegg er noen kolleger kjent med problematikken, men uenig i hvor alvorlig den er, eller i hvilken grad det er behov for eller ønskelig med en endring i måten vi forsker og publiserer på (se f.eks. Maxwell, Lau, & Howard, 2015; Stroebe & Strack, 2014). Sophie Scott (2013) frykter at nye retningslinjer kan hindre den typen utforskning av data som hun mener er viktig for psykologisk forskning. Det er også de som er skeptiske til at ens forskning blir gått etter i sømmene av andre, og det har blant annet blitt sådd tvil om motivasjonen og kvalifikasjonen til de som gjør replikasjoner. Redaktørene for *New England Journal of Medicine* (Longo & Drazen, 2016) mente at

forskere ikke er berettiget til å publisere basert på et forskningsdesign som andre har utviklet, eller på et datasett som andre har samlet inn. Det har utviklet seg en metadebatt (Singal, 2016), der forskere som Susan Fiske (2016) mener at det har vært en unødig krass og ufølsom tone i kritikken av ikke-replikerte studier og mistanke om uredelige forskningspraksiser. Andrew Gelman (21. september 2016) er uenig i dette, og peker på at det alltid vil være ubehagelig å bli kritisert, og at tilhengere av replikasjon har blitt avskrevet som «data-parasitter» og «metodologiske terrorister».

Vi skal være klar over at den type diskusjon som replikasjonskrisen bringer med seg, ofte følger de etablerte maktstrukturene i psykologisk forskning: Hvis man har bygget opp en forsknings- eller behandlingskarriere basert på studier utført i «den gamle modellen», kan det være at en har lite interesse av at den blir kritisk gjennomgått. Etablerte forskere, forlag, tidsskrift og vitenskapelige foreninger har gjerne også kommersielle interesser i å bevare status quo. Dette har ført til en generasjonseffekt, der unge forskere er mest opptatt av replikasjonskrisen, mens eldre forskere i større grad stritter imot.

Hva replikasjonskrisen ikke er

Å følge debatten om replikasjonskrisen i psykologien fra sidelinjen kan gi oss inntrykk av at vi ikke kan sette lit til noen av funnene våre, og at psykologifaget er i dyp krise. Riktignok er det forskere som Inzlicht (sitert i Engber, 2016) som mener at krisen er så alvorlig at istedenfor å gå gjennom hele litteraturen for å finne ut hva som lar seg replikere, burde vi heller kaste alt og erklære et «år én» for faget. Men dette synet bør nok nyanseres noe ved å se på hvilke slutninger vi *ikke* kan trekke fra replikasjonskrisen.

Ikke en avvisning av empirismen

Den omfattende debatten rundt replikasjonskrisen går riktignok mellom parter som alle mener at faget skal være empirisk basert. Det er altså enighet om at vi skal postulere hypoteser og teste dem opp mot data, selv om det har vist seg å være vanskeligere enn vi tidligere regnet med. Replikasjonskrisen gir derfor i liten grad støtte til påstander om at empirisk forskning ikke er mulig for psykologi (sist hevdet av Jan Smedslund i aprilutgaven av Psykologtidsskriftet). Det er riktigere å si at de selv-korrigerende mekanismene vi tidligere har antatt var til stede i psykologisk forskning, ikke har fungert optimalt.

På et mer detaljert nivå har de fleste sviktende replikasjoner blitt vist i laboratorieeksperimenter. Allikevel har ikke dette ført til at denne metodologiske tilnærmingen er svekket til fordel for andre metoder. Randomiserte, kontrollerte eksperimenter går fortsatt for å være gullstandarden for psykologisk grunnforskning og tilstøtende fagfelt, selv om det har blitt økt oppmerksomhet på hvordan eksperimentene gjøres, analyseres og rapporteres. Den reelle uenigheten går snarere på hvorvidt det skal forventes at vi alltid på forhånd annonserer alle studier og forventningene til funn, og alltid gir tilgang til dataene slik at andre skal kunne sjekke

analysen. Det er også variasjon i holdningene til å drive med replikasjoner, og hvor mye av forskernes tid som bør brukes på replikasjon, opp mot det som brukes på original forskning. Forskere kan ha ulike meninger om hvor mye tillit vi kan ha til én enkelt studie, eller om vi for eksempel bør forvente at to selvstendige laboratorier finner det samme funnet før et tidsskrift publiserer det.

«Unge forskere er mest opptatt av replikasjonskrisen, mens eldre forskere i større grad stritter imot»

Ikke bare for grunnforskning

I psykologien har det så langt vært mest snakk om sviktende replikasjoner for relativt «enkle» eksperimenter i lab-lignende situasjoner – der man tester effekten av en variabel på en annen. Hva med de mer komplekse mekanismene og fenomenene som vi ofte postulerer i psykologien, og som danner grunnlaget for hvordan de fleste psykologer jobber til daglig? Selv om replikasjonskrisen så langt i liten grad har handlet om slike komplekse sammenhenger, er det liten grunn til å forvente at replikasjonsraten skal være bedre for slik forskning. I spørreskjemastudier samles det vanligvis inn en rekke ulike variabler, uten at det på forhånd erklæres hvilke variabler vi forventer at skal virke på hverandre, med hvilke interaksjoner, og hvordan de skal operasjonaliseres. Mye forskning på terapi eller andre intervensjoner bruker i større grad kvalitative metoder, der tolkning av mellommenneskelig interaksjon står sentralt. På samme måte som kvantitativ metode har vist seg å bli påvirket av forskernes motivasjon og forventning, bør vi forvente lignende effekter i kvalitativ forskning, samtidig som effektene da vil være vanskeligere å legge merke til og etterprøve. Kvantitative kliniske studier vil også være utsatt for praksiser som ligner p-hacking og HARKing. Ofte benyttes en rekke ulike variabler til å måle resultatet av terapien, slik som et utvalg skalaer for symptomendring, psykologens eller pasientens vurdering av bedring, endring i livssituasjon, tilbakefall, medikamentbruk og behandlingens allianse eller varighet. Men i forskningsartikkelen utelates ofte de variablene som ikke viste et positivt resultat (såkalt outcome-switching, Chan, Hróbjartsson, Haahr, Gøtzsche, & Altman, 2004; Kirkham et al., 2010).

Heller enn at forskning med spørreskjema, intervensjoner, kvalitative og kvantitative kliniske metoder skal være mer pålitelige enn eksperimentelle studier, kan vi forvente at disse vil være enda vanskeligere å replikere. Samtidig er det vanskelig å se for seg hvordan mottiltak som pre-registrering skal kunne være like effektive her.

Ikke bare for psykologien

Det snakkes ofte om replikasjonskrisen *i psykologi*, og en kan få inntrykk av at dette er et problem som gjelder i større grad her enn for andre fagdisipliner, og at vi i dag står svakere sammenlignet med andre fag. Men vi kan innvende at ingen andre disipliner

har hatt like omfattende og organiserte forsøk på replikasjon som psykologien. I den grad det er forsøkt i andre disipliner, så er funnene ikke mer oppløftende der: En har bare klart å replikere funnene i 25 % av biomedisinsk forskning (Begley & Ioannidis, 2015), 11 % i kreftforskning (Begley & Ellis, 2012), 21 % i farmakologi (Prinz, Schlange, & Asadullah, 2011), og 33-49 % i økonomi (Chang & Li, 2015).

Heller enn at det står dårligere til med psykologifaget enn sammenlignbare fagfelt, kan det være at vi er tidligere ute enn andre med oppmerksomhet rundt den type problemer som i stor grad gjelder for all tilsvarende forskning. Det kan skyldes at vi i psykologien forsker på faktorer som påvirker våre handlinger, og at det derfor faller naturlig for oss å se kritisk på vår egen forskningsaktivitet. Gelman (22. september, 2016) mener at problemene i psykologien har fått uforholdsmessig mye oppmerksomhet blant annet fordi faget er av allmenn interesse. Det er også et sunnhetstegn at vi har høye forventninger til forskningen vår, og at det ser ut til å bekymre oss når replikasjonsraten er lav.

«Heller enn at det står dårligere til med psykologifaget enn sammenlignbare fagfelt, kan det være at vi er tidligere ute enn andre med oppmerksomhet rundt den type problemer som i stor grad gjelder for all tilsvarende forskning»

Forskningens formål

Replikasjonskrisen har kastet skygger over psykologifaget og pekt på noen svake sider ved «den gamle modellen» for forskning. Dette er noe vi bør ta alvorlig.

Hovedproblemet er ikke hvor dårlig det har gått de gangene vi har forsøkt å replikere funnene våre, men hvor sjelden vi i det hele tatt prøver å gjøre replikasjoner. For at forskningen skal være selv-korrigerende, må vi slutte å fokusere kun på å gjøre spennende nye enkeltstudier; vi må begynne å se på uavhengig replikasjon som en viktig del av forskningen.

Vi står på vakkende grunn i forsøket på å drive med vitenskapelig psykologi dersom vi tenker på alle publiserte studier og etablert praksis som uttrykk for «sikker kunnskap.» Siden effektene er systematiske, løser det ikke problemet å si at vi kun vil stole på funn som er entydige over flere studier, eller som har blitt støttet i metaanalyser. Vi bør heller ha en nøktern holdning til alle funn frem til de er replikert i uavhengige, pre-registrerte studier.

Replikasjonskrisen tvinger oss til å tenke grundig gjennom hvorfor vi i det hele tatt driver med forskning. Vi må se på om insentivene for forskerens karrierebygging står i veien for samfunnets nytte av forskningen. Vi må se på om det skal være mulig å

etablere seg som forsker uten å drive med forskning der målet alltid skal være å få nye og spennende funn. Robust empirisk forskning er mulig for psykologien, men replikasjonskrisen har vist at det krever at de selv-korrigerende mekanismene som vi tidligere har antatt var til stede, gjøres synlige og etterprøvbare.

Referanser

- Begley, C.G., & Ellis, L.M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531–533.
- Begley, C.G., & Ioannidis, J.P.A. (2015). Reproducibility in Science. *Improving the Standard for Basic and Preclinical Research*, 116(1), 116–126.
doi:10.1161/circresaha.114.303819
- Chan, A.-W., Hróbjartsson, A., Haahr, M.T., Gøtzsche, P.C., & Altman, D.G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *Jama*, 291(20), 2457–2465.
- Chang, A.C., & Li, P. (2015). Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not'. *FEDS Working Paper*, 2015-83.
- Doliński, D., Grzyb, T., Folwarczny, M., Grzybała, P., Krzyszycha, K., Martynowska, K., & Trojanowski, J. (2017). Would You Deliver an Electric Shock in 2015? Obedience in the Experimental Paradigm Developed by Stanley Milgram in the 50 Years Following the Original Studies. *Social Psychological and Personality Science*.
doi:doi:10.1177/1948550617693060
- Eklund, A., Nichols, T.E., & Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 201602413.
- Engber, D. (2016). Everything is crumbling. An influential psychological theory, borne out in hundreds of experiments, may have just been debunked. How can so many scientists have been so wrong. *Slate*. 6. mars 2016.
- Fiske, S.T. (2016). A call to change science's culture of shaming. *APS Observer*.
- Gelman, A. (21. september 2016, 21. September 2016). What has happened down here is the winds have changed. . *Statistical Modeling, Causal Inference, and Social Science blog*. Retrieved from <http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-wi...>
- Gelman, A. (22. september, 2016, 22. september 2016). Why is the scientific replication crisis centered on psychology? *Statistical Modeling, Causal Inference, and Social Science*. Retrieved from <http://andrewgelman.com/2016/09/22/why-is-the-scientific-replication-cri...>
- Inzlicht, M. (2015). Guest Post: A Tale of Two Papers. *Sometimes I'm wrong*. Retrieved from <http://sometimesimwrong.typepad.com/wrong/2015/11/guest-post-a-tale-of-t...>

Ioannidis, J.P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Jama*, 294(2), 218–228.

John, L.K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 0956797611430953.

Kahneman, D. (2012). A proposal to deal with questions about priming effects. *Nature*, 490.

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., ... Hess-Holden, C. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biol*, 14(5), e1002456.

Kirkham, J.J., Dwan, K.M., Altman, D.G., Gamble, C., Dodd, S., Smyth, R., & Williamson, P.R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *British Medical Journal*, 340, c365.

Lane, A., Mikolajczak, M., Treinen, E., Samson, D., Corneille, O., de Timary, P., & Luminet, O. (2015). Failed replication of oxytocin effects on trust: the envelope task case. *PloS one*, 10(9), e0137000.

Larsen, P.O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575–603.

Longo, D.L., & Drazen, J. M. (2016). Data Sharing. *New England Journal of Medicine*, 374(3), 276–277. doi:10.1056/NEJMe1516564

Maxwell, S.E., Lau, M.Y., & Howard, G.S. (2015). Is psychology suffering from a replication crisis? What does «failure to replicate» really mean? *American Psychologist*, 70(6), 487.

Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., du Sert, N.P., ... Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.

Norsk Psykiologforening. (1998). Etiske prinsipper for Nordiske psykologer. Hentet fra <https://www.psykiologforeningen.no/medlem/etikk/etiske-prinsipper-for-nor...>

Nosek, B. A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., ... Stodden, V. (20. april, 2017). Transparency and Openness Promotion (TOP) Guidelines. Hentet fra osf.io/9f6gx

Nuzzo, R. (2015). How scientists fool themselves-and how they can stop. *Nature*, 526(7572), 182.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., ... Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*, 10(9), 712–712.

Scott, S. (2013). Pre-registration would put science in chains. *Times Higher Education*: <http://www.timeshighereducation.co.uk/comment/opinion/science-in-chains/...>

Singal, J. (2016, 12. okt. 2016). Inside Psychology's 'Methodological Terrorism' Debate. *Series of us*. Retrieved from <http://nymag.com/scienceofus/2016/10/inside-psychologys-methodological-t...>

Smedslund, J. (2017). Det kommer an på. *Tidsskrift for Norsk psykologforening*, 54(4), 402–407.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71.

Szucs, D., & Ioannidis, J.P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, 15(3), e2000797.

Van Noorden, R. (2014). Global scientific output doubles every nine years. *Nature News Blog*.

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power – Commentary on Vul et al. (2009). *Perspectives on Psychological Science*, 4(3), 294–298.

Teksten sto på trykk første gang i Tidsskrift for Norsk psykologforening, Vol 55, nummer 7, 2017, side 640-647

TEKST

Bjørn Sætrevik, førsteamanuensis ved Institutt for samfunnspsykologi, Universitetet i Bergen