



Masteroppgave

Helsevitenskapelig fakultet

Atferdsvitenskap

Juni, 2022

Reduksjon i forekomst av bekreftelsesfellen

Effekt av to typer trening ved umiddelbare tilbakemeldinger på atferd

Kandidatnavn: 340535

Emnekode: MALK5000

30 studiepoeng

Takk

Jeg vil takke min samboer og utvidede familie for muligheten til å skrive master på tross av familiesituasjonen, for emosjonell støtte, akademisk pådriv og teknisk bistand. Jeg vil takke arbeidsplassen min for tilrettelegging, og kollegaer for støtte og bidrag. En takk også til min veileder Per Holth som har svar på e-post i helger, og arrangert forskningslabb under spesielle forhold.

Et forsøk på å redusere forekomst av bekreftelsestendensen ved to ulike og umiddelbare tilbakemeldinger på atferd. En studie av teknikk for å lære bedret beslutningstagning og unngå ensidig vurdering av fakta.

Elisabeth S. Johannessen

OsloMet- Storbyuniversitet

Fakultet for helsevitenskap, institutt for atferdsvitenskap

15. Juni 2022

Sammendrag

Bekreftelsesfellen innebærer en tilbøyelighet til å vektlegge informasjon skjevt i favør av egen overbevisning, eller ensidig søke verifisering av en hypotese uten å undersøke muligheten for falsifisering. Bekreftelsesfellen gjør seg gjeldene i trivielle hverdagsavgjøresler, men har også preget internasjonal politikk, påvirket historiske begivenheter, spiller inn på justis- og rettsystemets objektivitetsvurderinger, og medvirker til meningspolarisering via sosiale medier. Det fremstår derfor viktig og samfunnsnyttig å utrede teknikker som kan redusere forekomst av bekreftelsesfellen. I dette eksperimentet blir deltagerne utsatt for to former for tilbakemelding; generell tilbakemelding og utvidet/forklarende tilbakemelding. Formålet er å se på muligheten for å trene deltagerne til å falsifisere på lik linje med verifisere, og dermed lære ferdigheter for bedret beslutningstaking. Hovedfunnet fra datasett basert på 14 inkluderte deltagere tilsier at begge de uavhengige variablene har effekt på den avhengige variabelen, hvertfall på kort sikt. Hos noen deltagere hadde intervensjonen god effekt, og ferdighetene lot seg generalisere utover treningsbetingelsene, mens noen deltagere hadde ingen effekt av intervensjon. Det er undersøkt rekkefølgeeffekt mellom gruppene, og sett på noen mulige årsaker til hvorfor effekten varierte på individnivå. Det er benyttet et «*mixed design*».

Nøkkelord: Bekreftelsesfellen, redusere responsmønster, ulik tilbakemelding ved trening, beslutningstaking, justisfeil og atferdsvitenskap.

Abstract

Confirmation bias implies a tendency to emphasize information skewed in favor of one's own conviction, or unilaterally seek verification of a hypothesis without examining the possibility of falsification. Confirmation bias makes its appearance in everyday decisions, but has also influenced international politics, historical events and the judicial system's objectivity assessments, and contributes to opinion polarization via social media that takes us many decades back in time. It appears therefore important and socially beneficial to study techniques that can reduce the incidence of confirmation bias. In this experiment, participants are exposed to two forms of feedback; general feedback and extended / explanatory feedback. The purpose is to look at the possibility to train participants to falsify in the same way as verifying, and thus learn skills for improved decision making. The main finding from data sets based on 14 included participants indicates that both of the independent variables have an effect on the dependent variable, at least short term. In some participants, the intervention had a good effect, and the skills could be generalized beyond the training conditions, while some participants had no effect of intervention. The effect of order between the groups was examined, in addition to possible reasons why the effect varied on an individual level. It is used a "mixed design".

Key words: Confirmation bias, reduce response pattern, varied feedback by training, decision making, misconduct and behavioral science.

Innholdsfortegnelse

Takk	2
Sammendrag	4
Abstract.....	5
Innholdsfortegnelse.....	6
Oversikt over tabeller og figurer.....	7
Metode	16
Deltagere	17
Setting	18
Apparatur	19
Prosedyre.....	19
Resultat	23
T-tester for to uavhengige variabler.....	28
Diskusjon	30
Referanser	37
Appendiks	41

Etisk refleksjonsnotat.....	42
-----------------------------	----

Oversikt over tabeller og figurer

Tabell 1 A, B og C: Oversikt over kort med stimuli i alle 3 fasene.....	20
Figur 1: Flytdiagram over prosedyren	21
Tabell 2 A og B: Resultatoversikt gruppe 0 og gruppe 1	23
Figur 2: Individuell bedring fra baselinje 1 til baselinje 3.....	24
Figur 3 A og B: Antall forsøk deltakerne i gr. 0 / gr. 1 brukte på første sammenlignet med andre intervensjonsrunde	25
Figur 4: Oversikt over tiden deltakerne valgte å ha tilbakemelding tilgjengelig.....	27
Figur 5: Gjennomsnittlig endring fra intervensjon 1 til intervensjon 2	28
Tabell 3: Responsforekomst i gruppene ved første baselinjerunde før intervensjon.....	29
Tabell 4: "Selection task" presentert ved tre-term kontingens.....	36
Tabell 5: Rimelig tidsbruk på å lese instruksjoner før hver oppgave, eller tilbakemeldinger gitt etter hver hver oppgave.....	41

Bekreftelsesfellen innebærer en tilbøyelighet til å vektlegge informasjon skjevt i tråd med egen overbevisning, forventninger eller preferanse, samt ofte ignorere motstridende informasjon (Koriat et al., 1980). En faktasøker ensidig, og tolker systematisk bevis på en måte som støtter opp om allerede eksisterende overbevisning (Nickerson, 1998). Dette kan innebære at en låser seg til en bestemt teori der en med eller uten overlegg unnlater å sjekke alternative hypoteser (Bratholm & Eskeland, 2008). Nickerson (1998) presenterer muligheten for at en i slike situasjoner antar at alternative hypoteser må være feil, og at det ikke faller en inn å sjekke om dette stemmer. Informasjon som støtter egen overbevisning er mer forsterkende enn det å få informasjon som indikerer at en tar feil (Catania, 2013). I tillegg kan bekreftelsesfellen innebære et økt fokus på fenomener som er aktuelle for en selv og egen overbevisning, og dermed kan en få et skjevt bilde av dette fenomenets betydning eller omfang. Det skjeve fokuset kan øke tilbøyelighet for overtro eller selvoppfyllende profetier. Nickerson (1998) viser til at fenomenet bekreftelsesfellen er beskrevet allerede på 1600-tallet av Francis Bacon, da som en strategi for å sikre konformitet i informasjonsinnhenting, slik at nye opplysninger harmonerer med innledende hypotese.

Det er bred enighet om fenomenets eksistens, likevel hersker det uenighet rundt omtale og omfang. Kritikere som Klein (2019a) skriver at bekreftelsesfellen blir brukt for vidt, tillagt for mye vekt og mister fotfeste i akademia. Videre at bekreftelsesfellen er en bekreftelsesfelle i seg selv, og dens tilhengere er fanget i en fiksering rundt dens betydning (Klein, 2019b). Lite annen litteratur peker i retning av endret anseelse. Et litteratursøk i søketjenesten Oria datert til 22.10.21, gir en indikasjon på at bekreftelsesfellen tok seg opp i litteraturen omtrent fra 1920-tallet. Litteratursøket var begrenset til litteratur på Engelsk og Norsk. Fra 1921-1950 med søkeordet «Confirmation bias», er det 273 treff i Oria. Et tidsbegrenset søk utført 22.10.21 gir 181 833 treff. Den nyeste forskning virker å være spesielt rettet mot profesjonsyrker som legevirksomheten og politi- og rettsarbeid. Francis

Bacon er blant annet referert til i doktorgradsavhandlingen til Rachlew (2009) som omhandler justisfeil i politiets etterforskning. Det vektlegges at den første konklusjonen en lander på farger hele resten av informasjonsinnhenting mot «sannheten», dette sikrer konformitet i ledetrådene. Bekreftelsesfellen er godt omhandlet, har et moderne fokus og er godt gjort rede for, likevel gir et tilsvarende søk på «Debiasing» kun 5142 treff, og med emnebegrensningen «confirmation bias» ga søket 43 treff. Muligheten for at det eksisterer et tomrom i tilgjengelig forskning på å motvirke, påvirke eller forebygge fenomenet er til stede, og det vil være viktig å utrede og dokumentere effekt av en metode for å bedre beslutningstaking og unngå bekreftelsesfellen. Fischhoff (1982) presenterer fire utprøvde strategier for å minimere forekomst: (1) fortelle eller advare folk om dens eksistens, (2) beskrive utfall/konsekvens av feilslutning, (3) gi tilbakemelding på at noe kan være feilvurdert grunnet bekreftelsesfelle (4) gi trening med veiledning og tilbakemelding. De tre første strategiene ga lite eller ingen endring, mens den fjerde ga moderat bedring, også støttet av Milkman et al. (2009) og Rotgans og Schmidt (2019). Det som kan virke forebyggende i en spesifikk sak er dersom en får presentert tilstrekkelig og allsidig informasjon forut for avgjørelsen, diagnostiseringen eller konklusjonen (Morewedge et al., 2015; Rotgans & Schmidt, 2019).

Det hadde vært anvendelig dersom en kunne ty til enkel regelstyring ved råd eller milde advarsler, og at dette var tilstrekkelig for å få mennesker til å innse at de bør revurdere en beslutning. Atferd forekommer ikke i et tomrom, men formes av miljøet og dens stimulusendringer. Miljøbetingelsene er som samfunnet for øvrig, dynamisk, komplekst og i endring gjennom tid og rom (Cooper et al., 2016). Naturlig seleksjon ved fylogeneser forbereder bare organismen på en fremtid som likner fortiden. Likevel vil en økning eller nedgang av operant atferd antagelig være en funksjon av læringshistorie, med kulturell seleksjon og seleksjon ved atferdens konsekvenser i individets levetid (Furreboe & Sandaker,

2017). Det vil kanskje være grunn til å tro at bruken av heuristikker ville vært selektert vekkersom de kun medførte negative konsekvenser for individer eller grupper. Rachlew (2009) påpeker at forenklingsstrategier i hverdagen er nødvendig å ta i bruk dersom en skal få gjort små og mellomstore handlinger og unngå handlingslammelse. Ved interaksjon med omverden må en ty til snarveier for å håndtere den massive strømmen med stimuli en blir utsatt for, og bruken av heuristikker er ofte ubevisst (Ask & Granhag, 2005). Det å søke en meningsfull kausalitet er vanlig både i private sammenhenger, og profesjonelt. Både dommere og jurymedlemmer har vist seg å søke etter en årsaksforklaring som for dem selv er troverdig, for deretter å holde på denne som grunnlag for beslutning eller dom (Ask & Granhag, 2005). Det beste utgangspunktet for å vurdere fakta, er dersom en ikke har noen egeninteresser eller noe å tape av utfallet. Evnen til å fatte en balansert, nøytral og objektiv vurdering synker ved egeninvolvering og egeninteresse (Thurstone, 1924). Mennesker tenderer også til å lese bøker av forfattere de tidligere har kunnet assosiere seg med, eller trekker mot miljøer av folk de kan enes med (Nickerson, 1998). Dermed utsetter en seg for høyt trykk av informasjon som stemmer overens med allerede eksisterende overbevisning, og det hersker en viss fare for at de unngår forfattere eller miljøer de ikke føler samme grad av likhet med.

Beslutninger og valg kan formes til viktige utfall på individ- og samfunnsnivå (Milkman et al., 2009). Bekreftelsesfellen er et nært tilstedeværende fenomen i hverdagen som kan ta mange former (Nickerson, 1998). Opplevelsen av stress ved beslutninger øker i takt med hvor store konsekvenser det kan få om beslutningen er feil (Rachlew, 2009). En legger større vekt på positiv bekreftelse enn negativ avkreftelse, (Nickerson, 1998). Likevel skriver Nickerson at en hindrer seg selv i å finne sannheten ved å unngå å undersøke alternative hypoteser. Det er mulig at bevis for vår feilaktige overbevisning kan fungere som en aversiv stimulus, og påvirke hvordan vi tenker om oss selv, og oppleves som en

identitetskrise (The Decision Lab, 2020). Videre er mennesker sosiale vesener med trang til tilhørighet og søker konformitet i grupper. Dersom en tar et annet standpunkt enn resten av gruppen, vil en fort ikke passe inn. En vil kanskje ty til flukt- eller unngåelsesatferd som et ledd i en negativ forsterkningsprosedyre, i et forsøk på å unnsnippe aversiv stimuli. Å unnsnippe ubehag, vil øke sannsynligheten for mer av denne atferden.

Det ville vært nærliggende å tro at blant annet den ideelle, religiøse og politiske polariseringen som har preget krig og konflikter gjennom alle tider går mot slutten ved denne «kunnskapens» tidsalder. Likevel er det en fremvekst av nasjonalpolitiske og populistiske organisasjoner og partier i Norge. Det har vært flere terrorhendelser basert på høyreekstrem ideologi siden angrepet i 2011, og religion er fremdeles godt forankret til tross for at vitenskapen langt på vei har tilbakevist mye av det som hevdes i religiøse tekster. Mistillit til forskerne og autoriteter eller det å bortforklare pandemi og klimakrise med konspirasjonsteorier, forskningsfeil og hysteri kan få konsekvenser av betydning. Mange årsak virkning-analyser er mangelfulle, basert på pragmatisk sannhet, sosialt konstruert og korresponderende- eller logisk teori. Det debatteres rundt sosiale medier og algoritmers grad av påvirkning i polariseringsdebatten og spredning av konspirasjonsteorier, men det er vel liten tvil om at sosiale medier forenkler kontaktmuligheten med likesinnede og dermed treffer flere og bredere. Paradoksalt nok begrenser sosiale medier eksponering for bred og mangfoldig informasjon som belyser flere ulike perspektiver (Cinelli et al., 2021). Det dannes mer eller mindre lukkede virtuelle samfunn der menings-freder samles og forsterker hverandres allerede eksisterende meninger. Cinelli et al. (2021) omtaler dette som ekkoeffekten, og hevder at det i ekkokamre samles mennesker som fremsetter uniforme ideer og oppfatninger for deretter å få positiv sosial «*feedback*» på egne overbevisninger. Gruppepolariseringen som oppstår i slike ekkokamre forsterker meningene innad i gruppene og kan gjøre gruppens deltagere mer ekstreme i sitt syn (Cinelli et al., 2021). Med enorme

brukermasser er det tilnærmet umulig å regulere og kontrollere informasjonsflyten. Det har vist seg at falske nyheter sprer seg fortere enn reelle, og polariserte meninger får vokse seg økende ekstreme i fred (Cinelli et al., 2021).

Bekreftelsesfellen forekommer i de fleste profesjoner (Cook & Smallman, 2008). Kanskje særlig relevant innen yrker med mandat om maktutøvelse overfor andre mennesker. Eksempelvis Politi- og rettsvesen, helse og omsorg og i topp-politikken. I Norsk straffeprosess og moderne rett gjelder prinsippet om «fri bevisdømmelse». Beslutning skal treffes etter fri overbevisning på grunnlag av en samvittighetsfull prøvelse av de innsamlede bevisene, i tillegg til gitte prinsipper om sannsynlighet. Definisjonskravet til en sterk nok hypotese er at den anses å være «uten rimelig tvil», og det kan fattes en slutning dersom helheten av bevisene peker i samme retning (Bjerknes & Johansen, 2009). De lister opp validitetsutfordringer ved at troverdighetsvurderingen blir tatt på bakgrunn av bevisene som er samlet inn, ved at beslutninger blir fattet for fort eller på bakgrunn av den første historien en får presentert. Videre pekes det på bortforklaring eller tilpasning av bevis slik at de harmonere med ledene hypotese. Bratholm og Eskeland (2008) hevder at politiets etterforskning ofte tidlig penser seg inn på én teori, og at alternative ledetråder eller spor blir sett bort ifra. Etterforskningen setter sjelden søkelys på uskyldsspørsmålet, og er ofte sårbar for beslutningsfeil som i ytterste konsekvens ender i justisfeil. Justisfeil er en trussel for rettsstaten, rettsikkerheten og demokratiet som helhet. I enkelte subkulturer i politiet vil det være prestisjefyllt og forsterkende å finne informasjon som stemmer med egen hypotese, eller løse en sak fort. Rachlew (2009) peker på at politiets vektlegging av kriminalitetskontroll over rettfærdig rettsprosess, kan produsere justisfeil og medføre tap av tillitt i befolkningen. Oppklaringspress, tidspress, ressursmangel og manglende verifikasjons-kultur kan gjøre enkeltaktører i straffesakskjeden sårbare for stress og bruk av forenklingsstrategier, heuristikker og system 1-tankegang for å oppnå kravene.

I løpet av 2020-2021 har flere Norske kriminalsaker fått fornyet oppmerksomhet, noe av det som belyses i media er feil begått i etterforskningen. Dagsaktuelle saker vil eksempelvis være den kjente Baneheia-saken fra 19.mai 2000 som rystet hele nasjonen, og fikk bred mediedekning. Oppklaringspresset var høyt for allmennhetens trygghets- og rettferdighetsfølelse, og politiet kom med løfter om snarlig pågripelse. Det ble presentert teorier og fattet beslutninger basert på antagelser, og både tekniske og taktiske bevis ble vektlagt skjevt for å styrke den gjeldene teorien om to gjerningspersoner (Kvamme, 2021; Sommerset & Isungset, 2020). Det er mye som taler for at politi- og påtalemyndighetene i denne saken har siktet og domfelt en person til lovens strengeste straff på veldig mangelfullt grunnlag; teoretisk tvil, forhistorie og ønske om snarlig domfellelse. I den kjente Tengs-saken fra 1995 ble etterforskningen snevret inn omtrent umiddelbart i stedet for å omfavne flere teorier bredt. Etterforskningene bærer preg av å være tids- og prestisjepresset, grunnet høy ideell karakter. Det er fattet beslutninger basert på svært begrenset informasjonsgrunnlag med mål om hurtige siktelsener, i stedet for å belyse saken objektivt. Den kjente «fetteren» som raskt pektes ut som syndebukk og måtte flytte ut av landet, ble renvasket i 2021 etter 21 år med familiekonflikt og offentlig mistanke (Ighoubah, 2021). Beslutningsfeil er en betydelig ressurskostnad for både individer og samfunnet (Milkman et al., 2009). Dette stiller krav til storsamfunnet om å utarbeide verktøy og hjelpemidler som kan benyttes i situasjoner med «sårbarhetsfaktorer».

Det er diskrepans mellom hvordan mennesker rasjonelt sett burde foreta valg i forhold til normativ teori, kontra hvordan en i realiteten tar beslutninger (Angner, 2016). Begreper som «bounded rationality» omtales av flere primært innen atferdsøkonomien, og innebærer begrensninger i vurderingsevne blant annet ved godta løsninger som ikke er optimale. Kjente atferdsøkonomer som Richard Thaler, Daniel Kahneman og Herbert Simons uttaler seg om menneskets vurderingsevne som mindre mekanisk enn tidligere antatt, vi er mer komplekse

og dynamiske. Dette innebærer at mennesker ikke er rasjonelle maskiner som «prosesserer» informasjon logisk og matematisk. Vi er dårlige på å vurdere og predikere validitet, og kan ofte fatte beslutninger påvirket av hvordan noe fremstår, hvordan noen ser ut eller hvordan en ønsker at noe skal være (Tversky & Kahneman, 1974). Wason (1960) skriver at det fremstår som få «intelligente unge mennesker» på eget initiativ har fokus i hverdagen på å teste egen overbevisning. I forskningssituasjon kan en ved nøyere kalkulering vurdere andre alternativer til hypoteser, dersom en blir gjort oppmerksom på at en bør revurdere. Det som kreves for å endre på dette er at individet selv går imot egen overbevisning, og ønsker å teste troverdigheten til egne teorier ved falsifisering. Dette vil vel være en presentasjon av kontroll og rasjonell atferd som tidligere er beskrevet som noe menneske ikke behersker så godt. Wason (1960) anslår muligheten for at denne type beherskelse og kapasitet kan læres og kultiveres. Han har ved flere eksperimenter tatt for seg menneskets valgførelse, bekreftelsesfellen, rasjonalitet og «fornuft». De første eksperimentene var nok ment som en spe begynnelse på å kunne måle fenomenet bekreftelsesfellen, som ellers fremstår som abstrakt. Flere av studier foretatt av Wason (1966; 1968; 1969), Johnson-Laird og Wason (1970), og Wason og Shapiro (1971) blant annet Wason's «*four-card problem*», legger grunnlaget for dette eksperimentet. Deltageren blir presentert for fire kort med ulike stimuli fra to ulike kategorier (se stimuloversikt ved tabell 1), og en påstand formulert som «hvis P så Q». Eksempelvis vil den ene kategorien kunne være biltyper representert ved merke Toyota og Volvo, og den andre kategorien frukt representert av type eple og pære. Deltagernes oppgave blir deretter å undersøke om påstanden eller regelen er korrekt, eller kan falsifiseres ved å snu ett eller flere av kortene med stimuli referert til som; P, Q, \bar{P} eller \bar{Q} . Det er i tidligere studier vist til at majoriteten finner det nødvendig å snu P, som er korrekt og vil bekrefte påstanden. I tillegg velger flere å snu Q, men å snu dette kortet vil ikke kunne

avkrefte regelen eller tilføre ny informasjon. Det korrekte vil være å snu P og \bar{Q} , dette vil tilføre ny informasjon og en undersøker om hypotesen kan avkreftes, men de fleste er tilbøyelige til å kun verifisere og ikke falsifisere. Wason's fire-kort problem utfordrer deltagerne til å vurdere og velge hvilke kort som er nødvendige å snu, men privatatferden kan ikke observeres eller måles av andre. Catania (2013) har omtalt bekreftelsesfellen som "*confirmatory bias*", og forsøkt å presentere et observasjonsgrunnlag. Han foreslår å se prosessen som observerende responser da oppgaven består i å agere på diskriminativ stimuli med atferden «snu kort» for å sjekke hypotesen. Catania hevder at en ikke etterstreber å finne vilkårlig eller nøytral informasjon, men informasjon som leder til forsterkning. Catania (2013) drøfter muligheten for at vi over tid har lært at noe enten er enten riktig eller feil, og kan ikke være begge deler. Derfor vil en begrense gjetningen i tråd med denne teorien, og kanskje tro at en har bekreftet noe uten å vurdere muligheten for avkreftelse. Videre skriver Catania (2013) at den observerende responsen rettes mot det vi ønsker å vite, og kanskje ønsker en ikke å få informasjon om noe aversivt. Wason (1968) påpeker at vårt noe ensidige fokus på å kun verifisere en påstand er standhaftig og motstandsdyktig mot endring. Det er sjelden en går inn i en vurderingssituasjon med en utgangsholdning om at noe er feil, falskt eller ikke stemmer da læringshistorien «*prompter*» en til å forvente en relasjon av sannhet eller korrespondanse med påstanden.

Deltagerne i dette eksperimentet skal utsettes for to uavhengige variabler i et forsøk på å demonstrere effekt ved redusert bekreftelsestendens. Det diskuteres noen faktorer som kanskje kan belyse resultatet utover bedring i antall rett etter tilbakemelding(ene). Dette er rekkefølgeeffekt ved startbetingelsene, antall forsøk nødvendig for å fullføre oppgavene i de ulike startbetingelsene og tidsbruk på å lese tilbakemelding for utvalgte deltagere.

Metode

Det er benyttet kvantitativ metode da innsamlet data tar sikte på å være målbart, og skal presenteres i tabeller og grafer. Dataen skal egne seg for visuell inspeksjon som et ledd i kvalitetskontrollen, og visuell analyse for at leser enkelt og raskt kan få oversikt over resultatet. Designet er «*mixed*», med både sammenligning av to grupper, og enkeltsubjekt-design med deltageren som egen kontrollperson. Formålet med designet er å sikre eksperimentell kontroll, reliabilitet og validitet i innsamlingen og resultatet. Med enkeltsubjekt-designet er formålet primært å undersøke om intervensjonen har effekt på individets atferd, om effekten er replikerbar over flere av individene og om effekten lar seg generalisere utover treningen gitt i eksperimentsituasjonen. Det er hovedsakelig benyttet et ABACA og ACABA som i henhold til prosedyre ender ved ABA eller ACA i noen tilfeller. Det kartlegges et utgangspunkt, «*steady state*» hos deltagerne før intervensjon, for så å innføre en uavhengig variabel og observere effekt på den avhengige variabelen. Målet vil være å manipulere miljøet for så å kunne måle dette og demonstrere funksjonell relasjon mellom miljøbetingelsene/intervensjonen og atferden (Cooper et al., 2016). Det er tre baselinjerunder, og to intervensjonsfaser der deltagerne blir utsatt for trening ved to uavhengige variabler. Eksperimentet forutsetter ikke en reversering av effekt, men måler heller forskjellen i grad av endring. Sikkerheten ligger i ekskluderingskriteriene, der de som antagelig ikke vil ha noen nytte av intervensjon og vise noen endring blir ekskludert etter første baselinjerunde. Det vil være relevant å sammenligne de tre rundene med baselineoppgaver og kontrolloppgavene. Det skal også undersøkes om det er nærliggende å tro at deltageren har effekt av regelstyringen og tar seg tid til å lese instruks, eller holder på sin egen overbevisning om hva som er rett eller galt. Gruppe-designet tar sikte på å sammenligne rekkefølgeeffekten på presentasjonen av de uavhengige variablene. Det vil kunne være interessant å se på om den ene uavhengige variabelen produserer mer tendens til

mønsterlæring, eller økt refleksjon. Dataen som skal vektlegges i resultatdelen er primært antall rett og galt for individet, og sammenligne gruppene. Videre skal det presenteres mønster i seleksjon av kort som skal snus ved første baselinesesjon på gruppebasis, dette for å undersøke utgangssårbarhet for bekreftelsesfellen. Det er inkludert i oppgaven en pilotstudie vedlagt i appendiks (tabell 5) som tar for seg hvor lang tid det gjennomsnittlig tar å lese tilbakemeldingene og instruksjonen. Programmet registrerer hvor lang tid deltagerne bruker på å lese tilbakemeldingen de får. Deltagernes tidsbruk skal settes opp mot resultatet fra pilotstudien for å undersøke om det er nærliggende å tro at det er brukt tilstrekkelig tid på å lese tilbakemeldingen ved feil svar, eller om deltageren fortsetter å holde på egen overbevisning, gjetning eller mønsterlæring.

Deltagere

Deltagerne ble rekruttert på ulike måter. Eksperimentator tok kontakt med politihøgskolen i Oslo og politihøgskolen i Stavern, og koordinator for skolene la ut informasjon om prosjektet på student-hjemmesiden. Det ble kontaktet ulike soningsanstalter, og daglig leder ved den ene var behjelpelig i rekrutteringsprosessen. Det ble også forespurt i eksperimentators nettverk om noen ønsket å delta. Totalt antall deltagere var 21, hvorav 9 ble utsatt for betingelse 1 (utvidet tilbakemelding) og 12 ble utsatt for betingelse 0 (generell tilbakemelding). Det var ulikt antall deltagere pr. gjennomføring og variabelgruppe grunnet randomiseringsprosedyre og deltageres tilgjengelighet. I tillegg ble det foretatt én gjennomføring på et senere tidspunkt. Det er en kjønnsfordeling på om lag 50-50 menn og kvinner totalt i eksperimentet (10 menn og 11 kvinner), og en aldersspredning fra 21 til >70 år. Det er etterstrebet deltagere med bred diversitet både med tanke på alder, etnisitet, utdanning og yrkesbakgrunn. Det er bevisst unngått benyttelse av kun «*white male college students*». Ekskluderingskriteriene er; (1) Ikke fullført eksperiment, (2). minst fem av seks riktige ved første baseline, (3) deltager som opplyste om for store utfordringer med å forstå

instruks grunnet språklige vansker, (4) deltagers eget ønske om å trekke seg, (5) teknisk svikt som antas å kunne påvirke dataene. Totalt er 7 deltagere ekskludert, fire damer og tre menn. Alle deltagerne over 70år trakk seg grunnet vansker med det akademiske språket, med formen på eksperimentet (teknologien med «*tablet*») og én oppgave at han følte at det ble for anstrengende. Én deltager trakk seg etter fem baselineoppgaver, hun ønsket ikke fullføre. Én deltager trakk seg etter at det oppsto en teknisk feil, hun ønsket ikke fortsette da hun ikke hadde tid til å vente på at det ble rettet opp i. Tre deltagere ble ekskludert av eksperimentator da de hadde alle de innledende baselineoppgavene rett. Det inkluderte datagrunnlaget i studien er dermed basert på 7 deltagere i gruppe 0, og 7 deltagere i gruppe 1.

Setting

Deltagelse har foregått ved tre ulike anledninger grunnet vanskeligheter med å samle mennesker i pandemitid. Det er likevel etterstrebet å få til om lag samme miljøbetingelser i forhold til lyddisiplin eller andre forstyrrende elementer, tilrettelegging for selvstendig oppgaveløsning og identiske instruksjoner. Gjennomføring 1 foregikk i et klasserom på politihøgskolen i Stavern med fem deltagere. Deltagerne satt med god avstand til hverandre og kunne ikke se hverandres skjermer. De fikk ikke anledning til å prate sammen, og gjennomføringene ble kjørt omtrent samtidig eller så tett etter hverandre at eksperimentator hadde kontroll kommunikasjon. Gjennomføring 2 fant sted på personalkontoret til en soningsanstalt, alle de fire ansatte gjennomførte eksperimentet samtidig med noe ulik slutføringstid. Gjennomføring 3 fant sted på eksperimentators kontor, de fem deltagerne fikk ikke snakke sammen før alle hadde ferdigstilt gjennomføringen. Flexibilitet fra eksperimentators side var nødvendig grunnet Covid 19, og noen individuelle gjennomføringer var også nødvendig på ulike tidspunkt. Lokasjonen var eksperimentators kontor, og deltagere med individuelle gjennomføringer kjente ikke hverandre.

Deltagerne fikk informasjon om at det ikke blir samlet inn personopplysninger utover at eksperimentator er i besittelse av mailadressen som ble benyttet i rekrutteringen. Det ble ikke knyttet persondata opp mot eksperimentbesvarelsen. De ble informert om at resultatene og deltagelsen var anonym, og fikk oppgitt en kontaktperson ved spørsmål om personopplysninger. De spurt om de var villige til å delta, og informert om at de kunne uten fare for sanksjoner avslutte eksperimentet etter eget ønske. Videre ble det gitt informasjon om at eksperimentet dreier seg om elektronisk oppgaveløsning. Eksperimentator oppga et tidsestimert på opp mot 35 minutter, og at tidsfristen pr. oppgave er satt til 180 sekunder. Det er satt en tidsfrist av etiske hensyn, da det kan virke utmattende å sitte forlenge med én oppgave, eller eksperimentet i sin helhet. Det ble ikke gitt noen insentiver for deltagelse, men alle deltagerne fikk tilbud om faglig gjennomgang etter fullført deltagelse. Deltagerne ble av programmet bedt om å oppgi kjønn og omtrentlig alder. Deltagergruppene ble delt i to for hver gjennomføring, og utsatt for den ene startbetingelsen. Programmet bytter så betingelse automatisk ved fullført første del. Formålet med todelingen var å ivareta den eksperimentelle kontrollen ved å undersøke om eventuelle utslag på data skyldes rekkefølgen på de uavhengige variablene, eller annen miljøpåvirkning på nettopp den gruppen.

Apparatur

Microsoft Surface, DESKTOP-NLL9R18, Intel pentium CPU 44515Y 1,6ghz, 4gb RAM, 64-bit Operating system, Windows 10 pro

Program: ConfBB1

Datafil: Data blir automatisk registrert og lagret i egen datafil.

Prosedyre

Deltagerne fikk utdelt hver sitt «*tablet*» med programmet «Confbb1» allerede installert. En kunne velge mellom «*touch-screen*» eller bruke «*pad*» med musepeker. Eksperimentet startet med en innledende instruks der deltagerne ble forklart at de ville bli vist

ulike kort, fra to ulike kategorier med enten navn, symboler eller figurer på. I hver oppgave vil det bli presentert en påstand «hvis P på den ene siden, så vil det være Q på den andre siden». Deltagerens oppgave vil være å selektere kort som må snus for å undersøke om påstanden er korrekt, eller kan falsifiseres.

Tabell 1 A, B og C

Oversikt over kort med stimuli i alle 3 fasene

A) baseline

P	Q	\bar{P}	\bar{Q}
Sirkel	Rød	Trekant	Grønn
Volvo	Glass	Toyota	Kopp
Vokal (A)	Partall (4)	Konsonant (D)	Oddetall (7)
Galdhøpiggen	Hankjønnsymbol	Glittertind	Hunkjønnsymbol
Ost	Penn	Syltetøy	Blyant
Gorilla	Undulat	Sjimpanse	Ara-papegøye

B) Trening

P	Q	\bar{P}	\bar{Q}
Øl	22 år	Coke	16 år
Gitar	Eple	Trompet	Pære
Elefant	Tulipan	Kanin	Valmue
Snekker	Fiskeboller	Politi	Spagetti
Skrik	New York	Mona Lisa	Moskva
Joker	Vi selger ukeblader	Rema 1000	Vi selger ikke ukeblader

C) Generalisering

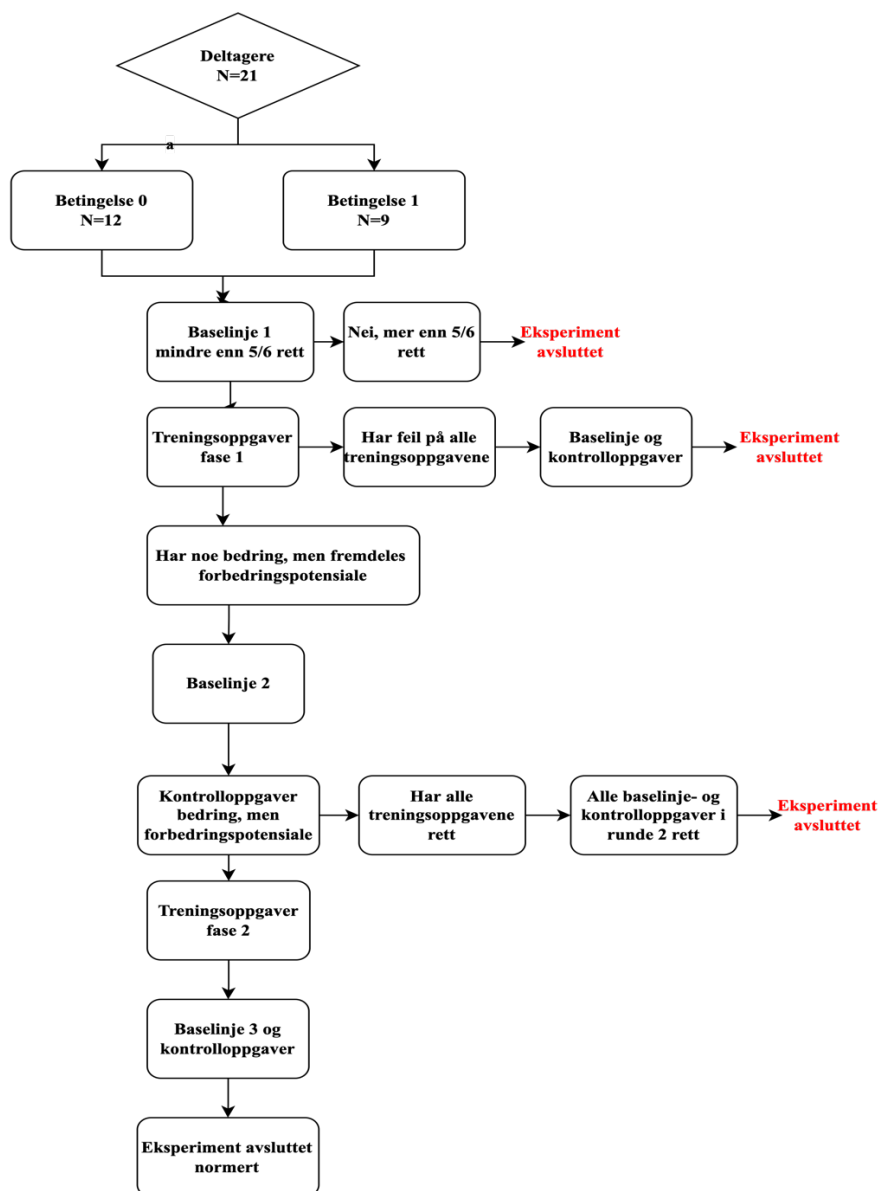
P	Q	\bar{P}	\bar{P}	\bar{P}	\bar{Q}	\bar{Q}	\bar{Q}
Eple	Queen	Kiwi	Pære	Druer	Abba	The Beatles	The Rolling stones
Yatzy	Bipolar lidelse	Stige-spill	Monopol	Sjakk	ADHD	Autisme	Schizofreni
Saturn	Italia	Mars	Uranus	Neptun	Ghana	Peru	Mongolia

Som vist i tabellen er fasene delt opp i tre med seks oppgaver og fire kort, med unntak av tre oppgaver og 8 kort i kontroll-og generaliseringsfasen. Det vil være interessant å undersøke om læringseffekten endres ved rekkefølge. Betingelse 0 innebar at deltagerne først

kun fikk tilbakemelding på om svaret var riktig eller galt. Betingelse 1 innebar at deltagerne ble forklart hvorfor svaret var rett eller galt. Kontrolloppgavene innebar utvidet mengde stimuli for å undersøke om ferdigheten kan generaliseres utover rammene i treningssituasjonen. Grunnet en programmeringsfeil ble det innhentet data fra deltager 5 og 9 som avviker fra tiltenkt prosedyre. Konsekvensene var mer data.

Figur 1

Flytdiagram over prosedyren



Deltagerne ble utsatt for ulik rekkefølge av de uavhengige variabler etter baseline betingelsene, enten ved at tilbakemeldingen først var generell eller først utvidet. De uavhengige variablene er de to formene for tilbakemelding i treningen som opererer på den avhengige variabelen, og søker å endre deltagerens responsmønster. Dersom en deltager svarer 0/6 rette på første treningsrunde, vil ikke programmet utsette deltageren for en ny runde med trening, men gå rett til baselineoppgavene i runde tre. Det vil ikke være etisk rett eller formålstjenlig å utsette deltagerne for langvarig eksperimentdeltagelse dersom de ikke mestrer oppgaven innen rimelig tid eller rimelig mengde regelstyring. Deltagerne fikk muligheter til å svare på kontrolloppgavene etter baselineoppgavene i runde to, og etter baselineoppgavene i runde tre. Formålet vil blant annet være å undersøke om deltagerne hypotesetester og falsifiserer, eller har lært seg et fast responderingsmønster uavhengig av formålet med treningen. Dersom det ikke forelå ekskluderingskriterie etter første runde med trening, ble deltagerne utsatt for den andre uavhengige variabelen for å undersøke rekkefølgeeffekt.

Deltagerne blir først presentert for 6 baselineoppgaver hvor det må gjøres minst én feil for å komme videre til første runde med treningsoppgaver. Hvis alle 6 baselineoppgavene er korrekt utført, vil det kunne forstås som at deltageren ikke har høy grad av bekreftelsestendens ved de gitte rammene for eksperimentet, og deltageren vil ikke bli utsatt for trening. Instruksen for del to gir beskjed om at de vil motta tilbakemelding på svaret, enten ved generell tilbakemelding, riktig/galt, eller en utvidet og forklarende tilbakemelding. De får maks tre minutter per oppgave, og informasjon om at de i første del ikke får beskjed om svaret er rett eller ikke. Den utvidede tilbakemeldingen forklarer både hvorfor et svar er riktig, og hvorfor et svar er feil. Deretter ble de utsatt for seks nye treningsoppgaver, der de ikke kom videre til neste oppgave før de enten svarte riktig, eller brukt opp tiden. Etter hver treningsoppgave kommer ny instruks og 6 baselineoppgaver og tre kontrolloppgaver.

Resultat

Av totalt 21 deltagere svarte 2 rett på alle baselineoppgavene og 1 svarte rett på 5/6 oppgaver. Disse 3 deltagerne ble ekskludert og ikke utsatt for treningsintervensjonene. Fire deltagere trakk seg av andre årsaker som beskrevet tidligere. Det er undersøkt effekten av treningen ved bedring fra baselinje 2 og eventuelt baselinje 3 fra første baselinje. Bedring måles i økning i antall rette responser fra utgangspunktet, og rett respons regnes som responskombinasjonen P og \bar{Q} sammen.

Tabell 2 A og B

A) Resultatoversikt gruppe 0, startbetingelse generell tilbakemelding

N	BL	Intr.v	BL	KO	Intr.v	BL	KO
1	0/6	5/6	5/6	3/3	6/6	6/6	3/3
2	0/6	3/6	6/6	0/3	6/6	6/6	0/3
3	0/6	3/6	4/6	1/3	6/6	5/6	3/3
4	0/6	2/6	1/6	0/3	0/6	3/6	0/3
5	0/6	4/6	6/6	3/3	6/6	5/6	3/3
6	0/6	3/6	4/6	3/3	4/6	5/6	3/3
14	0/6	4/6	6/6	0/3	6/6	6/6	3/3

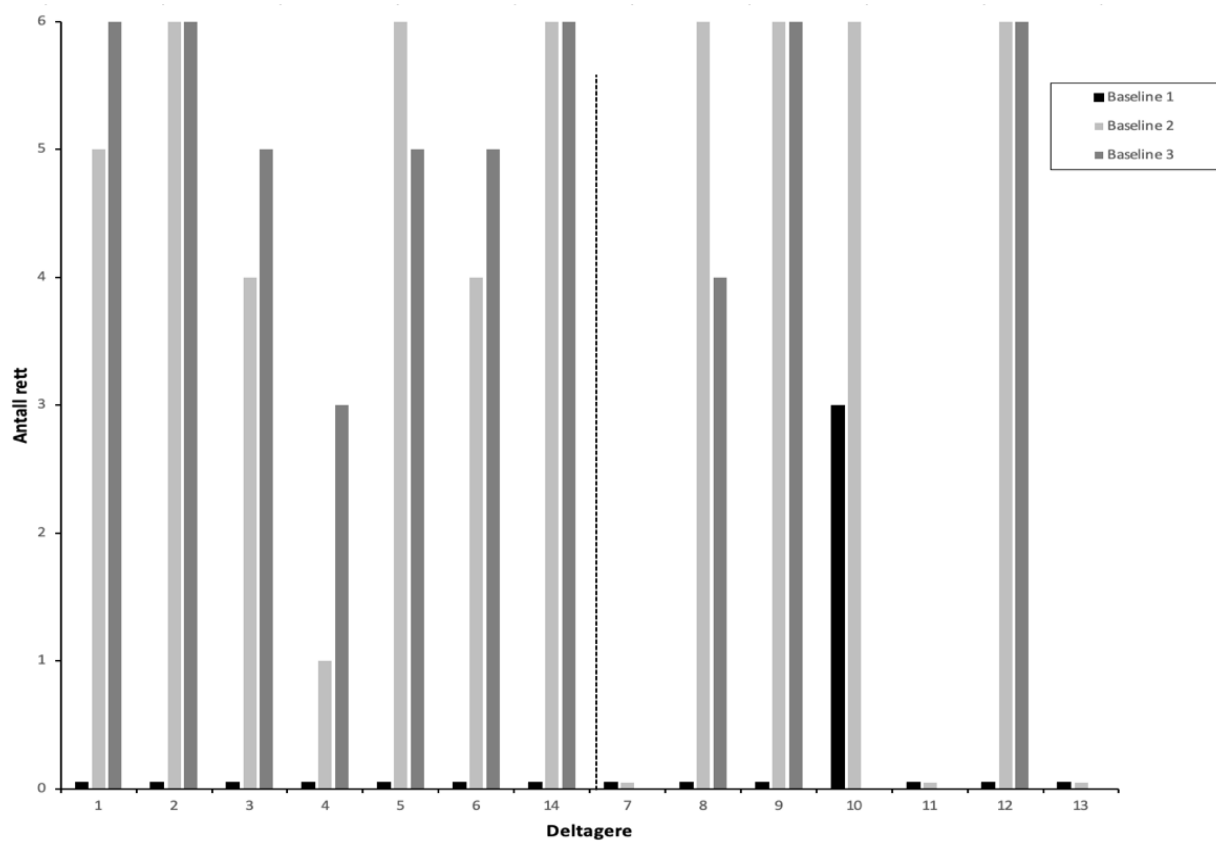
B) Resultatoversikt gruppe 1, startbetingelse utvidet/forklarende tilbakemelding

N	BL	Intr.v	BL	KO	Intr.v	BL	KO
7	0/6	0/6	0/6	0/3	-	-	-
8	0/6	3/6	6/6	0/3	4/6	4/6	0/3
9	0/6	5/6	6/6	3/3	6/6	6/6	3/3
10	3/6	3/6	6/6	3/3	-	-	-
11	0/6	0/6	0/6	0/3	-	-	-
12	0/6	4/6	6/6	1/3	6/6	6/6	0/3
13	0/6	0/6	0/6	0/3	-	-	-

Alle deltagerne i gruppe 0 har effekt av første treningsrunde, og utviser bedring fra baselinje 1 til 2, samt noe bedring fra baselinje 2-3. Tre deltagere har 100%bedring, én deltager 83% bedring, to deltagere har 66,7% bedring, og én deltager 16,7% bedring. Fra baselinje 2 til 3 har tre deltagere 16% bedring, én deltager 33% bedring og én deltager 16.6% nedgang. Resultatet fra kontroll-og generaliseringsrundene viser at 3/7 deltagere mestret disse oppgavene etter første runde, mens 5/7 mestret generaliseringsoppgavene etter andre runde. Etter første runde med trening hadde 4/7 i gruppe 1 100% bedring fra utgangspunktet, mens 3/7 hadde 0% bedring. Ved første runde kontroll-og generaliseringsoppgaver svarte 2 deltagere alle rett, og 5 alle feil. De deltagerne som hadde 0% bedring fra baseline 2, og 0/3 rett på kontroll-og generaliseringsoppgavene ble ekskludert for videre trening. Kun 3/7 ble utsatt for treningsbetingelse 2. Av deltagerne i gruppe 1 hadde 0% rett på de siste kontroll-og generaliseringsoppgavene. Grunnet programmeringsfeil ble deltager 9 utsatt for mer trening på tross av alle rett. Deltager 8 og 12 fikk mer trening i tråd med tiltenkt prosedyre, mens deltager 7, 11 og 13 ble avsluttet grunnet manglende progresjon. Deltager 8 viste reverserende effekt fra baselinje 2 til 3 (33% nedgang).

Figur 2

Individuell bedring fra baselinje 1 til baselinje 2 og til baselinje 3

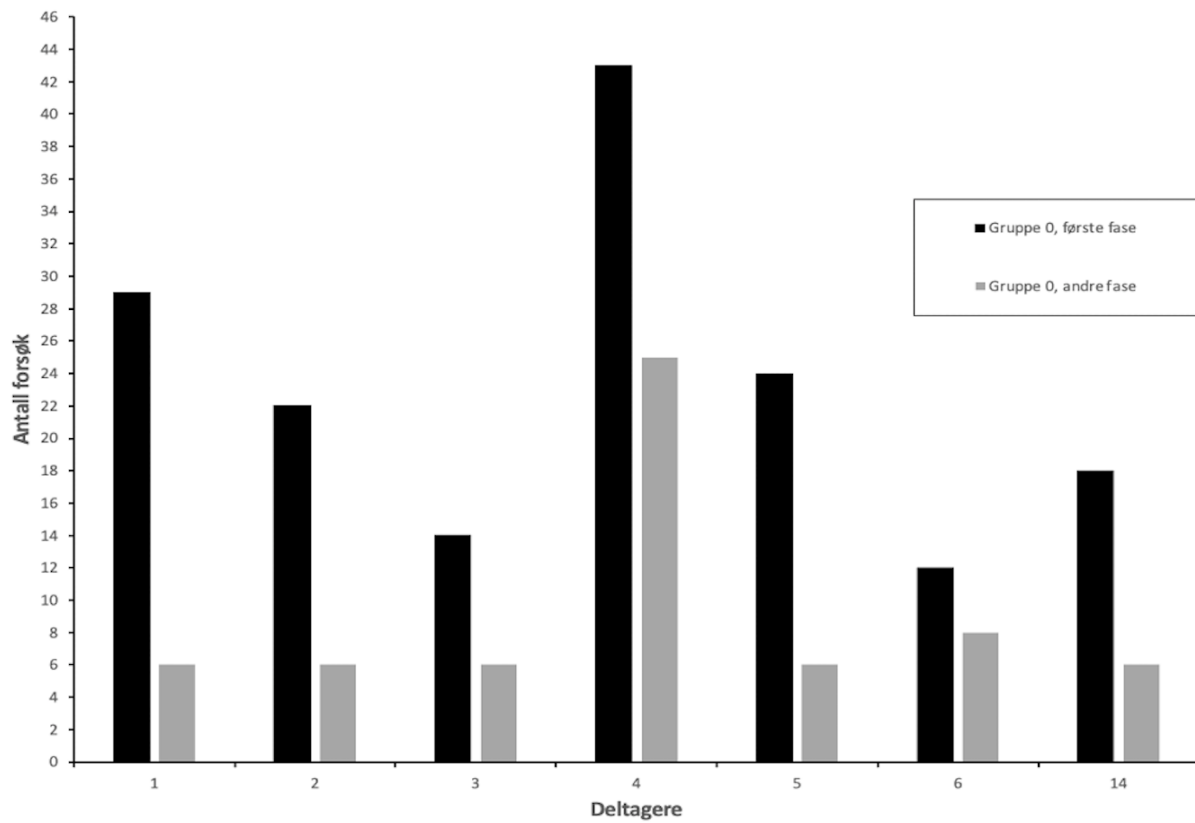


Note. Figuren viser differansen mellom antall rett i de tre respektive baselinjerundene per deltager. Delestreken markerer skille mellom deltagerne i gruppe 0 og gruppe 1. Deltager 7, 10, 11 og 13 har kun to markører da de ikke ble utsatt for baselinje 3.

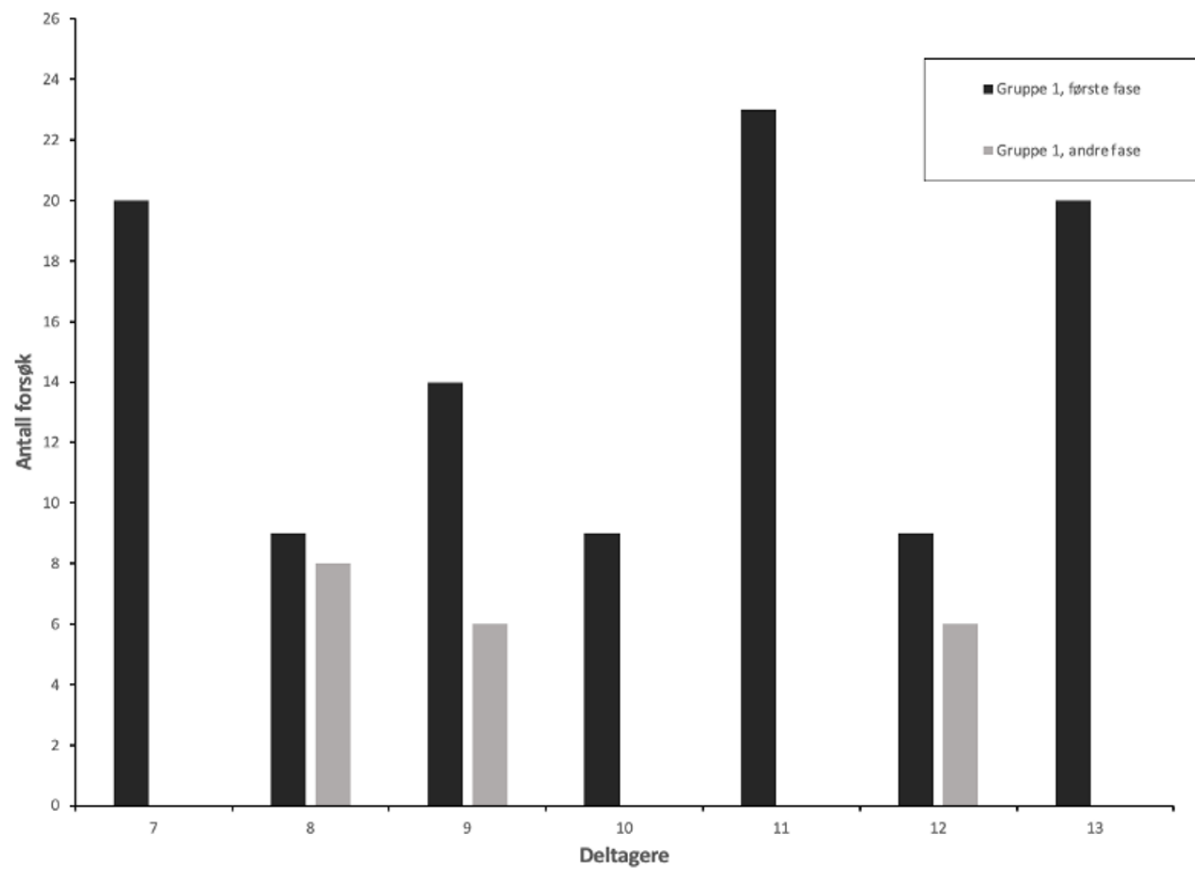
Det er undersøkt antall forsøk deltagerne i de respektive gruppene brukte på å fullføre treningsintervensjon 1, sammenlignet med treningsintervensjon 2. Formålet er å sammenligne mellom gruppene og alle individene med seg selv og mot de andre. Det er stor nedgang i forsøk fra runde 1 til 2 for alle deltagerne i begge gruppene.

Figur 3 A og B

A) Antall forsøk deltagerne i gr. 0 brukte på første sammenlignet med andre treningsfase



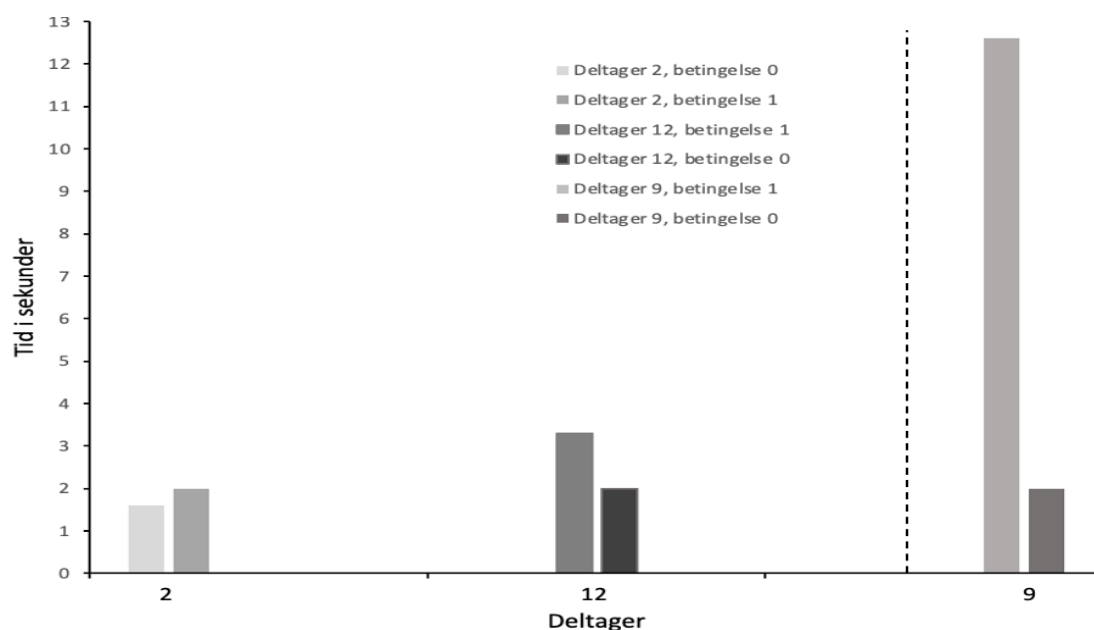
B) Antall forsøk deltagerne i gr. 1 brukte på første sammenlignet med andre treningsfase



To deltagere skilte seg ut i datamaterialet, det er foretatt en gjennomgang med formål om å belyse eventuelle faktorer som kan ha spilt inn. Inkluderingskriteriet var 6/6 rett på andre og tredje baselinjerunde, men 0/3 på begge rundene med kontroll-og generaliseringsoppgaver. Til sammenligning er det plukket ut deltager 9 som fikk 6/6 rett på baselinjeoppgavene, samt 3/3 rett på rundene med kontrolloppgaver. Det er målt tiden deltagerne har hatt tilbakemelding på svaret sitt tilgjengelig, til de har trykket «neste» oppgave. Deltager 2 (gruppe 0) brukte ved første rett respons 1,5 s og gjennomsnittlig 1,6 s per oppgave første treningsrunde. Ved ny treningsbetingelse brukte han gjennomsnittlig 2 s. Deltager 12 (gruppe 1) brukte i gjennomsnitt 3,3 s gjennom alle treningsoppgavene med utvidet tilbakemelding, og 2 s på den generelle tilbakemelding. Deltager 9 (gruppe 1) brukte i gjennomsnitt 12,6 s på å lese tilbakemelding ved to rette, og 2 s på den generelle tilbakemeldingen. Hun brukte 36 s fra første tilbakemelding om rett svar til hun trykket «neste».

Figur 4

Oversikt over tiden deltagerne valgte å ha tilbakemelding tilgjengelig

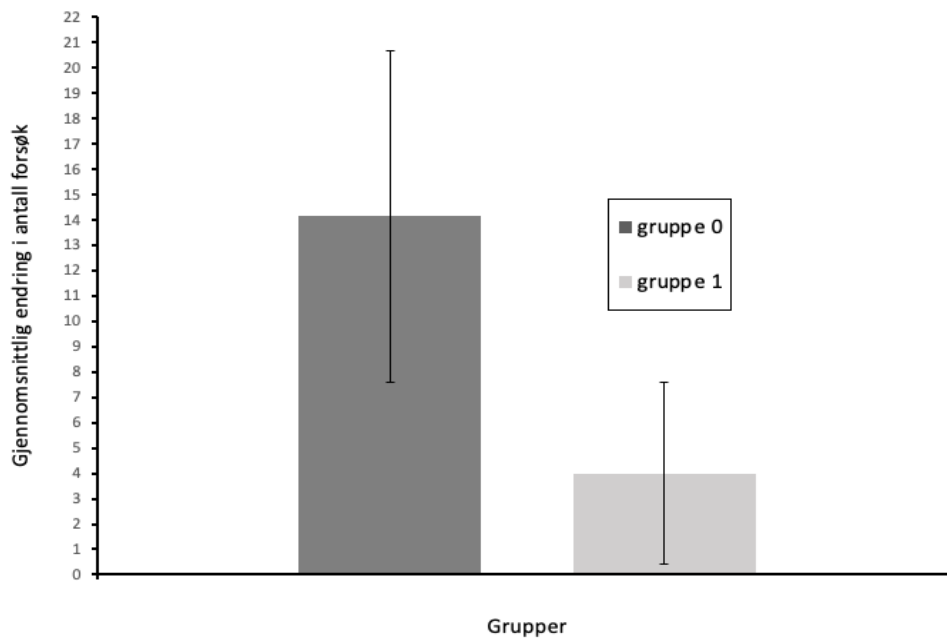


T-tester for to uavhengige variabler

Ved sammenligning av gjennomsnittlig endring i antall rett fra baselinje 1 til 2, ble det ikke funnet signifikant forskjell. Gruppe 0 ($M=4,57$, $SD=1,81$) sammenlignet med gruppe 1 ($M=3$, $SD=3,87$), $t(12)=0,97$, $p=.35$. Ved sammenligning av gjennomsnittlig endring av antall rette fra baselinje 2-3 ble det ikke funnet signifikant forskjell. Gruppe 0 ($M=0,57$, $SD=0,98$) sammenlignet med gjenværende deltagere fra gruppe 1 ($M=-0,67$, $SD=1,15$), $t(8)=1,75$, $p=.12$. Ved sammenligning av gjennomsnittlig differanse i antall forsøk fra intervensjonsrunde 1-2 ble det funnet signifikant forskjell mellom gruppe 0 ($M=14,14$, $SD=6,54$) sammenlignet med de 3 deltagerne i gruppe 1 ($M=4$, $SD=3,6$), $t(8)=2,47$, $p=.039$. Det er en større gjennomsnittlig endring i antall forsøk fra treningsfase 1 til 2 i gruppe 0 enn gruppe 1. Ved sammenligning av gjennomsnittlig antall forsøk i intervensjonsrunde 1 mellom gruppene gruppe 0 ble det ikke funnet signifikant forskjell, gruppe 0 ($M=23,14$, $SD=10,52$) sammenlignet med gruppe 1 ($M=14,86$, $SD=6,09$), $t(12)=1,8$, $p=.097$. Ved sammenligning av gjennomsnittlig antall forsøk i intervensjonsrunde 2 mellom gruppene ble det ikke funnet signifikant forskjell, gruppe 0 ($M=9$, $SD=7,09$) sammenlignet med de 3 deltakerne i gruppe 1 ($M=6,67$, $SD=1,15$), $t(8)=0,55$, $p=.60$.

Figur 5

Gjennomsnittlig endring fra intervensjon 1 til intervensjon 2



Note. Figuren viser differansen mellom gjennomsnittlig antall forsøk nødvendig fra intervensjon 1 til intervensjon 2 for hver gruppe, målt i antall forsøk det tok å fullføre runden. Det er signifikant forskjell mellom gruppene i differanse mellom intervensjonsrundene.

Tabell 3

Responsforekomst i gruppene ved første baselinjerunde før intervensjon

Gruppe	P	\bar{P}	Q	\bar{Q}
0 N=7	42	3	20	3
1 N=7	42	7	20	9

Begge gruppene viser tendens til å velge stimulus P enten alene eller i om lag halvparten av svarene sammen med Q. Respons \bar{Q} er valg tre ganger i den ene gruppen og ni ganger i den andre gruppen, men ikke sammen med P. Deltager 10 og 13 har valgt \bar{Q} fire ganger hver som eneste respons, og \bar{P} tre ganger hver. Utgangssårbarheten for bekreftelsesfellen fremstår lik i begge gruppene.

Diskusjon

I dette eksperimentet er det forsøkt å demonstrere korrelasjon mellom redusert forekomst av bekreftelsesfellen og to typer tilbakemelding. Dataene viser en markant nedgang i responsmønster preget av bekreftelsesfellen fra første til andre baselinje i begge gruppene. På individuelt nivå er det noe jevnere fordeling av bedring i gruppe 0 (startet med generell tilbakemelding) der alle deltagerne har utvist en viss grad av bedring. I gruppe 1 (startet med utvidet tilbakemelding) har deltagerne enten hatt en 100% bedring, eller en 0% bedring. Det er liten tvil om at intervensjonen hadde effekt på majoriteten av deltagerne, de som ikke viste effekt tilhørte gruppe 1. Ved visuell analyse av datasettene i gruppe 1, kan det virke som at det er en klar todeling i hvordan den utvidede tilbakemeldingen opererer på deltagerne. Enten etableres en klar forståelse av hvordan oppgaven skal løses ved hjelp av regelstyring, og de får 100% bedring, samt mestrer kontrolloppgavene. I motsatt tilfelle virker ikke den uavhengige variabelen inn på deltageren i særlig grad, det etableres liten forståelse og bedringen som eventuelt kommer ved baselinje 3 bærer preg av å være et produkt av en selvstendig mønsterlæring, heller enn korrekt problemløsning og tiltenkt regelstyring. Etter flere repetisjoner er det mulig at valg av kort som må snus blir kontingensformet, og deltagerne svarer rett respons i overensstemmelse med regelen uten at regelen spiller en rolle som egen variabel. Den individuelle forskjellen mellom gruppedeltagerne er noe overraskene og finner lite støtte i studiet til Holth (2019) som har sett på tilsvarende oppgaver og trening med tilbakemelding. De fant 71% bedring for begge gruppene i baselinje 2, med de samme tendensene til responsforekomst i første baselinje. Likevel er bedringen i den studien betydelig for alle individene i begge gruppene, og prestasjonen tilnærmet lik i begge gruppene fra baselinje 1 til baselinje 2. Tendensen er lik over alle deltagerne og mellom begge gruppene på antall rette, og hvor mange som gikk videre til neste runde med trening. Funnene til Holth (2019) viser også til et likt resultat

mellom gruppene på kontroll-og generaliseringsoppgavene på henholdsvis 50% og 53% mestring av disse. Gruppe 0 i dette eksperimentet steg fra 43% rett til 71% rett, mens de i gruppe 1 sank fra 29% rett til 0%. Den lave skåren for gruppe 1 er påvirket av frafall av deltagere etter første treningsrunde, men allerede ved første runde oppgaver er det større forskjeller mellom disse gruppene enn det som er resultatet i Holth (2019). Denne forskjellene mellom gruppene er likevel ikke regnet som signifikant. Deltager åtte viste reverserende effekt fra baselinje 2 til 3. En nærmere undersøkelse av dataene viser at den ene feilen kan ha vært et dobbeltrykk da han brukte 0,1 sekund på å svare og han ikke har valgt noen stimuli. Oppgave 15 er også besvart feil, men med identisk responsmønster som i innledende baselinje (P og Q). Data fra deltager fem etter andre baselinje er samlet inn grunnet en programmeringsfeil, men presenteres likevel da dataen først er hentet inn. Deltageren har hatt en nedgang i antall rett fra baselinje 2 til 3 som kanskje kan forklares som en feiltrykk grunnet høy trykk-frekvens på denne oppgaven, samt ulogisk responskombinasjon sett opp mot tidligere responsmønster.

En visuell analyse av deltagerdata som ikke viste noen effekt av intervensjonen i gruppe 1 ble gjennomført med tanke på å identifisere medvirkende faktorer. Deltager 7 var 46-50år, ikke student. Deltager 11 var 61-65år, ikke student. Deltager 13 var 26-30år, ikke student. Disse tre deltagerne brukte markant flere forsøk på å fullføre første treningsrunde enn de andre deltagerne på henholdsvis over 20 forsøk. Det er vanskelig å vurdere årsakssammenheng med så mange ukjente faktorer, likevel er de typiske eksperimentdeltagerne ofte relativt unge studenter i «studentmodus», oppdatert og kjent med et akademisk språk, vitenskapelig metode og vant med å sitte stille og lese akademiske tekster. På bakgrunn av dette veldig begrensede utvalget av deltagere kan det kanskje tenkes at alder eller akademisk bakgrunn kan spille inn på deltagerens prestasjon på oppgavene. Det er ikke sikkert at deltagerne var fortrolig med den språklige formuleringen av den utvidede

tilbakemeldingen. Den lengre og relativt kompliserte formuleringen kan ha skapt forvirring, irritasjon eller en følelse av lært hjelpesløshet i stedet for å fungere som oppklarende instruks. Valentine (1985) påpeker at det er få gode replikasjoner av Wason's eksperiment med identiske betingelser og identisk formulert instruks. Flere studier gjort av blant annet Hughes (1966) viser til noe bedret prestasjon ved annerledes formulering av instruks sammenlignet med Wason. Evans og Lynch (1973) hevder også at deltagerne i høy grad er tilbøyelige til å velger stimuliene som er nevnt i instruks. Wason (1968) har også gjort noen funn ved forskjell i «skarpere» instruks (finn kortet som falsifiserer), men forskjellen skal i følge Valentine (1985) ved abstrakte oppgaver heller gi en liten økning i tendens til å verifisere enn å avkrefte. Likevel er disse studiene i stor grad foretatt på unge studenter, og gir vel ikke noe entydig svar på om formulering og akademisk språk kan være utslagsgivende på resultatet.

Wason (1968) sitt datasett fra seleksjonsoppgavene konkluderte med at flere deltagere trengte mange forsøk (rundt 15) på å akseptere at det var nødvendig å vurdere en annen responskombinasjon enn det de først antok, og velge et «kontrapositivt» alternativ. Verifiseringsstimulien (Q) kommer i «veien» og fungerer som en «felle» for å vurdere andre alternativer. Utgangspunktet for antall forsøk nødvendig å bruke er et minimum på seks forsøk, likevel bruker alle deltagerne langt flere forsøk enn det, spesielt ved første treningintervensjon. De første oppgavene krever flest forsøk, også blir det ferre forsøk mot slutten av treningsrunden. Gjennomsnittlig brukte deltagerne i gruppe 0 og 1 henholdsvis 23 og 15 forsøk på å fullføre første treningsrunde. Forskjellen er ikke regnet ut som signifikant, selv om den generelle tilbakemeldingen i snitt krevde flere forsøk i likhet med resultatet i Holth (2019). Resultatene målt ved antall riktige i gruppe 0 både ved baselinjerundene og kontrollrundene tilsier også at det har skjedd noe læring i begge treningsrundene da det fremdeles var en økning i antall rett på tross av nedgang i antall forsøk. Det er også en nedgang i antall forsøk behøvd for gruppe 1, nedgangen er ikke like markant som i gruppe 0.

Differansen på nedgangen er ansett som signifikant. Dette kan kanskje indikere at det er noe mer læringseffekt ved å bytte til forklarende tilbakemelding enn å bytte til generell tilbakemelding, eller det kan være at deltagerne i gruppe 1 allerede hadde fått maks utbytte av første intervensjon.

Datasettet til deltager 2 og 12 harmonerer dårlig med pilotstudie (presentert i appendiks ved tabell 5) som tilsier at det skal ta rundt 4 sekunder å lese tilbakemeldingen «riktig/galt» og trykke «neste». Det skal ta 15 sekunder å lese to rette forklarende tilbakemeldinger, inkludert trykke «neste». I tillegg må en kanskje også anta at det vil ta noen sekunder å forstå hva tilbakemeldingen innebærer. Sammenlignet med deltager 9, ble det enda tydeligere at deltagerne 2 og 12 med stor sannsynlighet ikke har lest tilbakemeldingen som er gitt gjennom begge treningsfasene. Dette kan ha vært en medvirkende faktor til lav prestasjon.

Wason`s eksperimenter har vært utsatt for kritikk. Klein (2019a) hevder blant annet at det er poenngløst å prøve å «omprogrammere» folk, og at det ikke er viktig med «*de-biasing*» verktøy. Han hevder at tendenser til bekreftelsesfellen forsvinner ved kontekst i oppgavene, at folk svarer annerledes eller feil på oppgaver de ikke kan relatere seg til. Det er ingen kontekstbaserte oppgaver i innledende baselinjeoppgaver i dette eksperimentet. Den første treningsoppgaven er kontekstbasert i forhold til at stimuliene er mindre abstrakte, men det er vanskelig å sammenligne de etterfølgende oppgavene med denne da de har vært utsatt for påvirkning. Tretten av fjorten deltagere har en identisk responsforekomst ved oppgave 6 (den siste baselinjeoppgaven) som ved oppgave 11 (den første treningsoppgaven). Dette gir en indikasjon på at det ikke eksisterer den helt store forskjellen på responsen på tross av mindre abstrakt stimuli kombinasjon, hvertfall i dette eksperimentet. Wason og Shapiro (1971) har undersøkt tilsvarende, og funnet at det er noen flere som svarer rett på tematiske oppgaver enn helt abstrakte oppgaver. Disse oppgavene bør da være noe helt konkret som deltageren

kan relatere seg til, formulert på en gjenkjennelig måte. Klein (2019a) hevder at oppgavene i Wason's eksperimenter er så enkle, at de aller fleste tar for lett på oppgaven og svarer det første som virker logisk. Eksempelvis vil det fremstå som sannsynlig at tallkombinasjonen som fullfører regelen; 2-4-6, er 8-10-12, eller at svaret på rett stimulikombinasjon ligger i instruksen. Klein (2019a) peker på at når deltagerne i Wason's studie får beskjed om at dette er feil, så vil over halvparten svare rett på neste forsøk da de skjønner at svaret ikke er det «enkleste» og første de kommer på. Dette studiet gir ikke støtte til Kleins funn på dette punktet. Alle deltagerne bruker tre eller flere forsøk på den første treningsoppgaven uavhengig av type tilbakemelding. Flere prøver den samme stimuluskombinasjonen de allerede har fått tilbakemelding på at er feil flere ganger, også lengre ut i treningen. Dersom Kleins teori er rett, og deltagerne kun trengte å «ikke ta så lett på oppgaven» for skjønne konseptet, vil en vel kunne forvente et lavere antall forsøk etter flere runder med tilbakemelding. Wason og Shapiro (1971) foreslår eksperimentets rammer, formulering av instruks eller oppgavens design som variabler som må undersøkes nærmere i forhold til om det kan påvirke deltagerens prestasjon. Wason og Shapiro (1971) har sett på effekten av at deltagerne har fått erfaring med den logiske strukturen bak problemløsningsoppgavene. Heller ikke dette skal ha hatt særlig stor effekt, kunnskapen var ikke tilstrekkelig til å skape forståelse i eksperimentsituasjon.

Ifølge en studie gjort av Kruger og Dunning (1999) tror over halvparten av befolkningen at de er bedre enn gjennomsnittet på en rekke ferdigheter. Med slik overdreven tro på egen kompetanse vil det vel ikke være unaturlig at en også overvurderer egen evne til kritisk tenkning, beslutningstaking og rasjonalitet. Dersom en har gjort seg opp en mening om noe, vil en kanskje sette høyere tillit til egen vurdering enn andres vurdering. Deltagerne har fått ved forklarende tilbakemelding en begrunnelse for hvorfor noe er rett og galt, med sikte på å gi deltagerne innsikt eller kompetanse til å løse problemet. Det var flere av

deltagerne som spurte om hvordan svaret på oppgavene egentlig var «sunn fornuft». De visste ikke at de ikke var kompetent til å overkomme egen tilbøyelighet for bekreftelsesfellen. Det er forsøkt å kontingensforme atferd ved hjelp av forsterkningsprosedyrer med positive og negative forsterkere. Tilbakemeldingen «riktig» har til hensikt å forsterke mer av den «nye» atferden, eller det kan også virke negativt forsterkende med fravær av tilbakemeldingen «feil». Tilbakemeldingen «feil» kan virke som positiv straff, og fravær av tilbakemeldingen «riktig» kan virke som negativ straff med hensikt om å straffe vekk mer av responsmønster som ikke er hensiktsmessig i disse oppgavene. Målet vil være å både ekstingvere feil responsmønster, men også etablere påstanden som ny diskriminativ stimulus for målatferd (rett løsning på problemoppgaven). Antall forsøk nødvendig per oppgave varierer mellom gruppene. Deltagerne som startet med generell tilbakemelding der det er forsøkt å etablere ny atferd ved forsterkningsprosedyre alene, bruker flere forsøk enn gruppe 1 der det i tillegg er benyttet regelstyring med råd (begrunnelse med positive konsekvenser) eller milde advarsler (begrunnelse med negativ konsekvens, altså svaret er feil/ikke relevant og det vil fortsette å være feil om det ikke endres på). En kan kanskje også tenke seg at generell tilbakemelding ved kun riktig/galt gir større rom for «egen» regelstyring der en selv tester ut ulike teorier. Flere deltagere gir inntrykk av å gi litt opp, og ved tilnærmet lært hjelpesløshet bare trykke vilkårlig fort og mange ganger helt til de får responsen «riktig». Noen ser ut til å ha lært et mønster, og de falt igjennom på kontroll-og generaliseringsoppgavene. Det kan virke som at noen har måttet gjennomgå så mange repetisjoner at de til slutt har skjont regelen utfra mengdetrening. I likhet med studien til Holth (2019) så har de som forholdt seg til regel og instruks fått mange rett på oppgavene med kun fire kort, men ferdighetene lot seg ikke alltid generalisere utover rammene satt i treningssituasjon og de mestret ikke kontrolloppgavene.

Tabell 4

«*Selection task*» presentert ved tre-term kontingens

Foranledning	Atferd	Konsekvens
P på den ene siden	Snu kortet	Hvis Q på den andre siden, så stemmer regelen Hvis \bar{Q} , så er regelen feil
Q på den ene siden	Snu kortet	Kortet er ikke relevant for regelen Hvis P, så kan regelen stemme Hvis \bar{P} så er det på siden av relevans
\bar{P} på den ene siden	Snu kortet	Ikke relevant, og på siden av det regelen handler om uavhengig av hva som står på andre siden
\bar{Q} på den ene siden	Snu kortet	Hvis P, så er regel feil Hvis \bar{P} , så er informasjonen irrelevant for regel

Note. Figuren viser en oversikt over valgfaterd ved tre-terms kontingensen

Wason og Shapiro (1971) foreslår at det er en mulighet for at deltagerne heller lærer å bli bedre på å unngå Q, enn å aktivt velge \bar{Q} . Dette støttes av flere datasett i eksperimentet der deltagerne i kontroll- og generaliseringsrundene ikke velger Q, men heller ikke alle kortene med stimuliene \bar{Q} . I tillegg kan valg av kort som må snus ha blitt kontingensformet, og ikke lenger betinges av en regel. Regelen kan ha blitt gjort om fra en tiltenkt diskriminativ stimulus for å snu kort hvis P og/eller \bar{Q} , til en stimulus delta for å klikke snu kortene ved Q og/eller \bar{P} . Mer forskning er nødvendig for å utrede virkningsfulle «*de-biasing*» teknikker. Det kunne vært interessant å bruke en praktisk manuell tilnærming mulig uten hjelp av teknologi/ «pad» for å utelukke dataferdigheter som påvirkende faktor, for så å presentere ett og ett konkret kort og la deltagerne få fysisk snu på kortet selv. Det vil også være nødvendig å undersøke formulering av instruks, og utlede om en enklere formulering og muntlig tilbakemelding ville hatt større effekt.

Referanser

- Angner, E. (2016). *A Course in Behavioral Economics*. Red Globe Press.
- Ask, K. & Granhag, P. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, 2, 43-63. <https://doi.org/10.1002/jip.19>
- Bjerknes, O. T. & Johansen, A. K. H. (2009). *Etterforskningsmetoder - en innføring*. Fagbokforlaget.
- Bratholm, S. & Eskeland, A. (2008). *Justismord og rettssikkerhet*. Universitetsforlaget.
- Catania, A. C. (2013). *Learning*. Sloan Publishing.
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W. & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Cook, M. & Smallman, H. (2008). Human Factors of the Confirmation Bias in Intelligence Analysis: Decision Support From Graphical Evidence Landscapes. *Human factors*, 50, 745-754. <https://doi.org/10.1518/001872008X354183>
- Cooper, J. O., Heron, T. E. & Heward, W. L. (2016). *Applied behavior analysis: Pearson new international edition (2. utg.)*. Pearson.
- Evans, J. S. & Lynch, J. S. (1973). Matching bias in the selection task. *British journal of psychology*, 64, 391-397.
- Fischhoff, B. (1982). Debiasing. I D. Kahneman, P. Slovic & A. Tversky (Red.), *Judgment Under Uncertainty: Heuristics and Biases* (s. 422-444). Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.032>
- Furreboe, E. & Sandaker, I. (2017). Contributions of Behavior Analysis to Behavioral Economics. *The Behavior Analyst*, 40, 1-13. <https://doi.org/10.1007/s40614-017-0110-0>

- Holth, E. (2019). *Confirmation bias: Prevalence and debiasing techniques* [Masteroppgave, OsloMet]. Open Digital Archive. <https://oda.oslomet.no/oda-xmlui/handle/10642/8012>
- Hughes, R. N. (1966). Stimulus-Change Effects of Forcing in Alternation Behavior. *Psychological Reports*, 19(2), 515-518. <https://doi.org/10.2466/pr0.1966.19.2.515>
- Ighoubah, F. (2021, 26. september). *Lille julaften 1997 fikk Sigurd Klomsæt oppdrag i en av de mest omtalte drapssakene i Norge*. TV2 Nettavisen. <https://www.nettavisen.no/birgitte-tengs/krim/drap/lille-julaften-1997-fikk-sigurd-klomsat-oppdrag-i-en-av-de-mest-omtalte-drapssakene-i-norge/f/5-95-300117>
- Johnson-Laird, P. N. & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1(2), 134-148. [https://doi.org/https://doi.org/10.1016/0010-0285\(70\)90009-5](https://doi.org/https://doi.org/10.1016/0010-0285(70)90009-5)
- Klein, G. (2019a, 5. mai). *The curious case of confirmation bias*. Psychology Today. <https://www.psychologytoday.com/us/blog/seeing-what-others-dont/201905/the-curious-case-confirmation-bias>
- Klein, G. (2019b, 11. juni). *Escaping from fixation*. Psychology Today. <https://www.psychologytoday.com/us/blog/seeing-what-others-dont/201906/escaping-fixation>
- Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980). Reasons For Confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118. <https://doi.org/10.1037/0278-7393.6.2.107>
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. <https://doi.org/10.1037/0022-3514.77.6.1121>

- Kvamme, O. D. (2021, 18. februar). *Jeg har skrevet om Baneheia-saken i 20 år. Nå har jeg svekket tro på rettsstaten*. M24. <https://m24.no/baneheia-baneheia-saken-kommentar/jeg-har-skrevet-om-baneheia-saken-i-20-ar-na-har-jeg-svekket-tro-pa-rettsstaten/318575>
- Milkman, K. L., Chugh, D. & Bazerman, M. H. (2009). How Can Decision Making Be Improved? *Perspectives on Psychological Science*, 4(4), 379-383. <https://doi.org/10.1111/j.1745-6924.2009.01142.x>
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H. & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 129-140. <https://doi.org/10.1177/2372732215600886>
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Rachlew, A. (2009). *Justisfeil ved politiets etterforskning* [Doktorgradsavhandling, Universitetet i Oslo]. DUO Vitenarkiv. <https://www.duo.uio.no/handle/10852/22587>
- Rotgans, J. I. & Schmidt, H. G. (2019). There is no shortcut to de-biasing biases. *Medical Education*, 53(11), 1064-1066. <https://doi.org/10.1111/medu.13958>
- Sommerset, T. & Isungset, O. (2020, 4. oktober). *Var Viggo i bua eller i Baneheia?* NRK. https://www.nrk.no/sorlandet/xl/var-viggo-i-bua-eller-i-baneheia_-1.15136771
- The Decision Lab. (2020). *Why do we favor our existing beliefs? The Confirmation Bias explained*. <https://thedecisionlab.com/biases/confirmation-bias/>
- Thurstone, L. L. (1924). *The Nature of Intelligence*. Routledge & Kegan Paul. <https://doi.org/10.4324/9781315010298>

- Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/doi:10.1126/science.185.4157.1124>
- Valentine, E. R. (1985). The effect of instructions on performance in the Wason selection task. *Current Psychological Research & Reviews*, 4(3), 214-223. <https://doi.org/10.1007/BF02686572>
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129-140. <https://doi.org/10.1080/17470216008416717>
- Wason, P. C. (1966). Reasoning. I B. Foss (Red.), *New Horizons in Psychology* (s. 135-151). Penguin Books.
- Wason, P. C. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273-281. <https://doi.org/10.1080/14640746808400161>
- Wason, P. C. (1969). Regression in reasoning? *British journal of psychology*, 60 4, 471-480.
- Wason, P. C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23(1), 63-71. <https://doi.org/10.1080/00335557143000068>

Appendiks**Tabell 5**

Rimelig tidsbruk på å lese instruks/tilbakemelding etter hver oppgave

Tid	Instruks	Feedback ved 2 feil, utvidet tilbakemelding	Feedback ved 2 rett, utvidet tilbakemelding	Feedback, «riktig», «galt»
Sekunder	33	23	15	4

Note. Tabellen viser resultatet av en pilotstudie med formål om å anslå omtrentlig tid det tar å lese instruks og tilbakemeldingen en får gjennom eksperimentet. Studiet er basert på gjennomsnittet av tidsbruken til to deltagere.

Etisk refleksjonsnotat

I eksperimentet oppgaven er bygget på var det nødvendig at reelle eksperimentdeltagere deltok ved oppmøte. Eksperimentet er derfor vurdert meldt inn til Regional komité for medisinsk-og helsefaglig forskningsetikk (REK), og Norsk senter for forskningsdata (NSD). Veileder tok en telefon til NSD og vurderte søknad på bakgrunn av prosjektskissen og ROS-analysen, og meldte tilbake at det ikke var nødvendig da det ikke skal samles inn navn eller persondata som kan knyttes til besvarelsene. Ved soningsanstalter og politihøyskolen som flere av deltagerne er fra, er det ikke mulig å bare dukke opp og rekruttere deltagere. Derfor er det nødvendig med en viss form for kontakt med rekruteringsformål. Jeg tok kontakt med koordinator for politihøyskolen som la ut en forespørsel om deltagelse i eksperiment på politihøyskolens studentside, og ba om å få beskjed dersom noen meldte seg. Koordinator misforsto, og la ved kontaktinformasjonen min i forespørselen. Deltagerne tok derfor direkte kontakt med meg via sin student-epost. Epostadressen ble dermed benyttet som middel for å avtale tid og sted, men studentene fikk informasjon om at epostadressen ville bli slettet etter gjennomføringen. Dersom de ønsket oppgaven tilsendt etter innlevering, ba jeg om samtykke til å beholde e-post-korrespondansen frem til det. Ved soningsanstalten var det daglig leder som koordinerte kontakten med deltagerne i forkant av oppmøte, det var ikke nødvendig med noen form for kontaktopplysninger fra resterende deltagere. Deltagerne oppga kjønn og omtrentlig alder innenfor en 5-års ramme, ingen andre personopplysninger ble samlet inn. Ingen opplysninger som kan identifisere person ble knyttet opp mot datamateriale. Deltagelsen baserte seg på en engangsgjennomføring, og det var ikke nødvendig å notere på noen måte hvilken deltager dataen tilhørte. Det ble heller ikke samlet inn noe biologisk materiale, tatt bilder, video, lyd eller lignende. Det var ikke behov for å innta medikamenter, eller gjøre inngrep på kropp. På grunnlag av dette ble det ikke vurdert nødvendig å søke godkjenning eller melde inn forsøket til REK eller NSD. Det er likevel foretatt en ROS-

Analyse, i forhold til fare for tap av data. Denne er sendt inn og godkjent av veileder (Per Holth, som videresendte denne til Torunn Lian) i forkant, jeg har ikke mottatt noe referansenummer.

Ved oppmøte fikk deltagerne opplest noen rettigheter i eksperimentsituasjon, som at deltagelse er frivillig og uforpliktende, og det er lov å trekke seg uten konsekvenser. De fikk opplysninger om at tidsbruken kunne variere, men opp mot en maksimaltid på 30 minutter. Deltagerne fikk oppgitt en kontaktperson de kunne henvende seg til ved videre spørsmål om personvern hensyn, og ble tilbudt en individuell gjennomgang av egne resultater etter eksperimentets slutt. Det ble også tilbudt en faglig gjennomgang av teoriene bak eksperimentet, og mulighet for å bli tilsendt egne data eller oppgaven i sin helhet etter 15. Juni, 2022.