

MASTEROPPGAVE
Læring i komplekse systemer
November 2018

Kognitiv psykologi vs. atferdsanalyse:
Valgatferd og belief-bias effekten

Miriam S. Østenheden



OsloMet – storbyuniversitetet

Fakultet for helsevitenskap
Institutt for atferdsvitenskap

Anerkjennelser

Tusen takk til min veileder gjennom denne prosessen, Gunnar Ree, for tålmodighet og kloke ord.

Tusen takk til Constance Nordenstam for å ha vært der for meg hele veien og for å ha hjulpet meg med å samle inn data. Jeg er veldig glad for at vi har fullført denne prosessen sammen.

Tusen takk til Sofia Alderete for god støtte gjennom hele studietiden og skriveprosessen.

Takk til Linn Syrrist, Trondheims mest lovende lektorstudent for hjelp til oversettelser og formuleringer når google translate har sviktet meg, og for uvurderlig mental støtte gjennom hele studietiden.

Takk til familien min som har støttet meg og vært tålmodige.

Innholdsfortegnelse

Oversikt over tabeller	i
Sammendrag.....	ii
Artikkel 1	1
Sammendrag	2
Abstract	3
1 Innledning.....	4
2 Rasjonalitet og deduksjon	6
2.1 Avgrenset rasjonalitet.....	6
2.2 Deduktiv resonnering	7
3 Kognitiv psykologi	8
3.1 Heuristikker og bias	9
3.2 Forventet nytteverdi-modellen	10
3.3 Prospektteori.....	12
4 Atferdsanalyse	13
4.1 Seleksjon ved konsekvenser	15
4.2 Verbalatferd.....	16
4.3 Regelstyrt og kontingensformet atferd	16
5 Valgatferd.....	19
Referanser.....	27
Artikkel 2	1
Sammendrag.....	2
Abstract	3
1 Innledning.....	4
2 Metode.....	14
2.1 Deltakere	14
2.2 Prosedyre.....	19
2.3 Mellom-observatørenighet	20
2.4 Registrering av data.....	20
2.5 Resultater.....	22
3 Diskusjon.....	24
Referanser.....	31

Oversikt over tabeller

Tabeller til artikkel 2:

Tabell 1.....33

Tabell 2.....34

Tabell 3.....35

Sammendrag

Kognitiv psykologi er opptatt av menneskelig atferd og de mener at det som skiller mennesker fra andre dyr er evnen til kognisjon. Kognisjon innebærer mentale prosesser og strukturer mennesker bruker for å tenke seg frem til ting og ta avgjørelser. De jobber innenfor et mekanistisk rammeverk hvor prosesser blir delt opp i mindre biter og undersøkt bit for bit, til man finner en årsak som kan forklare en observert atferd. Valgatferd har tradisjonelt sett blitt forklart gjennom en økonomisk modell av rasjonalitet, men nyere kognitive teorier tar høyde for at mennesker ikke kan rasjonalisere optimalt hele tiden, og at vi faktisk er begrenset i vår tilgang til informasjon. Atferdsanalysen har tradisjonelt sett lett etter årsak til atferd i miljøet, determinanter som påvirker individet utenifra. Private hendelser var lenge ekskludert som mulig årsak, men i nyere tid har det blitt inkludert i forklaringer av atferd. Valgatferd fra et atferdsanalytisk syn er avhengig av læringshistorie, forsterkningsbetingelser, og regelfølgning. Ved å sammenligne disse fagdisiplinenes syn på valgatferd blir det klart at atferdsanalysen og kognitiv psykologi kan gagnes av å gjøre tolkninger basert på vitenskapelige prinsipper. Gjennom artikkel 1 legges det et teoretisk grunnlag for atferdsanalytisk tolkning av kognitive fenomener. Heuristikker er mentale snarveier vi bruker når vi vil ta raske avgjørelser, og kognitive bias er de systematiske feilene som kan oppstå ved å bruke heuristikker for mye. Artikkel 2 tar for seg et tradisjonelt kognitivt eksperiment på belief-bias effekten og en atferdsanalytisk tolkning av hvorfor effekten oppstår.

Nøkkelord: kognitiv psykologi, atferdsanalyse, valgatferd, belief-bias

Running head: RASJONALITET OG VALGATFERD

Rasjonalitet og valgfærd

Miriam S. Østenheden

15. november 2018

OsloMet – storbyuniversitetet

Sammendrag

En av de største temaene for undersøkelse innenfor kognitiv psykologi er menneskelig resonnering. Mennesker resonnerer for å oppnå mål eller for å løse logiske problemer, men resonnering kan ikke alltid være optimalt. Avgrenset rasjonalitet erkjenner begrensningene vi har i tilgang til informasjon og at det er umulig for mennesker å ta ytelses-maksimerende valg hele tiden. Type 1-prosessering er det raske, automatiske systemet hvor vi behandler informasjon ubevisst, mens type 2-prosessering er det trege, analytiske systemet vi tar i bruk når type 1 ikke strekker til. Heuristikker er mentale snarveier, strukturer adaptert fra miljøet, som type 1 tar i bruk for å ta raske avgjørelser, ofte er avgjørelsene gode og hjelpsomme, andre ganger leder det til at vi gjør systematiske feilslutninger i form av kognitive bias. Valgatferd og ulike forklaringsmodeller som forventet nytteverdi-modellen og prospektteori, er ulike deskriptive teorier som forklarer valg fra et kognitivt ståsted. Innenfor atferdsanalytisk fagdisiplin vil det være hensiktsmessig å snakke om funksjonelle relasjoner mellom atferd og miljø, samt læringshistorie og forsterkning. En av de mest undersøkte kognitive biasene er belief-bias, når vurderingen av gyldigheten av et argument blir påvirket av troverdigheten av argumentet. Atferdsanalyse har de siste tiårene inkludert private hendelser i forklaringer av atferd, men hypotetiske konstrukter slik som belief-bias blir ikke sett på som gode forklaringer.

Nøkkelord: resonnering, avgrenset rasjonalitet, valgatferd, prospektteori, atferdsanalyse, regelstyrt atferd.

Abstract

One of the biggest topics for research in cognitive psychology is human reasoning. People use reasoning as a way to achieve goals or solve logical problems, but reasoning cannot always be perfect. Bounded rationality recognizes the limitations we have in accessing information and that it is impossible for humans to make performance-maximizing choices all the time. Type 1 processing is the fast, automated system where we process information unconsciously based on past experiences with similar information, while type 2 processing is the slow, analytical system we use when type 1 isn't good enough. Heuristics are mental shortcuts, structures adapted from the environment type 1 uses to make quick decisions. Decisions made from using heuristics are often good and helpful but, sometimes we make systematic failures in the form of cognitive biases. Different explanatory models for decision making behavior such as expected utility model, prospect theory, and different strategies for choice are descriptive theories that explain choices from a cognitive point of view. Within the behavior analytic discipline it would be useful to talk about functional relationships between behavior and the environment, as well as learning history and reinforcement. One of the most examined cognitive biases is the belief bias, when the assessment of the validity of an argument is influenced by the believability of the conclusion. Behavior analysis has for a while included private events in explanations of behavior, but hypothetical constructs such as belief bias are not seen as good explanations.

Keywords: reasoning, bounded rationality, choice behavior, prospect theory, behavior analysis, rule-governed behavior.

Rasjonalitet er et gjennomgående tema i kognitiv psykologi og hvordan man definerer det virker å være i kontinuerlig diskusjon. I tester av rasjonalitet blir det ofte testet for deduktiv resonneringsevne i oppgaver bestående av å sette sammen mindre biter av informasjon til en konklusjon, som for eksempel ved bruk av syllogismer. En syllogisme består av to premisser og en konklusjon, og i eksperimenter med bruk av syllogismer er det ofte deltakerens oppgave å bedømme om konklusjonen følger logisk av premissene (Stanovich, West, & Toplak, 2016). På daglig basis tar vi mange avgjørelser, og valgatferd er observerbar når et valg mellom alternativer blir tatt. Oppfattede eller potensielle konsekvenser kan komme umiddelbart eller utsatt i tid, de kan variere i hvor sannsynlig det er at de kommer, og valgatferd kan sies å være motivert av dem (Baum, 2005). I atferdsanalyse vil det være mest hensiktsmessig å snakke om regelstyrt atferd for å forstå hvordan mennesker tar valg. Regelstyrt atferd er atferd under kontroll av en verbal foranledning, og det er vår erfaring med tidligere regler og deres konsekvenser som avgjør om vi følger regler vi blir gitt eller ikke. Læringshistorie er en uhyre viktig faktor for å forstå et individs atferd, og for atferdsanalyse er det en uvurderlig determinant som må bli hensyntatt. De to vanligste typene regelfølgning, følgning og sporing, forklarer hvorfor og hvordan vi følger regler som blir gitt oss. Atferd må bli forklart i relasjon til miljøvariabler, til kontingensen atferden skjer i (Hayes & Ju, 1998).

Kognitiv psykologi har flere forklaringsmodeller hva gjelder hvordan mennesker tar valg, en økonomisk teori kalt forventet nytteverdi-modellen forklarer hvordan mennesker optimalt bør ta valg og den baserer seg på at mennesket kan ha full oversikt over all tilgjengelig informasjon og beregne seg frem til de mest hensiktsmessige valgene hele tiden. I motsetning til denne teorien står prospektteori foreslått av Kahneman og Tversky (1979) som tar utgangspunkt i at mennesker

er begrenset i sin evne til å være rasjonelle, som forklart av teorien om avgrenset rasjonalitet forelå av Herbert Simon (1983).

Når mennesker tar raske avgjørelser sier kognitiv psykologi at det skyldes bruk av heuristikker, også kalt tommelfingerregler. Heuristikker kan lede til gode og dårlige avgjørelser, og når vi gjør logiske feilslutninger kan kognitiv psykologi identifisere dem i form av ulike biaser/forutinntatttheter som forstyrrer resonneringsprosessen (Tversky & Kahneman, 1974). En av mange tester som blir brukt for å undersøke grad av rasjonalitet hos mennesker er tester i deduktivt resonnement. Syllogismer blir brukt som oppgaver i logikk og for å teste for belief-bias effekt. Belief-bias er en av de mest empirisk støttede biasene kognitiv psykologi taler for og den går ut på at i oppgaver hvor deltakere blir bedt om å bedømme om en konklusjon følger logisk fra premisser, vil troverdigheten av konklusjonen ha en så stor effekt på deltakeren at de gjør feilslutninger (Evans, Handley, & Harper, 2001).

Kognitiv psykologi har som tradisjon at strukturelle og mekanistiske forklaringer av atferd er gode nok så lenge de samsvarer med den virkelige verden. Atferdsanalyse på sin side må spesifisere funksjonelle relasjoner og se på atferd i konteksten den oppstår i (Dougher, 1995). Private hendelser kan inkluderes som en del av atferdsanalytisk forklaring, men ikke i form av hypotetiske konstrukter som kognitiv psykologi bruker. Private hendelser blir av atferdsanalysen sett på som atferd på lik linje som offentlige, men man må prøve å oppnå høyest mulig grad av kontroll og prediksjon, eller i det minste en fortolkning, for at det skal være hensiktsmessig å ha de med i forklaring (Hayes & Brownstein, 1986; Palmer, 1991).

I denne teoretiske artikkelen skal jeg ta for meg kognitiv psykologi og atferdsanalytisk syn på valgatferd. Artikkelen legger grunnlaget for en empirisk studie i hva angår en spesifikk kognitiv bias; belief-bias. Forklaringsmodeller fra kognitiv psykologi og atferdsanalyse vil bli

satt opp mot hverandre med mål om å skape en større forståelse for hva som skjer når vi gjør logiske feilslutninger.

Rasjonalitet og deduksjon

Rasjonalitet stammer fra økonomisk teori og er en type normativ teori. Normativ teori beskriver hvordan mennesker skal eller bør oppføre seg hele tiden for å maksimere nytteverdi, og den går ut i fra at vi hele tiden har tilgang til all informasjon relevant for oss. Å resonere betyr å trekke slutninger eller å på logisk vis tenke seg frem til noe (Store norske leksikon, 2012). Det skilles på to forestillinger av rasjonalitet; rasjonalitet₁ som beskrives som rasjonalisering for å oppnå mål, for eksempel; til hvilken tid man skal reise hjemmefra for at det skal klaffe best med t-bane og trikk til jobb, eller hvordan å spare inn penger på matbudsjettet en måned for å få råd til ny hvitevare. Rasjonalitet₂ er resonnering i tråd med normativ teori og logisk prosess (Evans, Over, & Manktelow, 1993). Når mennesker vi ser på som smarte og fremgangsrike tar dårlige avgjørelser blir vi først overrasket, og vi stiller oss spørsmål som; «hvorfor i all verden gjorde han det», «hva tenkte han på», «han visste jo bedre», men mennesker er først og fremst motivert av mål, og vi er villig til å gjøre mye rart for å oppnå målene våre – selv om det strider i mot logisk resonnement.

Avgrenset rasjonalitet

«Kun når vi forstår begrensningene [av rasjonalitet] kan vi utlede prosedyrer for å effektivt bruke kreftene menneskelige resonneringsevner gir oss» (Simon, 1983, s. 3, egen oversettelse). Simon (1983) var en av de tidligste kritikerne av et totalt rasjonelt og egoistisk menneske. Han introduserte begrepet *avgrenset rasjonalitet* som en motsetning til den økonomiske modellen av rasjonalitet, for å beskrive hvordan et menneske ikke kan ta totalt rasjonelle og ytelsesmaksimerende valg til enhver tid. Han søkte å forklare organismers, og

spesielt menneskelig, rasjonalitet ut i fra dets begrensede tilgang til informasjon og evne til å beregne seg frem til sannsynligheter. Begrepet avgrenset rasjonalitet referer til kognitive restriksjoner vi legger på oss selv for å kunne ta alle valg vi må gjøre i hverdagen (Simon, 1983). Rasjonaliteten vår er bokstavelig talt bundet av det faktum at vår tilgang til informasjon er bundet. For eksempel velger vi hvilket universitet vi vil søke oss inn på uten å undersøke alle universiteter i hele verden, eller ofte også landet. Eller så velger vi hvilken bil vi vil kjøpe uten å undersøke alle biler som er tilgjengelige for oss. En kan argumentere for at det å søke etter alle mulige alternativer tilgjengelig for oss hele tiden er en irrasjonell idé.

I Simons (1983) forklaring legges det vekt på de interne og de eksterne begrensningene som påvirker hvordan en organisme ser et rasjonelt problem. Han antyder at vi i behandling av informasjon, som på den ene siden består av søken etter og prosessering av informasjon, og på den andre siden valget av det beste alternativet, utvikler adaptive strategier for balanse og kompromiss mellom de to sidene. Simon, blant andre forskere, forsøkte å skape et bilde av en type «verktøykasse» med psykologiske adaptasjoner, strategier, og lite strevsomme perseptuelle-og hukommelsesprosesser mennesker kan bruke for å ta valg. Denne verktøykassen er en del av menneskets kognitive system og man velger strategi ut ifra hva slags problem man skal løse (Todd & Gigerenzer, 2000).

Deduktiv resonnering

I søken etter forståelse av menneskelig resonnering er tester av deduktiv resonnering en av de viktigste verktøyene kognitive psykologer har. Evans et al. (2001) skriver at deduktiv resonnering blir sett på som en av de ypperste prestasjonene innen rasjonell tankegang og den kan beskrives som prosessen hvor vi setter sammen biter av kunnskap, og trekker implisitte eller tidligere latente konklusjoner som passer med hva vi allerede tror. I motsetning til deduksjon står

induksjon som fungerer på motsatt måte, at vi gjør antagelser basert på spesifikk observasjon. Vi bruker deduktiv resonnering for å underbygge det vi allerede tror i form av å lage regler og generaliseringer, i stedet for å måtte huske en stor mengde overflødig informasjon. Disse reglene og generaliseringene vi lagrer etter å ha resonnert deduktivt kan vi trekke frem og bruke når det passer oss. Evans et al. (2001) er derfor ikke overrasket over mengden bevis kognitive psykologer har funnet på at vi trekker dårlige og uriktige slutninger i tester av deduktivt resonnement.

Innen kognitiv psykologi er en av de mest brukte metodene for å teste deduktiv resonnering bruken av syllogismer. En syllogisme består av to premisser og en konklusjon, konklusjonen er enten troverdig eller ikke, og gyldig eller ikke. Deltakeren skal bedømme om konklusjonen kan trekkes logisk fra premissene, om den er valid. Den type valg som testes for med syllogismer har en mål-struktur, og subjekter blir motivert til å finne den logisk korrekte responsen ved at de får instruksjoner om det eller fordi de ikke ønsker å fremstå som dumme (Evans et al., 1993).

Kognitiv psykologi

Innenfor kognitiv psykologi legges det til grunn at vi har to forskjellige måter å prosessere informasjon på. Teorien kalles for to-prosess teorien, og skiller på type 1- og type 2-prosessering. Type 1 beskrives som rask og automatisk, hvor vi bruker heuristikker, også kalt kognitive snarveier, for å behandle informasjon. Årsaken til at type 1-prosessering beskrives som automatisk er blant annet at behandlingen er rask og den ikke krever at vi bevisst søker etter svar. De stimuli som setter i gang type 1-prosessering gjør det fordi vi enten har behandlet dem før eller fordi de ligner på noe vi kjenner fra før, og det innebærer assosiative og implisitte læringsprosesser, samt behandling av følelser. Vi kan kjøre flere type 1-prosesser samtidig og

behandle flere tanker samtidig, uten at vi blir slitne av den grunn (Toplak, Stanovich, & West, 2012).

Sammenlignet med type 1-prosessering er type 2-prosessering langsommere og mer krevende. Den type behandling av informasjon krever at vi er bevisste og bruker stor del av hjernekapasiteten vår på å løse et problem. På grunn av hvor stor plass type 2-prosessering tar, blir den også begrenset av vår evne til å holde oss fokusert og bevisst over lengre tid. I motsetning til type 1-prosessering kan vi ikke behandle flere tanker om gangen, og det krever såpass mye bevissthet at vi heller må gjøre det psykologene kaller for serie-prosessering, behandle tanker etter hverandre. Når det kommer til hvor like eller forskjellige mennesker er eller tenker, finner vi som regel mange likheter i type 1 prosessering i motsetning til type 2-prosessering, der folk skiller seg fra hverandre i større grad. En av hovedfunksjonene til type 2-prosessering er å overstyre type 1 når vi skjønner at det ikke strekker til. Faren med for mye type 1-prosessering er at vi avgir responser som ikke er passende for situasjonen, som enten er feil eller irrasjonelle. Det er derfor store forskjeller mellom mennesker på når type 2-prosessering settes i gang hos individet (Toplak et al., 2012).

Heuristikker og bias

Heuristikker, eller tommelfingerregler, kan beskrives som enkle regler for hvordan vi skal tenke oss fram til ting når vi har begrenset med tid og/eller informasjon. Todd og Gigerenzer (2000) skriver at enkle heuristikker gjør det bra fordi de er adaptert fra strukturer i miljøet. Hvordan vi bruker heuristikker er individuelt fra person til person, og hva vi velger å bruke vil være et resultat av tidligere erfaring og lærdom, og hvilke signaler vi plukker opp fra miljøet rundt oss. Bruk av heuristikker er en måte for mennesker å takle det faktum at rasjonaliteten vår er bundet av vår begrensede tilgang til informasjon. To av de mest kjente heuristikkene er

representativitetsheuristikken og tilgjengelighetsheuristikken, som begge fungerer for å gi raskt svar når vi har begrenset med tid eller informasjon. Et eksempel på representativitetsheuristikken kan være når man sitter på kafe og observerer en person som ser sliten og medtatt ut, med tung sekk plassert ved siden av seg og stirrer tomt ut i luften. Da ville man antatt at personen er student på mastergradsnivå rett før oppgaven skal leveres. Denne heuristikken går ut på at jo mer X ligner på medlemmene i klasse Y, vil vi anta at X tilhører klasse Y.

Tilgjengelighetsheuristikken bruker vi når vi skal vurdere sannsynligheten for at en hendelse kan skje, og det avhenger av hvor lett vi kan se for oss situasjonen i hodet. For eksempel så vil vurderingen av hvor sannsynlig det er at man havner i en bilulykke, avhenge av om det har skjedd deg eller noen du kjenner før. Bruk av heuristikker leder som oftest til gode utfall, men det kan også lede til systematiske og forutsigbare feil i form av biaser (Tversky & Kahneman, 1974). Begrepet "bias" refererer til systematiske feil vi gjør når vi tar avgjørelser og estimerer sannsynligheter (Stanovich et al., 2016). I kognitiv psykologi snakker man ofte om rasjonalitet, resonnering, heuristikker og biaser i sammenheng med hverandre. I hvilken grad vi er rasjonelle når vi resonnerer og hvilke bias som virker ødeleggende for at vi tenker logisk. Kahneman og Tversky (1974) introduserte konseptet for å sette ord på og identifisere hva som skjedde når mennesker tok irrasjonelle valg eller gjorde logiske feilslutninger.

Forventet nytteverdi-modellen

Innenfor kognitiv psykologi er forventet nytteverdi-modellen i stor grad omtalt som et godt utgangspunkt for å forstå hvordan mennesker tar valg, og den er generelt akseptert som en normativ modell for rasjonelt valg (Kahneman & Tversky, 1979). Modellen stammer fra tradisjonell økonomi og går ut i fra at mennesker er rasjonelle i vurderingen av alternativer og at den subjektive nytteverdien er forskjellig fra person til person, og varierer fra situasjon til

situasjon, men at den maksimeres ved hvert valg (Simon, 1955). Et valg består av å velge mellom alternativer. Alternativer består av ulike handlinger, for eksempel å dra på fest eller å lese til eksamen, bli sittende oppe og se tv eller gå og legge seg tidlig. En del av det å ta valg er å tenke seg til og vurdere alternativer, mye i likhet med problemløsningsatferd. Derfor er slutningstaking og problemløsning to viktige temaer når man snakker om rasjonalitet. En annen viktig del av det å ta valg er vår tro om hva konsekvensene blir av valget vi tar. Vi vurderer alternativene tilgjengelig for oss og bestemmer oss for hva vi skal gjøre basert på hva vi tror utfallet av avgjørelsen blir og hvor sannsynlig det er at vi oppnår det vi vil. I valget mellom å dra på fest og lese til eksamen vil det spille en rolle i hvilken grad vi tror det å lese den kvelden vil forbedre sjansen for å gjøre det bra på eksamen. I tillegg kan det vurderes om man trenger en pause fra lesingen, eller hvor mye man har lyst til å møte vennene sine. Det blir en kost-nytteverdi-vurdering av alternativene, og vi velger det alternativet som er minst kostbart og mest nyttig.

I kognitiv forskning på atferd har forventet nytteverdi-modellen blitt brukt for å predikere responser i situasjoner hvor deltakere har blitt bedt om å spille om penger. Vanligvis besto alternativene av forskjellige pengesummer det kunne spilles om. Hva deltakerne trodde om sannsynligheter var basert på informasjon gitt av eksperimentator, og konsekvens var kontaktutbetalinger til deltakeren. Forventet nytteverdi-modellen kan ikke forklare hva som skjer når mennesker tar dårlige valg, når vi «burde ha visst bedre». Dersom vi får valget mellom en mulighet til å vinne 200 kr, eller en sikker 150 kr, velger vi det siste. Mennesker er tiltrukket av sikre konsekvenser, og modellen kan ikke forklare hvorfor. Selv om mennesket i teorien skal kunne beregne seg frem til hvilke valg som gir mest utbytte ut ifra hva som er viktig for en, så er det ikke alltid at vi klarer det. Vi er begrenset i vår kapasitet til å tenke nøye igjennom alt vi gjør hele tiden og alle mennesker er forskjellig, vi kan være rasjonelle, men kun til et visst punkt (Simon, 1983)

Prospektteori

At handlingene våre blir bestemt av mentale representasjoner av situasjoner, og ikke av situasjonene direkte, er en av de grunnleggende antakelsene i kognitiv psykologi. Hvordan vi oppfatter og tolker miljøet rundt oss er avgjørende for vår atferd. Vektleggingen av mentale representasjoner er en del av prospektteori, en av de mest kjente teoriene fra kognitiv psykologi, som forklarer hvordan mennesker tar valg under risiko og usikkerhet. Teorien ble utviklet av Kahneman og Tversky (1979) som en utfordring til forventet nytteverdi-modellen, som de mente ikke ga et realistisk bilde av menneskelig valg. Prospektteori tar utgangspunkt i at det er to faser i valgprosessen; redigering og evaluering. Redigeringsfasen består av at vi først forsøker å forstå de mulige alternativene/prospektene ved å ramme inn vilkårene for et valg, det kan forklares som at man forenkler og kombinerer noen mengder, og evaluerer mulige tap og gevinster, og da ofte ender opp med en enklere representasjon av prospektene. Evalueringsfasen består av at vi evaluerer de nye, enklere prospektene til vi finner den som har høyest verdi og velger denne. Redigeringsfasen består av flere operasjoner, koding; vi koder om mulige utfall av alternativene til tap eller vinning ut i fra et referansepunkt. Referansepunktet er ofte individets nåværende posisjon i forhold til valget, og behøver ikke å være konstant, det kan forandre seg akkurat som miljøet rundt oss er i kontinuerlig forandring. Dårligere utfall enn individets referansepunkt blir sett på som tap mens bedre utfall blir sett på som vinning. Videre operasjon i redigeringsfasen er: kombinasjon; vi kombinerer sannsynlighetene assosiert med ekvivalente utfall, dvs. utfall som betyr det samme, deretter følger segresjon; hvor man deler prospektenes risikokomponenter fra de som er risikofrie, og til slutt kansellering; å forkaste komponenter som er delt mellom de mulige prospektene (Kahneman & Tversky, 1979).

Atferdsanalyse

Atferdsanalyse er en vitenskap og behaviorisme er filosofien om den vitenskapen. John B. Watson regnes som grunnleggeren av behaviorisme og han skapte filosofien som et svar på komparativ psykologi. Han var uenig i psykologiens fokus på «sinnet» da han mente at verken introspeksjon eller analogier til dyrenes bevissthet kunne produsere pålitelige resultater på lik linje med metoder brukt av andre vitenskaper. Han argumenterte for at psykologi måtte fokusere på studier av observerbar atferd for at det skulle kunne bli en naturvitenskap. Kritikere av atferdsanalyse har poengtert at determinismen som er lagt til grunn i den behavioristiske filosofien ikke er forenelig med hvordan mennesker fungerer. De mener det blir feil å legge til grunn at all atferd er et resultat av genetisk disposisjon og miljøvariabler, og at fri vilje må regnes med som en faktor. Atferdsanalysen mener på sin side at den mystiske naturen til fri vilje er det som gjør at den ikke bør regnes med i forklaringer av atferd, og at det kun kan bidra til mer forvirring og misforståelser (Baum, 2005).

Moderne atferdsanalyse er basert på radikalbehaviorisme hvor det ønskes at vår naturlige erfaring av verden og atferd blir forklart med et begrepsapparat som er kjent for folk flest. Her skilles det ikke på en indre og en ytre verden, men vi ser heller at all atferd skjer i en verden. Fokuset ligger på å formulere begreper og konsepter gode nok for beskrivelse og videre forskning, og å oppnå en grad av prediksjon og kontroll slik at teorier kan bli testet eksperimentelt. Tidlige behaviorister hadde et litt annet verdenssyn, realisme hadde som fokus å beskrive den «sanne» verden. Hvis et objekt kunne observeres, eksisterte det, men det var et skille mellom den sanne verden og vår oppfattede verden. I motsetning til realisme står pragmatisme, forestillingen om at den virkelige nytten av vitenskapelig oppdagelse er hvilke

muligheter det gir oss videre. Dagens atferdsanalyse har et pragmatisk verdenssyn, og tror at vitenskap hjelper oss å forstå hva det er vi observerer og opplever med sansene våre, og at det fører til handling og problemløsning. Et viktig holdepunkt innenfor pragmatismen er at ingen kunnskap er helt sikker og den legger mye vekt på muligheten vitenskap har til å predikere fremtidige hendelser og kontrollere hva som skjer. For å finne ut om et spørsmål eller en idé er verdt å undersøkes må det vurderes om svaret på spørsmålet kan hjelpe fagfeltet og vitenskapen videre. Hvis å besvare spørsmålet ikke vil ha noen praktisk betydning, er det ikke verdt undersøkelsen i det hele tatt (Baum, 2005).

Lenge var atferdsanalysen kritisert for å ikke inkludere mentale hendelser i forklaringer av atferd. Skinner (1976) kom frem til en type dualistisk forklaring av atferd som skjer innenfor og utenfor huden. Han skilte på private og offentlige hendelser, hvorav det eneste skillet mellom de er hvem som kan observere atferden, andre eller kun en selv. Selv om vi snakker om atferd som kun en selv kan observere, er det atferd på lik linje med den atferden andre kan observere at en gjør – atferden er like naturlig. En «mentalistisk forklaring», som Skinner kalte det, innebærer at årsaken til atferd er noe inne i kroppen, som ikke er spesifikt nok til at det er målbart. En hendelse er privat når bare én person er vitne til den, for eksempel at du er alene hjemme og hunden din bjeffer, men hvis du ikke er alene hjemme når det skjer så blir hendelsen kategorisert som offentlig. Mentalistiske forklaringer refererer til sinnet, noe vi ikke kan observere på samme måte som tanker, det er noe hypotetisk, men som vi ofte refererer til allikevel. Baum (2005) skriver at minimumskravet for vitenskapelig studie burde være at objektet finnes i tid og rom, slik som for eksempel hjertet eller hjernen, mens «sinnet» er en fiktiv enhet, og siden det ikke kan observeres fungerer det ikke som forklaring. Gradvis har et skille mellom atferd formet direkte av dens konsekvenser og atferd under kontroll av regler blitt utviklet og skapt en link mellom atferdsanalytisk forskning og kognitive prosesser.

Seleksjon ved konsekvenser

Atferdanalyse diskuterer atferd på tre nivåer av seleksjon; naturlig, operant, og kulturell (Skinner, 1981). Naturlig seleksjon skjer på genetisk plan, hvor gener til de individene som er mer tilpasningsdyktige blir selektert til å reprodusere fremfor genene til de individene som er mindre tilpasningsdyktige. Eksempler på egenskaper til dyr som er selektert frem via naturlig seleksjon er pelsen til en hare som skifter farge avhengig av årstid, eller giraffens lange hals for å nå mat i høyder andre dyr ikke kan nå. Operant seleksjon er det andre nivået av seleksjon. Atferdanalyse har stort fokus på operant atferd, det vil si atferd som er selektert av sine konsekvenser. En konsekvens kan være forsterkende for atferden, det betyr at den vil øke sannsynligheten for at atferden forekommer igjen og med høyere frekvens, eller den kan være straffende, den vil minske sannsynligheten for at atferden forekommer igjen. Et enkelt eksempel på denne formen for seleksjon er å ta ta med seg paraply når det regner ute, eller å skru på lysbryter når det er mørkt. Atferd blir formet av omgivelsene og miljøet rundt oss, fra øyeblikket vi blir født vil atferden vår bli forsterket eller straffet, og atferdsrepertoaret vårt blir formet. I det siste nivået av seleksjon, kulturelt nivå, blir atferd selektert av de sosiale kontingensene i en kultur og forsterkere blir levert av det sosiale samfunn, eksempelvis tradisjoner. Kulturelle praksiser blir selektert når atferd blir videreført fra et individ til et annet (Skinner, 1981). Når atferdsanalyse skal forklare hvorfor atferd forekommer er læringshistorie en svært viktig faktor fordi den til varierende grad kan fortelle oss hva som har forsterket eller straffet en atferd tidligere, da snakker vi ofte om tidligere selektert atferd.

Verbalatferd

Fra et atferdsanalytisk perspektiv er menneskets evne til å kommunisere med hverandre på forskjellige måter det som gjør oss overlegne andre arter. Vi kan snakke med hverandre over lange avstander via telefon, vi kan snakke til mange mennesker samtidig som ved en tale eller over radio, og via skrevne ord kan språk bevares over tid. Tidlig forskning på atferd innenfor atferdsanalyse hadde som vane å undersøke effekten av direkte erfarte kontingenser og hvordan manipulasjon av kontingensene påvirket atferd. Det var først på 70- og 80-tallet at forskere ble interessert i effekten språk har på menneskelig atferd. Skinner (2014) var tidligere ute med å identifisere denne sammenhengen mellom språk og mennesket, han definerte verbal atferd som atferd forsterket gjennom en annen person, og han skilte på snakker og lytter. Skinners forståelse av verbale hendelser tok utgangspunkt i at det kun var nyttig å analysere atferden til snakkeren, siden lytterens atferd ikke nødvendigvis er verbal og det eneste lytteren gjør er å forsterke atferden til snakkeren. Hayes og Hayes (1989) skriver at atferden til lytteren er like kompleks som atferden til snakkeren, og kanskje til og med enklere å analysere. En eksperimentator kan lett arrangere noe for lytteren å lytte til, for så og se effekten i lytteatferd. Lytteren vil ikke bare være påvirket av det som blir lyttet til, men også sin egen læringshistorie, men allikevel vil mulighetene for kontroll og manipulasjon være større i lyttesituasjonen enn i snakkesituasjonen (Hayes & Hayes, 1989).

Regelstyrt og kontingensformet atferd

Kontingensformet atferd er atferd direkte styrt av sine konsekvenser, både forsterkerkende og straffende, for eksempel første gang du møter på en hund og du erfarer at det å løfte armene opp i været kan gjøre at hunden hopper opp på deg for å leke (straffende), eller når du begynner å

kjøre bil ute i trafikken og du må erfare å gjøre alle nødvendige bevegelser og handlinger flytende slik at kjøreturen blir så behagelig som mulig (forsterkende). Atferd kan også kontrolleres av sine foranledninger, det vil si foranledninger som signaliserer at en respons fører til en konsekvens. Mulige verbale foranledninger kan være regler eller instruksjoner, og disse har en effekt fordi de signaliserer at en konsekvens er tilgjengelig gitt en spesifikk respons. Atferden å kjøre bil, som beskrevet over, vil i starten være regelstyrt fordi du vil få instruksjoner om hva du skal gjøre; hvordan å starte bilen, gasse, bremse, gire, og se i speil. Atferden vil så gå over til å bli kontingensformet i det øyeblikket reglene ikke gir eksakt nok informasjon til å takle omgivelsene – som når du skifter fra å øve på en stor, åpen parkeringsplass til å kjøre i trafikken. Når vi følger en verbal regel, noe som forteller oss hva og når, og konsekvensen av å gjøre en handling, så kalles det regelfølgning. Atferd under kontroll av verbale foranledninger kalles regelstyrte og de er ikke like sensitive til konsekvensene som følger atferden som atferd formet i direkte kontakt med kontingenser (Catania, Shimoff, & Matthews, 1989). Kontingensene som opprettholder regelfølgning kan overstyre andre konsekvenser av den beskrevne atferden, ta for eksempel regelen; du må ta en blodprøve for å sjekke etter infeksjon så vi kan finne en passende behandling. Denne regelen vil for de fleste overstyre de mulige konsekvensene; smerten ved et nålstikk, ubehaget ved å dra til legen, ubehaget ved å spørre sjefen om fri fra jobb, osv.

Prinsippet om regelstyrt atferd har sine røtter i Skinners forklaring av verbal atferd fra boken *Verbal Behavior* (2014). Der snakker han for første gang om hvordan effekten av en instruksjon kan være annerledes enn å være i direkte kontakt med kontingensene. Han snakker om to typer operant atferd, atferd som er formet i direkte kontakt med kontingensene, og atferd som blir endret som et resultat av beskrivelse av kontingensene. I hverdagen bruker vi ofte instruksjoner eller regler fordi læring i direkte kontakt med kontingensene kan være farlig (f.eks. å kjøre bil eller å svømme), kontingensene er svake (f.eks. å be små barn om å vaske hender før

middag), eller når kontingenser konkurrerer med hverandre (f.eks. å prioritere å lese til eksamen versus å se på tv-serie). Om vi følger kommandoer, instruksjoner, og regler som blir gitt til oss vil fra et atferdsanalytisk synspunkt komme an på hva slags forsterkningshistorie vi har med det fra tidligere. Og om en atferd er kontingensformet eller regelstyrt kan vi ikke være sikre på dersom regelen samsvarer med kontingensen. Selv om en atferd er under kontroll av en foranledning, vil den med stor sannsynlighet bli påvirket av andre konsekvenser enn de signalisert av foranledningen på et eller annet tidspunkt fordi kontingenser og miljøet rundt oss stadig er i endring. Skinner (1976) skriver at en person som følger en verbal instruksjon aldri kan oppføre seg eksakt slik instruksjonen sier, fordi beskrivelsen av kontingensene aldri er fullstendig eller eksakt, ofte forklarer vi på enklest mulig måte slik at instruksjoner eller regler skal bli enklere å forstå for lytteren. Hvor sensitiv en atferd er til endring i kontingensene står i sammenheng med hvor sensitiv en regel er til endringer i kontingensene. Baum (2005) skriver at et sikkert tegn på at en atferd er kontingensformet er når du ikke kan forklare den til andre.

Det er to typer regelfølgning som forklarer hva som motiverer individet til å følge regelen, *pliance*; på norsk «følging» og *tracking*; på norsk «sporing». Når du følger en regel på grunn av konsekvensene av å følge eller ikke følge regelen, og ikke konsekvensene av atferden regelen beskriver i seg selv, er det *følging*. For eksempel hvis en lærer sier til sine elever «bruk blyant når dere svarer på prøven, ikke kulepenn», og elevene følger bruker blyant fordi de vet at læreren er i godt humør så lenge regelen blir fulgt, så er det *følging*. Hvis elevene hadde brukt blyant fordi de vil kunne erstatte svarene sine dersom de skriver noe feil, så er ikke det *følging*, for da følger elevene regelen på grunn av en mulig konsekvens av å ikke gjøre selve atferden regelen beskriver. Den andre typen regelfølgning, *sporing*, er regelfølgning under kontroll av i hvilken grad verden, eller de naturlige kontingensene, samsvarer med regelen som blir gitt. Et eksempel kan være å gi veibeskrivelse, hvis en lytter som blir gitt en veibeskrivelse følger beskrivelsen fordi

den korresponderer med den faktiske veien, er det sporing. Hvorvidt en lytter sporer, er avhengig av hva slags historie lytteren har med at naturlige kontingenser har korrespondert med tidligere beskrivelser/råd den har blitt gitt tidligere, og andre variabler som kan påvirke viktigheten av korrespondansen mellom beskrivelse og verden (Hayes & Ju, 1998).

Malott (1989) skriver om regler som motivasjonelle operasjoner. Hvis vi ser på en regel som en motivasjonell operasjon vil det si at regelen har en effekt på verdien av konsekvensene av å følge regelen, enten forsterkningsverdi eller aversiv verdi. Ta eksempelvis denne regelen «hvis ikke du begynner å jobbe med artikkelen snart vil du ikke bli ferdig med masteroppgaven». Denne uttalelsen kan fungere som en motivasjonell operasjon ved å øke den aversive verdien til stimuli assosiert med å ikke jobbe med artikkelen. Så vil det å starte å jobbe med artikkelen redusere den aversive verdien av andre aktiviteter, og heller forsterke jobbing.

Valgatferd

«En avgjørelse er et valg mellom muligheter» (Smith & Kosslyn, 2007, s.367, egen oversettelse). Vi snakker ofte om enkle eller vanskelige, og gode eller dårlige avgjørelser. Om en avgjørelse er noen av disse tingene avhenger av antall alternativer vi har å velge mellom og hvor sikre eller usikre vi er på utfallet av avgjørelsen vår. Vi snakker ofte om gode og dårlige avgjørelser, og at en god avgjørelse er når vi velger det alternativet som er mest sannsynlig å gi oss det vi vil oppnå, men i følge Evans et. al (1993) er det ikke så enkelt. Vi kan svært sjeldent være helt sikre på utfall i den virkelige verden, og om en avgjørelse er god eller dårlig må vurderes ut ifra hvor mye teoretisk støtte den har, og ikke av utfallet. Noen avgjørelser er enkle å ta fordi handlingen vi velger virker som det åpenbart riktige, men når ting ikke er åpenbart for oss blir det vanskeligere. For eksempel kan det å velge mellom å dra på fest eller lese til eksamen oppleves som en enkel avgjørelse når det er to dager til du skal ha eksamen, men to uker før eksamen blir det et vanskeligere valg. Den kognitive tradisjonens svar på hva som er gode og

dårlige valg er forventet nytteverdi-modellen som nevnt tidligere, hvor det er forventet at man tar valg som maksimerer nytteverdi for beslutningstakeren. Man kan på en måte si at menneskearten har utviklet seg til å bli jordas mektigste og mest kompliserte art ved å ta gode avgjørelser, noe som kan tyde på at rasjonell beslutningstaking er naturlig selektert til å være en del av oss. Fra et atferdsanalytisk perspektiv er hvorvidt en avgjørelse er bra eller dårlig ikke viktig, vi vil heller si noe om de funksjonelle relasjonene som sier noe om hvorfor vi gjør atferden. På den måten kan valg undersøkes ytterligere og målet er med høyest mulig grad av prediksjon og kontroll (Skinner, 1976). Slutningstaking ble tradisjonelt forklart med normativ teori, hvordan vi bør oppføre oss og ta avgjørelser, mens dagens forskning mer og mer dreier seg om deskriptive teorier – hvordan og hvorfor vi faktisk, i realiteten, tar avgjørelser.

Skinner (1976) mener at problemet med valg er ikke å velge mellom alternativer, men å unnsnippe ubeslutsomhet. Det er i utgangspunktet ikke vanskelig å finne alternativer til et problem, men når det er snakk om konsekvenser som er viktige for oss, og disse konsekvensene veier like mye, oppstår det et problem som vi må løse på en eller annen måte.

Atferdsanalyse foreslår ikke strategier slik det gjøres i kognitiv psykologi, men vi foreslår heller at sannsynligheten for at en person utfører en atferd finnes i deres læringshistorie. Skinner (1976) skriver at når vi står ovenfor et valg mellom flere alternativer, og konsekvensene av alternativene er av like stor verdi for oss, må vi inngå i problemløsningsatferd. En slik situasjon kan bli aversiv for individet, og enhver beslutningsatferd som får oss ut av situasjonen blir forsterket. Mens kognitiv psykologi legger forklaringer av atferd til strukturer i hjernen eller «sinnet», vil atferdsanalysen kun inkludere private hendelser som en del av en større forklaring som alltid vil søke ut mot miljøet individet er i relasjon med (Hayes & Brownstein, 1986).

Atferdsanalyse har som tradisjon et ønske om å oppnå en grad av prediksjon og kontroll av de fenomenene som blir studert, og det jobbes innenfor et relasjonelt eller kontekstualisert

rammeverk, det vil si at vi ser på atferd i kontekst, i relasjon til miljøet (Dougher, 1995). Atferd er den avhengige variabelen vi vil se endringer i avhengig av manipulerbare miljøvariabler (Hayes & Brownstein, 1986). Atferdsanalysen er opptatt av pragmatisme, hvilken nytte oppdagelsene vi gjør kan ha for oss, og om det kan føre til handling og problemløsning. Kognitiv psykologi jobber på den annen side innenfor et mekanistisk rammeverk, det vil si at forklaringer av fenomener er sanne i den grad de korresponderer med observerte hendelser, og beskrivelse og teoretisk prediksjon blir sett på som en god vitenskapelig forklaring. Atferdsanalysen valgte i mange år å se bort i fra mentale prosesser/hendelser som en del av forklaring av atferd, mye fordi det krevde at man kunne identifisere variabler i miljøet som hadde en effekt på disse interne hendelsene, og i beste tilfelle kunne man kun innhente vage beskrivelser av kognitive prosesser og historiske hendelser. Skinner skrev at mentalistiske forklaringer kan fungere like bra som miljømessige forklaringer når det kommer til prediksjon av atferd.

En av skillene mellom mekanistisk og kontekstualisert rammeverk for forskning er analyseenhetene de ulike tilnærmingene velger å legge vekt på. Mens mekanistisk psykologi legger vekt på hypotetiske konstrukter eller enheter som ikke er direkte observerbare, legger funksjonell atferdsanalyse vekt på relasjoner mellom atferd og kontekst. Atferd og kontekst kan ikke forstås isolert, det er kun sammen man kan finne funksjonelle relasjoner. Vi skiller på topografi og funksjon, det vil si at en stimulus som har en funksjon i en setting, trenger ikke ha samme funksjon i en annen, selv om den er topografisk lik i begge settinger (Dougher, 1995).

Et annet skille mellom de to rammeverkene er hvordan de forklarer og hvor de legger årsak til atferd. Det meste av psykologisk forskning går ut på å se på en pågående prosess, dele den opp i mindre biter, og legge årsaken til atferd i de enkelte bitene. Det er ikke vanlig å ta med individets læringshistorie som en del av årsaken, og atferd blir sett på som et resultat av et indre system. Atferdsanalyse på sin side ser på mønstre av atferd og hva som forsterker atferden

umiddelbart eller i løpet av et visst tidsrom (Dougher, 1995). I motsetning til kognitiv psykologi er læringshistorie en veldig viktig del av forklaringen innenfor atferdsanalyse fordi den sier noe om hvilke betingelser som har opprettholdt eller straffet en atferd historisk og umiddelbart.

Individer blir formet og endret av sine erfaringer, og forskjellige personer vil oppleve forskjellig effekt av kontingenser. Konteksten atferden opptrer i kan fortsette bakover eller fremover i tid, fra de fjerneste foranledninger til de mest utsatte forsterkere (Hayes & Brownstein, 1986).

Atferdsanalysen ser på funksjonelle relasjoner mellom atferd og miljø (kontekst), identifiserer determinanter, forklarer atferd ut ifra dens determinanter (eller gjør eksperimenter hvor kontroll og manipulasjon er viktig), og så tolker årsakssammenhengene og forsøker å finne ut i hvilken grad dette funnet kan brukes til å forklare mer av menneskelig atferd.

Det atferdsanalytiske synet på forklaring versus beskrivelse, er at en forklaring skal referere til prediksjon og kontroll av tilstrekkelig omfang og presisjon. En forklaring uten kontroll er det Skinner (1976) kalte for «tolkning» - det vil si at det kan gis en god beskrivelse av et fenomen, men at det ikke er mulig å oppnå kontroll. Kognitiv psykologi, eller det mekanistiske synet på forklaring – er strukturelle beskrivelser av atferd. Strukturelle beskrivelser av atferd kan så brukes til å undersøke hva som faktisk er årsak til atferden, men det er ikke det samme som kontekstuell forklaring – det vil ikke lede til kontroll.

Atferdsanalysen har lenge vært kritisert for å utelukke mentale hendelser som forklaring av atferd, men Skinner (1976) skriver om hvordan hendelser som å tenke eller å føle kan være en del av forklaringer av atferd, og hvordan vi kan skille på skjult og åpenbar atferd, for eksempel vil en forekomst av tenkning bli ansett som skjult atferd, dog like vitenskapelig tilgjengelig. Verdien av en observasjon blir bestemt av i hvilken grad den muliggjør prediksjon og kontroll (Hayes & Brownstein, 1986). Når en forklaring innebærer at en spesifikk tanke blir tenkt før atferden forekommer, må en atferdsanalytiker stille spørsmål ved kontingensene som fører til

relasjonen mellom tanken og atferden, og hva determinantene av den tanken er. Tanken, den skjulte atferden, blir sett på som atferd på lik linje med den åpenbare atferden den fører til, og den kan forklares i sammenheng med konteksten den skjer i, i motsetning til mekanistisk forklaring som kun ville fokusert på atferd-atferd relasjonen (Hayes & Brownstein, 1986). I tilfellene det ikke er mulig å oppnå kontroll må atferdsanalysen fortolke. Fortolkning må ikke forvirres med spekulasjon, spekulasjon har ingen begrensninger mens fortolkning må begrense seg til eksperimentell analyse. Det er ingen tvil om at fortolkning bidrar til forståelse av atferd, i følge Palmer (1991) er det noen ganger vanskelig å skille mellom fortolkning og eksperimentell analyse. Når fenomener er for store eller for vage til å ha kontroll over, er fortolkning et godt alternativ fordi det bygger på vel etablerte vitenskapelige prinsipper. Når private hendelser skal brukes som en årsak til atferd skal det ikke tas med som data eller brukes til å generere prinsipper, men det er under fortolkning at private hendelser blir aktuelt.

For å sette disse fagdisiplinene opp mot hverandre vil jeg bruke et eksempel på en valgsituasjon, for eksempel en students valg mellom å gå på fest eller lese til eksamen som ble nevnt tidligere. I dette eksempelet tar jeg også som utgangspunkt at eksamen er to uker frem i tid. Hvordan dette valget tas vil i følge forventet nytte-verdi modellen handle om hvilket valg som gir mest utbytte for den som velger. Velgeren er nødt til å gjøre en vurdering av de usikre utfallene av valgene og beregne seg frem til hvilket av alternativene som har mest nytteverdi. Modellen tar utgangspunkt i at mennesker er totalt rasjonelle, men mange kognitive psykologer har beveget seg bort fra denne antakelsen. Prospektteori som ble nevnt tidligere, foreslått av Kahneman og Tversky (1979) vil anta at en person som står mellom valgene å lese til eksamen eller dra på fest tar utgangspunkt i situasjonen personen står i når valget skal tas. I tillegg legges det til grunn at folk tar valg basert på de mulige tapene eller gevinstene av de ulike valgene, heller enn sluttresultatet. Atferdsanalytisk tilnærming til valgatferd er noe annerledes, når en person står

ovenfor et valg, enten det er å gå til høyre eller venstre, eller det handler om hvilken bil man vil kjøpe, så vil for det første læringshistorie ha en betydning. Vår læringshistorie med å ta valg vil enten øke eller minske sannsynligheten for å gjøre et valg, og det kommer an på om atferden vår har blitt forsterket eller straffet tidligere. Utgangspunktet for studenten er altså ikke situasjonen han eller hun er i der og da, men det er historien av forsterkning eller straff for å lese til eksamen og å dra på fest.

Forventet nytteverdi modellen er en normativ modell, den sier noe om hvilke valg som er de optimale, mens prospektteori er en deskriptiv modell. Modellen ble utviklet av Kahneman og Tversky (1979) fordi de mente den ville gi en mer psykologisk nøyaktig beskrivelse av valg i den virkelige verden, sammenlignet med forventet-nytteverdi teorien. Deskriptive modeller er en standard innenfor kognitiv psykologi og det passer med det mekanistiske rammeverket de jobber innenfor. I atferdsanalyse er ikke deskriptive modeller bra nok for å forklare atferd, der letes det etter årsakssammenhenger og funksjonelle relasjoner. Det er ikke nok å beskrive hvordan valg tas, vi er ute etter å kunne forklare relasjonen mellom atferden å ta valg og konteksten valget blir tatt i – og om dette er et mønster av atferd som kan studeres og danne grunnlaget for eksperimentell analyse.

Ut i fra prospektteori må studenten først finne et referansepunkt som utgangspunkt for videre valgprosess, det gjøres ved at de mulige utfallene av avgjørelsen blir rangert i henhold til en heuristikk, de blir redusert til forenklete verdier, og at studenten bestemmer hvilke utfall som er ekvivalente til hverandre, og derfra setter referansepunktet. Referansepunktet for valget kan enten være hvor bra studenten forventer å gjøre det på eksamen per hvor mye av fagstoffet han eller hun har akkumulert per dags dato, eller så kan et mulig referansepunkt være det verst tenkelige utfallet basert på forventingen om hvor mye studenten får lest dersom han eller hun drar på festen, for eksempel at studenten kun får lest 80 % av fagstoffet. Ut i fra referansepunktet

vurderer studenten dårligere utfall som tap og bedre utfall som gevinster. I evalueringsfasen blir de nye forenklete utfallene vurdert og den med høyest gevinst vinner frem. Når de ulike prospektene skal evalueres mot referansepunktet mener Kahneman og Tversky (1979) at tap er mer smertefullt enn gevinster føles bra, dette blir kalt tapsaversjon. I motsetning til forventet nytteverdi modellen hvor et individ er likegyldig til utgangspunktet de skal ta avgjørelsen fra og alt som betyr noe er den høyeste oppnåelige nytteverdien, mener Kahneman og Tversky at det er mer realistisk at mennesker vil ta hensyn til utgangspunktet deres før de tar en avgjørelse som involverer risiko og usikkerhet.

For å finne ut av hvilket valg studenten kommer til å ta må det vurderes nøyere om atferden å lese til eksamen har blitt forsterket nok tidligere. Dersom studenten har en historie av forsterkning av å lese til eksamen vil det være sannsynlig at han eller hun velger nettopp dette, men det vil også ha noe å si hvor viktig konsekvensen av å gjøre det bra på eksamen er for studenten akkurat nå. Det vil være samme tilfelle med atferden å dra på fest, hvis studenten har en læringshistorie av at det å dra på fest er bra (at det har gitt ønskede konsekvenser tidligere som f.eks. å ha det gøy med venner) vil valget avhenge av hvilken konsekvens som er mest attraktiv for studenten; å få en god karakter på eksamen eller å ha det gøy med venner. Hvor attraktiv konsekvensene er for studenten kommer an på hva tidsperspektivet på levering av de forskjellige konsekvensene, å dra på fest er en umiddelbar forsterker, mens å gjøre det bra på eksamen er en utsatt forsterker. Det er bevist gjentatte ganger i atferdsanalytiske eksperimenter at forsterkere som leveres umiddelbart har en stor effekt på atferd, mens utsatte konsekvenser ikke har like stor effekt. La oss ta som utgangspunkt at studenten har erfart at det å lese godt og sammenhengende til eksamen de siste to ukene før eksamensdato har ført til at studenten har fått en karakter A. Selv om karakteren er en utsatt forsterker kan den være av større verdi for studenten enn den umiddelbare forsterkeren å dra på fest. Kanskje mer sannsynlig er det at atferden å lese til

eksamen er under kontroll av negativ forsterkning – ønsket om å unnsnippe en aversiv stimulus som i dette tilfelle vil være å gjøre det dårlig på eksamen.

Atferdsanalyse og kognitiv psykologi har forskjellige måter å undersøke og forklare fenomener på. Lenge har de vitenskapelige disiplinene kritisert hverandre for å ha feil fokus i forskning, kognitiv psykologi har kritisert atferdsanalysen for å ha for stort fokus på kontekst eller miljøet som forklaring av atferd, og atferdsanalysen har på sin side kritisert kognitiv psykologi for å ha hovedfokus på mentale prosesser og hypotetiske konstrukter (Dougher, 1995). Det er lite atferdsanalytisk forskning på kognitive fenomener til tross for at det har blitt mer vanlig å inkludere private hendelser som en mulig årsak for atferd. Belief-bias effekten i logisk resonnement har en stødig og stor mengde forskning bak seg i den kognitive tradisjonen, men det foreligger ingen atferdsvitenskapelig fortolkning av fenomenet. Ved å gjøre slike fortolkninger kan de vitenskapelige disiplinene sammen øke forståelsen av menneskelig atferd.

Referanser

Baum, W. M. (2005). *Understanding behaviorism: behavior, culture, and evolution* (2 ed.).

Oxford: Blackwell publishing.

Catania, C. A., Shimoff, E., & Matthews, B. A. (1989). An experimental analysis of rule-

governed behavior. In S. C. Hayes (Ed.), *Rule-governed behavior : Cognition, contingencies, and instructional control* (s. 119-150). New York: Springer.

Dougher, M. J. (1995). A bigger picture: Cause and cognition in relation to differing scientific

frameworks. *Journal of Behavior Therapy & Experimental Psychiatry*, 26(3), 215-219.

Evans, J. S. B. T., Handley, S. J., & Harper, C. N. (2001). Necessity, possibility and belief: A

study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology A:*

Human Experimental Psychology, 54A(3), 935-958.

doi:<http://dx.doi.org/10.1080/02724980042000417>

Evans, J. S. B. T., Over, D., & Manktelow, K. (1993). Reasoning, decision making and

rationality. *Cognition*, 49, 165-187. doi:[http://dx.doi.org/10.1016/0010-](http://dx.doi.org/10.1016/0010-0277%2893%2990039-X)

[0277%2893%2990039-X](http://dx.doi.org/10.1016/0010-0277%2893%2990039-X)

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a

behavior-analytic view of the purposes of science. *The behavior analyst*, 9(2), 175-190.

Hayes, S. C., & Ju, W. (1998). Rule-governed behavior. In W. O'Donohue (Ed.), *Learning and*

behavior therapy (s. 374-391). Boston: Allyn & Bacon.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk.

Econometrica, 47(2), 263-292.

- Malott, R. W. (1989). The achievement of evasive goals In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (s. 269-322). New York: Springer.
- Palmer, D. C. (1991). A behavioral interpretation of memory. In J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (s. 261-279). Reno, NV: Context press.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69, 99-118.
- Simon, H. A. (1983). *Reason in human affairs*. California: Stanford University Press
- Skinner, B. F. (1976). *About behaviorism* New York City: Vintage books / Random house.
- Skinner, B. F. (1981). Selection by consequences. *science*, 213(4507), 501-504.
- Skinner, B. F. (2014). *Verbal behavior*. Cambridge, USA: BF Skinner Foundation.
- Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: Mind and brain* (Vol. 6): Pearson/Prentice Hall Upper Saddle River.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: toward a test of rational thinking*: MIT Press.
- Store norske leksikon. (2012). Resonnere. Retrieved from <https://snl.no/resonnere>
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and brain sciences*, 23(5), 727-741.
- Toplak, M. E., Stanovich, K. E., & West, R. F. (2012). Intelligence and rationality. In R. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence (3rd Edition)* (s. 784-826). Cambridge UK: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Running head: BELIEF-BIAS I SYLLOGISTISK RESONNERING

Belief-bias i syllogistisk resonnering og ulike forklaringsmodeller

Miriam S. Østenheden

15. november 2018

OsloMet – storbyuniversitetet

Sammendrag

Mennesker resonnerer for å predikere sannsynligheter eller for å ta avgjørelser, men vi kan ikke resonnerer optimalt hele tiden. Heuristikker, mentale snarveier vi bruker for å ta raske avgjørelser er hjelpsomme, men de kan også føre til systematiske feilslutninger, kalt kognitive bias. Belief-bias effekten er når vi lar vår kunnskap eller tro om verden forstyrre vurderingen av et arguments gyldighet. Mange studier av belief-bias har vist at effekten eksisterer og flere forklaringsmodeller er lagd for å forklare det, men det er enda ikke en atferdsanalytisk fortolkning av fenomenet der ute. En systematisk replikasjon av Markovits og Nantel (1989) sitt eksperiment med fire spørreskjemaer, syllogismer i evaluerings- og produksjonsparadigme, med positivt og nøytralt innhold ble konstruert. Data fra 56 deltakere fra OsloMet – storbyuniversitetet underbygger allerede eksisterende data om fenomenet. Resultatene bekrefter primærhypotesen om at belief-bias effekten er mest fremtredende når innholdet er positivt, men den avkrefter sekundærhypotesen om at belief-bias effekten er størst under produksjonsparadigme. Den atferdsanalytiske forklaringen legger vekt på læringshistorie, forsterkning, og regelfølgning som forklaring av effekten.

Nøkkelord: resonnering, valgførd, heuristikker, bias, belief bias, atferdsanalyse, regelfølgning

Abstract

People use reasoning to predict probabilities or to make decisions, but we cannot reason optimally all the time. Heuristics, mental shortcuts we use to make decisions are helpful, but they can also lead to systematic failures, called cognitive biases. Belief bias is when we let our knowledge or belief about facts of the world interfere with the assessment of an argument's validity. Many studies of belief bias have shown that the effect exists and several explanatory models have been made to explain it, but not yet has it been constructed a behavioral analytical interpretation of the phenomenon. A systematic replication of Markovits and Nantel's 1989 experiment with four questionnaires, syllogisms in evaluation and production paradigms, with positive and neutral content was constructed. Data from 56 participants at OsloMet – Oslo Metropolitan University supports existing data on the phenomenon. The results confirm the primary hypothesis that the belief bias effect is most prominent when content is positive, but it rejects the secondary hypothesis that the belief bias effect is greatest under the production paradigm. The behavior analytic interpretation emphasizes learning history, reinforcement contingencies, and rule-following as an explanation of the effect.

Keywords: reasoning, choice behavior, heuristics, bias, belief-bias, behavior analysis, rule-following

Mennesker resonnerer for å trekke slutninger eller for å finne løsninger på problemer (Stanovich, West, & Toplak, 2016). Rasjonalitet kommer fra økonomisk teori hvor det er forventet at mennesket kan vurdere alle mulige alternativer og beregne seg frem til den høyeste nytteverdien. Ideen om avgrenset rasjonalitet ble introdusert av Herbert Simon på 80-tallet og teorien forklarer hvordan mennesket er begrenset i sin evne til å være rasjonell hele tiden. Den økonomiske modellen av rasjonalitet tok utgangspunkt i at mennesker kan være totalt rasjonelle og kontinuerlig ta ytelses-maksimerende valg, men Simon (1983) var ikke enig i dette. Han mente at vår rasjonalitet er bundet av det faktum at vår tilgang til informasjon er begrenset, vi kan ikke vite alt om alt.

Kognitiv psykologi legger til grunn at vi bruker to prosesser for resonnering. Toplak, Stanovich, og West (2012) beskriver type 1 som det raske, automatiske systemet hvor vi trekker raske konklusjoner og hvor vi hele tiden registrerer og tolker miljøet rundt oss. Det er et komplekst system som mye av tiden fungerer ubevisst og ligger til grunn for intuitive responser og feilslutninger. Eksempler på hva som behandles av type 1-prosessering er å identifisere hvor lyden av en sirene kommer fra, finne svaret på 3×3 , tolke ansiktsuttrykk eller toneleie, forstå enkle setninger, eller gjenkjenne objekter i et bilde. Type 2-prosessering er det langsommere, analytiske systemet som skal ta over resonnering når type 1 ikke strekker til. Det som skiller mennesker fra hverandre i forskning på resonnering er når type 2 tar over problemløsningsprosessen. Eksempler på type 2-prosessering kan være å finne en person i et rom fullt av andre mennesker, å formulere en god setning om noe faglig, eller fastsette gyldigheten av et logisk argument.

Stanovich et al. (2016) skriver at en mengde forskning over en rekke tiår har etablert at menneskelig resonnering avviker fra den standard normative forestillingen om hvordan vi bør

løse resonneringsproblemer. Det er i anvendelsen av type 1-prosessering at vi har en tendens til å gjøre feilslutninger. Kahneman (2011) skriver at type 1-prosessering vedder på et svar når situasjonen er usikker, og at hva systemet vedder på avhenger av tidligere erfaring. Heuristiske prosesser, eller type 1 prosesser i resonnering, er raske, ubevisste prosesser som er tungt basert på kontekst. Bruk av heuristikker, også kalt tommelfingerregler, skjer når vi er nødt til eller ønsker å ta en avgjørelse raskt, enten fordi vi bevisst ikke vil eller kan bruke tid på det, eller ubevisst. Bruk av arbeidsminnet er begrenset i hvor mye informasjon det kan prosessere på en gang og heuristikker er en måte å få unna informasjon raskt (Tversky & Kahneman, 1974).

Vi tar et valg når vi velger et alternativ blant to eller flere alternativer, og et godt valg regnes ofte som (innenfor kognitiv psykologi) å være det alternativet som gir oss mest utbytte – mest av det vi vil ha. For å ta et valg må det foreligge en situasjon som må vurderes og en må forsøke å finne alternativer som kan løse situasjonen på best mulig måte. Det blir vanskelig for oss å ta en avgjørelse når vi er usikre på situasjonen og/eller alternativene, eller når alternativene virker å være like gode. Forventet-nytteverdi modellen fra kognitiv psykologi forklarer valg som en vurdering av alternativer opp mot hva som gir mest nytteverdi for individet. Modellen går ut i fra at alle mennesker er ulike, men at vi kan maksimere nytten av hvert valg vi tar ved å være totalt rasjonelle (Kahneman & Tversky, 1979). Simons (1955) teori om avgrenset rasjonalitet står i motsetning til denne modellen av valgførelse. Prospektteori, en annen kognitiv modell for forklaring av valgførelse, ble utviklet av Kahneman & Tversky (1979) som et alternativ til forventet nytteverdi-modellen. Teorien beskriver hvordan mennesker tar valg under risiko basert på potensielle tap eller vinnings, heller enn en mulig konsekvens senere, og valgene blir tatt ved bruk av heuristikker. Prospektteori og forventet nytteverdi-modellen motstrider hverandre hva gjelder referansepunktet, eller utgangspunktet for valget. I følge prospektteori vil smerten ved å tape noe vi allerede har bety mer for oss sammenlignes med hvordan å vinne noe føles godt,

Kahneman og Tversky (1979) kalte dette for tapsaversjon. I motsetning til forventet nytteverdi-modellen er vi ikke likegyldige til referansepunktet. Denne modellen er blitt populær for økonomer og et utgangspunkt for atferdsøkonomi, studiet av hvordan kognitive, emosjonelle, kulturelle, og sosiale faktorer påvirker økonomiske avgjørelser.

Atferdsanalyse er deterministisk og pragmatisk i søken etter forklaringer av atferd. Det innebærer at vi prioriterer forklaringer som kan bringe vitenskapen videre, og vi søker å oppnå en grad av prediksjon og kontroll over de fenomenene vi undersøker. Fokuset til atferdsanalysen er å formulere begreper og konsepter gode nok til at det kan bli testet eksperimentelt, og på den måten bygge et begrepsapparat som er tett knyttet sammen med den faktiske verden (Baum, 2005). En av atferdsanalysens prinsipper er at atferd blir selektert av sine konsekvenser. Vi undersøker atferd ved å spesifisere funksjonelle relasjoner mellom stimuli og responser, og responser og konsekvenser. Hvor viktig en konsekvens er for oss kan variere fra et tidspunkt til et annet, og innenfor atferdanalyse brukes motivasjonelle operasjoner som forklaring på atferd. For eksempel hvis et individ står ovenfor et valg om å være sosial eller å bli hjemme alene vil det å være deprivert for sosial kontakt være en etablerende operasjon for å søke kontakt med andre mennesker, og ha en evokerende effekt på atferden å være sosial. Sosial kontakt vil være av høy forsterkningsverdi for oss. Når sosial kontakt er oppnådd til et punkt hvor individet er fornøyd og behovet er dekket, vil sosial kontakt miste sin forsterkningsverdi og metning oppstår. Neste gang individet står ovenfor et valg mellom sosial kontakt og å være alene vil valget kanskje være annerledes. Vi søker alltid utenfor individet for å forklare atferden, til miljøet og kontingensene individet er i. Atferdsanalysen har lenge blitt kritisert for å ikke inkludere mentalistiske hendelser som en del av sine forklaringer. B. F. Skinner (1976) belyste problemet, og selv om han ikke trodde på at mennesker har fri vilje så anerkjente han at forklaringer av atferd også kan komme

fra «under huden», ellers kalt private hendelser. Å snakke om valgfaterd fra et atferdsanalytisk ståsted, innebærer å snakke om læringshistorie, forsterkningsbetingelser, og regelstyring.

Regler er noe som forteller oss hva og når vi skal gjøre noe, og hva som skjer når vi gjør det. Regelfølgning er når vi gjør det regelen forteller oss å gjøre (Hayes & Hayes, 1989). En forutsetning for regelstyrt atferd er verbal atferd, evnen til å kommunisere med andre mennesker gjennom verbale handlinger. Skinner (2014) forklarte verbalatferd i tråd med sin teori om operant betingning og sa at kjennetegnet ved verbalatferd var at konsekvensen av atferden ble mediert gjennom en annen person. Han skilte på snakker og lytter, og til tross for at han mente at det var snakkernes atferd som var viktig å analysere, har andre i etterkant også lagt fokus på lytteren (Hayes & Hayes, 1989). Kontingensformet atferd, atferd direkte styrt av sine konsekvenser, forekommer når vi ikke har noe erfaring med hvordan vi skal oppføre oss i en gitt situasjon, for eksempel første gang du prøver å sykle. Regelstyrt atferd er atferd under kontroll av foranledende stimuli som regler eller instruksjoner, for eksempel første gang du skal svømme uten å bli holdt av en lærer, så vil du følge instruksjonene om å puste jevnt, bevege armene og beina i takt og med spesifikke bevegelser, eller hvis du får en veibeskrivelse til et sted du ikke har vært før. Det er to typer regelstyring vi ofte snakker om; (1) følgning, når vi følger en regel på grunn av konsekvensene av å følge eller ikke følge regelen, og ikke konsekvensene av atferden regelen beskriver, og (2) sporing, når vi følger regelen til den grad den samsvarer med de naturlige kontingensene den beskriver (Hayes & Ju, 1998). Selv om en atferd i utgangspunktet er regelstyrt kan den etterhvert bli kontingensformet da miljøet stadig er i forandring. I tillegg skrev B. F. Skinner (1976) at en person som følger en verbal instruksjon eller regel aldri kan oppføre seg eksakt slik som regelen beskriver fordi vi ofte forenkler beskrivelser for å gjøre de mer forståelig for den som lytter.

Mens atferdsanalysen ser på atferd innenfor et kontekstualisert rammeverk, atferd i konteksten den oppstår i og de funksjonelle relasjonene, har kognitiv psykologi som utgangspunkt et mekanistisk rammeverk, det vil si at så lenge forklaringen av atferd korresponderer med det som observeres, er jobben gjort (Dougher, 1995). Mye av forskningen innenfor psykologi går ut på å ta prosesser i fra hverandre og undersøke bitene for å finne årsaken til atferd. Ofte stammer årsaken fra indre prosesser, og læringshistorie blir ikke sett på som en viktig del av forklaringen. Atferdsanalytisk forskning ser derimot etter mønstre av atferd over tid og hvilke forsterkningsbetingelser som opprettholder atferden, læringshistorie blir derfor en viktig del av forklaringen. Der hvor kognitiv psykologi ville forklart årsaken til atferd med tanker, vil atferdsanalyse se på tanker som atferd på lik linje med annen atferd, og forsøke å forklare den ut ifra konteksten atferden skjer i. Selv om atferdsanalysen søker å oppnå prediksjon og kontroll over fenomener som studeres, er det ikke alltid det er mulig. Når fenomener er for vage eller komplekse til å studeres er fortolkning et godt alternativ som kan bidra til vitenskapen, nettopp fordi fortolkning skal bygge på etablerte vitenskapelige prinsipper (Palmer, 1991).

Heuristikker forklarer hvorfor vi gjør feilslutninger, mens bias beskriver de systematiske feilene vi gjør i slutningstaking og sannsynlighetsestimering. Tversky og Kahneman (1974) introduserte biaser som en forklaring på når menneskelig resonnering avviker fra slik det forventes. Når vi tar til oss informasjon om miljøet rundt oss og behandler det med type 1-prosessering blir det gjort en rask analyse basert på tidligere erfaringer. Hvordan vi tolker omgivelsene våre er altså basert på hva vi har sett og lært før, altså læringshistorie. Biaser, forvrengninger av virkeligheten, fører til avgjørelser som ikke er logisk rasjonelle, men som er verdifulle når tiden vi bruker på å ta avgjørelsen er viktigere enn nøyaktigheten av den. Bruken av bias kan også være et produkt av begrensningene mennesker har i prosessering av informasjon, relatert til teorien om bundet rasjonalitet.

Belief-bias kan forklares som den logiske feilslutningen mennesker gjør når vi lar vår tro om hvordan verden fungerer overstyre vurderingen av et logisk argument; vi bedømmer argumenter som gyldige hvis de er troverdige for oss og avviser argumenter som ikke er troverdige. Begrepet belief-bias blir brukt gjennom hele artikkelen da det ikke er noen god oversettelse på norsk. Det er bevist at denne effekten eksisterer (Evans, Barston, & Pollard, 1983; Markovits & Bouffard-Bouchard, 1992; Markovits & Nantel, 1989; Newstead, Pollard, Evans, & Allen, 1992; J. Oakhill, Johnson-Laird, & Garnham, 1989; J. V. Oakhill & Johnson-Laird, 1985), og det var M. C. Wilkins som i 1928 først rapporterte om den. Det var dog først på 80-tallet at videre forskning på fenomenet virkelig ble satt i fokus av Evans et al. (1983) da de beviste med god eksperimentell kontroll at deltakernes vurdering av konklusjonene ble påvirket av logisk gyldighet og troverdighet (Markovits & Nantel, 1989).

Testing av belief bias har som tradisjon å bruke syllogismer. En syllogisme er en form for deduktivt argument, det vil si at en konklusjon trekkes ut ifra premisser. En syllogisme består av stort begrep (major term), et mindre begrep (minor term), og et middels begrep (middle term) – i denne rekkefølgen, og den sier noe om relasjonene mellom begrepene. Det er fire typer uttalelser syllogismer blir satt sammen av: (1) den universelle bekreftende: Alle X er Y, (2) den spesifikke bekreftende: noen X er Y, (3) den universelle negative: ingen X er Y, og (4) den spesifikke negative: noen X er ikke Y (Evans et al., 1983). Det som gjør at tester med syllogismer blir gjort om fra en test av resonnement til en test av belief bias er at forskere legger til en belief bias komponent. I eksempler hvor syllogismen er gyldig legger de til en konklusjon som ikke er troverdig, og med syllogismer som ikke er gyldige legger de til konklusjoner som er troverdige (Stanovich et al., 2016). Det som gjør eksperimentet effektivt i å påvise belief bias er at det utfordrer deltakerens kunnskap om verden på en slik måte som kan evokere feilslutninger, til tross for klare instruksjoner som skulle motvirket det. Eksperimenter på belief bias effekten med

bruk av syllogismer har blitt gjort på forskjellige måter; eksperimenter hvor deltakere blir bedt om å evaluere gitte konklusjoner (Evans et al., 1983) og eksperimenter hvor deltakere blir bedt om å trekke egne konklusjoner (J. Oakhill et al., 1989). Stanovich et al. (2016) skriver i sin søken om en omfattende test av rasjonalitet at i tester av belief-bias må det være en konflikt mellom naturlige, intuitive responser og mer gjennomtenkte, normative responser. Syllogismer med en belief-bias komponent utfordrer deltakeren til å holde fokus på resonnement gjennom et problem, uten å bli distraheret av den naturlige tendensen mennesker har til å gjøre det som er enklest – å bruke kognitive snarveier. Et klassisk eksempel på en syllogisme er;

Premiss 1: Alle blomster har kronblader.

Premiss 2: Roser har kronblader.

Konklusjon: Roser er blomster.

Denne syllogismen er troverdig, fordi alle vet at roser er blomster, men hvis man analyserer premissene vil man kunne se at konklusjonen ikke er gyldig. Premiss 1 sier at alle blomster har kronblader, ikke at alt som har kronblader er blomster. Det er her konflikten mellom vår kunnskap om verden og en nøye vurdering av det som står skrevet kommer i spill, de fleste ville svart at syllogismen er gyldig – uten å faktisk vurdere premissene. Stanovich et al. (2016) skriver at mennesker har en naturlig tendens til å være kognitivt gjerrige – at vi velger måter å tenke på som er lettere og mindre krevende, på bekostning av en mer anstrengende vurdering eller gjenfinning av fakta. Vi kan velge når vi vil bruke kognitive snarveier og ikke basert på hvilke avgjørelser som er viktige for oss. For eksempel vil vi ikke være kognitivt gjerrige når vi skal vurdere om vi skal kjøpe hus eller bil, eller velge fritidsaktivitet til barna, men gjerrigheten vil vise seg mer i hverdagslige problemer som hva å lage til middag eller hvilken film man skal se den kvelden.

Det er en lang tradisjon med forskning på belief-bias effekten innenfor kognitiv psykologi. Evans et al. (1983) gjorde fagfeltet oppmerksom på denne biasen da de pekte ut tre svakheter ved tidligere forskning som de kontrollerte for i sine eksperimenter; (1) konverteringseffekt, dvs. at når deltakerne ble vist «alle A er B» gikk de ut i fra at også «alle B er A», (2) atmosfærebias, dvs. at deltakerne viste en tendens til å velge konklusjoner som delte samme setningsoppbygging som premissene uten hensyn til den logiske gyldigheten, og (3) figurbias, dvs. tendensen til at rekkefølgen informasjonen i syllogismene blir presentert påvirker deltakeren. Resultatene deres viste tre klare tendenser; (1) deltakerne gikk god for flere gyldige enn ugyldige konklusjoner, (2) deltakerne gikk god for flere troverdige enn ikke-troverdige konklusjoner, og (3) effekten av belief bias viste seg større på ugyldige konklusjoner enn på gyldige. Evans et al. skriver at funnene deres underbygger to-prosess teorien om resonnering og at det virker å være en konkurranse mellom logiske og ikke-logiske tendenser. Høyest grad av belief-bias effekt og lavest grad av logikk ble funnet hos de deltakerne som fokuserte mest på konklusjonen og refererte til irrelevant informasjon under avgjørelsesprosessen. Evans et al. kalte denne effekten for respons bias, og det er en av de mest populære forklaringene av hva som skjer når man observerer belief-bias effekten. Denne forklaringen av belief-bias har i senere tid blitt navngitt selektiv undersøkelsesmodellen og dens grunnleggende antakelse er at deltakerne tar konklusjonen i syllogismen som utgangspunkt og aksepterer den hvis den er troverdig, uten å vurdere om den følger logisk fra premissene.

Oakhill og Johnson-Laird (1985) fant også bevis for belief-bias effekten, men i motsetning til Evans et al. (1983) hvor deltakerne var nødt til å evaluere gitte konklusjoner, valgte Oakhill og Johnson-Laird å be deltakerne sine produsere konklusjonene selv. I de tilfelle premissene ledet til en gyldig konklusjon var det større sjans for at de produserte den selv hvis den var troverdig enn ikke troverdig. I tillegg valgte de å lage konklusjoner som enten var usanne

i følge definisjon, det vil si at de brøt med definisjoner av begreper, eller empirisk usanne, at de brøt med kjente fakta. De premissene som ikke kunne lede til gyldige konklusjoner førte til at deltakerne svarte «ingen gyldig konklusjon» hvis premissene la grunnlag for en usann i følge definisjon-konklusjon. Oakhill og Johnson-Laird brukte resultatene sine til å støtte opp under mentale-modeller teorien som forklaring av belief-bias effekten. Modellen tar utgangspunkt i at tro kan påvirke resonnering på tre måter; under tolkning av premissene, ved å begrense søket etter alternative modeller, og ved vurdering av konklusjoner som genereres. Det som skjer under resonnering, i følge teorien, er at individet lager en mental representasjon ut i fra premissene og dermed sjekker om konklusjonen går overens med modellen. Konstruering av modeller kan sammenlignes med hvordan en arkitekt tegner en bygning som passer med omgivelsene eller hvordan diagrammer settes opp i fysikk, modeller blir konstruert ut ifra situasjonen de skal representere. Dersom konklusjonen passer inn i modellen går individet videre til å vurdere troverdigheten av konklusjonen, men blir ikke konklusjonen vurdert troverdig starter en videre prosess som innebærer å søke etter alternative modeller for premissene. Syllogismer kan være en-modell syllogisme (hvor bare en modell er mulig å konstruere) eller multipl-modell (det er mulig å konstruere flere modeller). Konklusjoner som passer med de alternative mentale modellene blir akseptert, ellers blir de avvist.

Newstead et al. (1992) undersøkte effekten av belief-bias for å sette to forklaringsmodeller opp mot hverandre; selektiv undersøkelsesmodell (som nevnt over) og misforstått nødvendighetsmodell. Den sist nevnte modellen har som utgangspunkt at deltakerne misforstår hva som er ment med logisk nødvendighet. I introduksjonen, før syllogismene blir presentert, blir deltakeren instruert i å evaluere (eller produsere) om en konklusjon nødvendigvis følger fra premissene. Ordet «nødvendigvis» er ment å signalisere til deltakeren at en konklusjon må evalueres gyldig (eller produseres) bare hvis den følger fra premissene. Selv om det bare er en

gyldig konklusjon som kan trekkes fra premissene, er det noen ganger det kan virke som at det er flere mulige konklusjoner. Denne forklaringsmodellen tror at når deltakere gjør en logisk analyse av syllogismene og de ikke intuitivt vet om en konklusjon er gyldig eller ikke, vil de basere avgjørelsen (eller produksjonen) på deres kunnskap om verden. Modellen predikerer at ved ubestemte, ugyldige syllogismer så vil deltakerne raskt innse at ingen konklusjon kan følge fra premissene. Dog vil det være vanskelig for dem å svare «ingen gyldig konklusjon» på grunn av misforståelse av ordet «nødvendig», og de vil heller svare i henhold til tidligere erfaring med å måtte avgi et svar. Fra dette følger prediksjonen om at hvis alle ugyldige syllogismer som blir brukt er bestemt falske (dvs. leder til konklusjoner som definitivt er feil) så vil interaksjonen mellom logikk og tro forsvinne. Resultatene til Newstead et al. støttet opp mot misforstått nødvendighetsmodellen da prediksjonene modellen gjør passet med resultatene deres, men heller syntes de at mentale modeller-teorien kunne forklare belief-bias effekten bedre. Newstead et al. foreslår at årsaken til belief-bias effekten er at deltakerne ikke klarer å konstruere alternative modeller når den første modellen ikke passer med konklusjonen.

Markovits og Nantel (1989) baserte sin studie på Evans et al. (1983) og Oakhill og Johnson-Laird (1985) for å sammenligne belief-bias effekten fra et evalueringsparadigme og produksjonsparadigme. Det foreliggende eksperimentet er en replikasjon av Markovits og Nantel sitt eksperiment. Replikasjonen er en systematisk replikasjon fordi materiale ble oversatt til norsk da det var norske studenter som skal testes. Kazdin (2011) definerer systematisk replikasjon som en repetisjon av eksperiment ved å systematisk la variabler variere (s. 375, egen oversettelse). Replikasjon er en viktig del av vitenskapelig forskning av flere grunner; det gjør at empiriske funn kan underbygges, styrkes, og generaliseres. Kazdin nevner at systematiske replikasjoner kan bidra til å vise at funnene er gjeldende over et bredere spekter av forhold.

Resultatene Markovits og Nantel (1989) fikk viste at belief-bias effekten generelt var av høyere grad i produksjonsparadigme enn i evalueringsparadigme. Metoden valgt av Markovits og Nantel var å lage fire par av kategoriske syllogismer hvor alle åtte syllogismene startet med det universelt bekreftende premisset «alle A er B». Det andre premisset kunne ha en av fire følgende former: (1) «X er A», (2) «X er B», (3) «X er ikke A», og (4) «X er ikke B». X varierte mellom å være et nøytralt begrep, et oppfunnet ord, som deltakerne ikke hadde noen tidligere erfaring med eller klar tro om, og et positivt begrep, et ord deltakerne kunne fra før og som de ville ha en eller annen formening om i forhold den empiriske sannheten av argumentet. Primærhypotesen i det foreliggende eksperimentet er at belief-bias effekten vil være av høyest grad i de syllogismene som er av positivt innhold (med begreper alle har kjennskap til) enn de av nøytralt innhold (oppfunnede begreper). Sekundærhypotesen er at belief-bias effekten vil være mer fremtrende i produksjonsparadigme enn i evalueringsparadigme, i likhet med resultatene til Markovits og Nantel.

Belief bias har blitt forsket mye på innenfor kognitiv psykologi, det er bevist at fenomenet eksisterer, men det finnes enda ikke en atferdsanalytisk beskrivelse av fenomenet. Formålet med denne empiriske studien er å replikere et tradisjonelt kognitivistisk eksperiment av belief-bias effekten og gjøre en atferdanalytisk tolkning av den.

Metode

Deltakere

Totalt 56 studenter var med i studien og ble rekruttert i klasserommet de satt i etter endt forelesning, og etter avtale med dagens foreleser. Studentene ble rekruttert fra tre forskjellige klasser, i forskjellige studieretninger. Studentene fikk instruksjoner muntlig i plenum, samt i spørreskjemaene, og tilbud om å notere seg ned spørreskjemanummer for å kunne følge opp

resultater eller trekke seg fra studien senere. På denne måten ble studentene anonymisert og ingen liste over navn på deltakerne eksisterer.

Materiale

Fire spørreskjemaer ble konstruert etter metoden til Markovits og Nantel (1989). Første side av hvert spørreskjema var generell informasjon om studien og kontaktinformasjon, på side 2 ble det presentert fire abstrakte resonneringsproblemer som var like for alle fire spørreskjemaer.

På toppen av siden med de abstrakte problemene sto det:

Gå ut i fra at det er sant at:

Alle XAR er YOF

og svar på multiple choice spørsmålene.

Etter dette ble det presentert fire multiple choice oppgaver. Den første så slik ut:

- (A) Hvis en bok er en XAR, kan du si
- a. med sikkerhet at boken er en YOF
 - b. med sikkerhet at boken ikke er en YOF
 - c. usikkert om boken er en YOF eller ikke.

Deretter så de tre andre multiple choice oppgavene slik ut:

- (B) Hvis en koie er en YOF, kan du si
- (C) Hvis en bok ikke er en XAR, kan du si
- (D) Hvis en koie ikke er en YOF, kan du si

Intensjonen med å starte spørreskjema med å løse abstrakte problemer uten å teste for belief-bias effekt er å oppmuntre til å resonnerer logisk. En slik prosedyre ble utformet av

Markovits og Nantel (1989) for å redusere eventuelle belief-bias effekter og for at effekten av instruksjonene som blir gitt før de påfølgende resonneringsproblemene ikke skal bli for tunge for deltakerne.

Av de fire forskjellige spørreskjemaene var to konstruert med et evalueringsparadigme, og to med et produksjonsparadigme. På side 3 av spørreskjemaene sto instruksjoner, og på side 4 og 5 ble det presentert åtte problemer fordelt på de to sidene. Fire av de åtte problemene presentert hadde en konklusjon som var logisk gyldig, mens de fire andre ikke hadde det. På første siden med problemer ble instruksjonene gjentatt.

Instruksjonene for evalueringsskjemaene var som følger (adaptert fra Markovits & Nantel, 1989):

Du kommer til å bli presentert med åtte problemer. Du må bestemme hvorvidt den gitte konklusjonen til hvert problem *følger logisk* fra premissene eller ikke.

Du må *gå ut ifra at alle premissene er sanne* og du må begrense deg selv til informasjonen gitt i premissene. Dette er veldig viktig.

På side 4, første side med syllogismer, sto disse instruksjonene på toppen av siden:

For hvert problem må du bestemme om den gitte konklusjonen følger logisk fra premissene. Sett ring rundt JA hvis, og bare hvis, du bedømmer at konklusjonen utvilsomt kan utledes fra premissene, ellers sett ring rundt NEI.

Det positive evalueringsskjemaet besto av følgende åtte syllogismer fordelt på to sider:

(1)

Premiss 1: Alle ting som blir røyka er bra for helsa.

Premiss 2: Sigaretter blir røyka.

Konklusjon: Sigaretter er bra for helsa.

(2)

Premiss 1: Alle som er arbeidsløse er fattige.

Premiss 2: Røkke er ikke arbeidsløs.

Konklusjon: Røkke er ikke fattig.

(3)

Premiss 1: Alle blomster har kronblader.

Premiss 2: Roser har kronblader.

Konklusjon: Roser er blomster.

(4)

Premiss 1: Alle dyr med fire bein er farlige.

Premiss 2: Pudler er ikke farlige.

Konklusjon: Pudler har ikke fire bein.

(5)

Premiss 1: Alle pattedyr kan gå.

Premiss 2: Hvaler er pattedyr.

Konklusjon: Hvaler kan gå.

(6)

Premiss 1: Alle østlige land er kommunister.

Premiss 2: Canada er ikke et østlig land.

Konklusjon: Canada er ikke kommunister.

(7)

Premiss 1: Alle dyr elsker vann.

Premiss 2: Katter liker ikke vann.

Konklusjon: Katter er ikke dyr.

(8)

Premiss 1: Alle ting som har en motor trenger olje.

Premiss 2: Biler trenger olje.

Konklusjon: Biler har motorer.

Det nøytrale evalueringsskjemaet var identisk med det positive, men byttet ut med følgende nøytrale begreper:

- (1) Ramioner er bra for helsa.
- (2) Nulle er ikke fattig.
- (3) Penner er blomster.
- (4) Argomeller har ikke fire bein.
- (5) Lapitarer kan gå.
- (6) Sylvania er ikke kommunistisk.
- (7) Sløver er ikke pattedyr.
- (8) Ontobiner har motorer.

Markovits og Nantel (1989) undersøkte med en uavhengig gruppe studenter om konklusjonene av syllogismene de lagde virket troverdige eller ikke. Formen av de åtte syllogismene er fordelt slik: 1 og 5 har formen «alle A er B, C er A, C er derfor B», 4 og 7 har formen «alle A er B, C er ikke

B, C er ikke A», i disse to formene ble konklusjonene rangert som ikke troverdige av de uavhengige studentene, til tross for at de er logisk gyldige. Syllogisme 3 og 8 har formen «alle A er B, C er B, C er A», og 2 og 6 har formen «alle A er B, C er ikke A, C er ikke B», i disse to formene ble konklusjonene vurdert som troverdige selv om de ikke er logisk gyldige.

De to produksjonsskjemaene var identiske til evalueringsskjemaene, men uten en skrevet konklusjon. På side 2 av spørreskjemaet sto instruksjonene:

Du kommer til å få presentert en serie av åtte problemer. Du må *produsere* en konklusjon som *følger logisk* fra premissene.

Du må *gå ut i fra at alle premissene er sanne* og begrense deg selv til informasjonen gitt i premissene. Dette er veldig viktig.

På toppen av side 3 av produksjonsskjemaene, før de første fire syllogismene ble presentert, sto det følgende:

For hvert problem skal du produsere en konklusjon som følger logisk fra premissene. Du skal kun formulere en konklusjon hvis du bedømmer at det er mulig å utlede uten tvil fra premissene. Hvis du tror det ikke er mulig å lage en konklusjon som følger logisk og utvilsomt fra premissene skal du ikke lage en, skriv bare INGEN.

Prosedyre

De fire forskjellige spørreskjemaene ble stokket om til randomsert rekkefølge og gitt ut til studentene. De ble informert om at det ikke forelå en tidsfrist og at de kunne holde på så lenge de følte for det, eller trekke seg når de måtte ønske. Skjemaene ble delt ut og samlet inn av eksperimentator og en medhjelper.

Mellom-observatørenighet

En uavhengig forsker fikk tilgang til et utvalg av spørreskjemaene (50 %) med fasit for å sjekke at rettingen av spørreskjemaene ble gjort riktig. Enighet ble kalkulert ved at antall enigheter mellom den uavhengige forskeren og eksperimentator ble delt på det totale antallet enigheter og uenigheter, deretter ganget med 100 for å finne prosent enighet. $56 \text{ spørreskjemaer} / 2 = 28$. $28 \times 8 \text{ (syllogismer per skjema)} = 224$. $\text{Enighet} = 224/224 = 1 = 100\%$.

Registrering av data

I henhold til metoden for registrering av data Markovits og Nantel (1989), ble deltakerne scoret på de abstrakte resonneringsproblemene, samt de åtte syllogismene. Scorene som ble registrert fra de åtte syllogismene blir kalt belief-scorer og indikerer om deltakeren svarte i henhold til troverdigheten av konklusjonen.

Abstrakte problemer. De abstrakte problemene ble scoret ut i fra hvordan deltaker hadde besvart alle fire problemer og ble kategorisert i en av fire mulige kategorier; kondisjonal (de som svarte riktig på alle fire problemer), mellomliggende (de som svarte riktig på modus ponens, og feil på en eller flere av de andre formene), bidireksjonal (de som svarte på alle fire problemene som om relasjonen var symmetrisk), og usikker (de som svarte feil på modus ponens).

Positivt evalueringsskjema. I de positive evalueringsskjemaene ble deltakerne gitt ett poeng hver gang de svarte at en troverdig respons var gyldig og en ikke-troverdig respons var ugyldig. Hvert skjema fikk derfor en score mellom 0 og 8. For eksempel ville en deltaker som bestemte at «sigaretter er bra for helsa» var en ugyldig konklusjon få 1 poeng, og at «Røkke er ikke fattig» var en gyldig konklusjon få 1 poeng.

Nøytralt evalueringsskjema. I de nøytrale evalueringsskjemaene ble det scoret på tilsvarende måte, det vil si at en deltaker som bestemte at «ramioner er bra for helsa» var ugyldig fikk 1 poeng eller at «penner er blomster» var gyldig fikk 1 poeng. Scorene i de nøytrale skjemaene indikerte ikke det samme som i de positive evalueringsskjemaene, altså troverdigheten av konklusjonen, men heller at de valgte den troverdige responsen av årsaker relatert til den logiske formen eller det globale innholdet i syllogismene.

Positivt produksjonsskjema. For å forklare scoringen av de positive produksjonsskjemaene på enklest mulig måte vil den logiske formen bli referert til på samme måte beskrevet av Markovits & Nantel (1989), med henvisning til det mindre premisset (for eksempel vil premissene «A er B, C er A» bli forkortet til «C er A»). Under scoringen av det positive produksjonsskjemaet ble det skilt på de ubestemte («C er B» og «C er ikke A») og de bestemte syllogismene («C er A» og «C er ikke B»). På besvarelse av de ubestemte syllogismene fikk deltakerne 1 poeng hver gang de lagde en ugyldig, men troverdig konklusjon, for eksempel på syllogismen «Alle østlige land er kommunistiske, Canada er ikke et østlig land» fikk deltakerne som lagde konklusjonen «Canada er ikke kommunistisk» 1 poeng. På besvarelse av de bestemte syllogismene fikk deltakerne 1 poeng dersom de ikke produserte den logisk gyldige, men ikke-troverdige konklusjonen, eller hvis konklusjonen de ga var usikker, for eksempel på syllogismen «Alle dyr med fire bein er farlige, pudler er ikke farlige» fikk deltakerne 1 poeng dersom de for eksempel svarte «Pudler har kanskje ikke fire bein» eller «INGEN».

Nøytralt produksjonsskjema. De nøytrale produksjonsskjemaene ble scoret på lik måte som de positive produksjonsskjemaene, 1 poeng hver gang de produserte eller ikke produserte konklusjoner i likhet med forklaringen over. For eksempel på syllogismen «Alle østlige land er kommunistiske, Sylvania er ikke et østlig land» fikk deltakerne som lagde konklusjonen «Sylvania er ikke kommunistisk» 1 poeng, og på syllogismen «Alle dyr med fire bein er farlige,

argomeller er ikke farlige» fikk deltakerne som svarte «INGEN» eller lagde en konklusjon som var usikker, 1 poeng.

Resultater

Tabell 1 indikerer gjennomsnittlige belief-skårer for hver av de fire logiske formene (to problemer per form i hvert skjema) som en funksjon av type spørreskjema (evaluering eller produksjon) og innhold (positiv eller nøytral). Funnene styrker teorien om belief-bias effekten i resonnering da skjemaene med positivt innhold, med begreper alle har tidligere erfaring med, har høyere total belief-skåre enn de med nøytrale begreper. I skjemaene med positivt innhold er skåren for både evaluering (4,43) og produksjon (3,82) høyere enn for de nøytrale skjemaene (2,73 og 2,88). I tillegg indikerer summen av belief-skåre på hver av de logiske formene at de ubestemte syllogismene som ikke kan ha logisk gyldig konklusjon er de som oftest blir bedømt med effekt av belief-bias (ubestemt skårer: 4,94 og 4,03 mot bestemt skårer: 1,81 og 3,08).

Mer spesifikt kan man observere at for de nøytrale problemene i evalueringsparadigme skårer deltakerne høyere på belief bias på de ubestemte problemene, de uten logisk gyldig konklusjon (0.80 og 0.87), enn på de bestemte, med logisk gyldig konklusjon (0.40 og 0.67). For de positive problemene i evalueringsparadigme er belief-skåren høyest for syllogismene med konklusjonene «roser er blomster» og «biler har motor», troverdig, men ugyldig konklusjon (1.71), og «pudler har ikke fire bein» og «katter er ikke dyr», ikke troverdig, men gyldig konklusjon (1.0).

For de nøytrale syllogismene i produksjonsparadigme er det høyest utslag av belief-bias effekt under produksjon av konklusjon på de to ubestemte formene «C er B» (1,06) og «C er ikke A» (1,13). For syllogismene med positivt innhold er det liknende resultat, 1,36 og 1,18 (ubestemt) mot 0,36 og 0,91 (bestemt).

Tabell 2 viser andel av observerte responser for de fire logiske formene, fordelt på mulige responser (responser som indikerer belief-bias eller logisk korrekthet) i produksjonsskjemaene som en funksjon av positivt og nøytralt innhold. Under inspeksjon av tabellen kan man se at for de to ubestemte formene («C er B» og «C er ikke A») er det høyere andel responser som samsvarer med belief-bias enn som samsvarer med den logisk korrekte responsen, dette gjelder for både nøytralt og positivt innhold. For de bestemte formene («C er A» og «C er ikke B») er det en vesentlig andel av logisk korrekte responser. Skillene er mest tydelig for syllogismene med nøytralt innhold, men mindre tydelig for de med positivt innhold, for eksempel er skåren på den bestemte formen «C er ikke B» fordelt 0,78/0,22 med nøytralt innhold, og 0,55/0,46 med positivt innhold.

En annen observasjon som er verdt å poengtere er at av de to bestemte formene viser tallene en sterk tendens til at deltakerne velger ingen spesifikk konklusjon heller enn å produsere den lett tilgjengelige troverdige (men logisk ugyldige) responsen. Det samme fant også Markovits og Nantel (1989) i sine tall. For eksempel, i stedet for å gi konklusjonen «pudler har fire bein» som er den mest troverdige responsen, viser tallene at deltakerne heller velger å skrive «ingen».

Tabell 3 viser gjennomsnittlig belief-skåre, kombinert evaluerings- og produksjonsskema, fordelt på nøytralt og positivt innhold som en funksjon av deltakernes respons på de abstrakte problemene. Deltakerne ble skåret basert på hvordan de hadde svart på alle fire problemene, med utgangspunkt i den første logiske formen, modus ponens. Det var kun 6 deltakere som hadde valgt riktig alternativ på alle fire problemer, gjennomsnittlig belief-skåre for nøytralt innhold var 0.80 mens det for positivt innhold var 1. Størst andel deltakere, 32 stykker, svarte riktig på modus ponens, men hadde en eller flere feil på de tre andre problemene, disse ble kategorisert som mellomliggende og gjennomsnittlig belief-skåre for nøytralt innhold 2.94 og 3.93 for positivt innhold. 16 deltakere behandlet relasjonen som en bidireksjonal, altså som at

hvis XAR så YOF og hvis YOF så XAR, skåren for nøytralt innhold var 3.56 og for positivt innhold 5.14. Kun to deltakere svarte feil på modus ponens, begge med positivt innhold, gjennomsnittlig belief-skåre var 4.

Diskusjon

Resultatene av den empiriske studien underbygger tidligere funn om at belief-bias har en effekt på resonnering og den såkalte konflikten mellom logikk og tro i syllogistisk resonnering (Evans et al., 1983; Oakhill & Johnson-Laird, 1985). Funnene i tabell 1 bekrefter primærhypotesen ved at belief-bias er mer fremtredende i syllogismene med positivt innhold enn med nøytralt. I evalueringsparadigme under nøytralt innhold var gjennomsnittlig belief-skåre 2.73 mens den under positivt innhold var 4.43. Den samme effekten kan observeres under produksjonsparadigme med en belief-skåre på 2.88 for nøytralt innhold og 3.82 for positivt innhold. Sekundærhypotesen om at belief-bias vil være mer fremtredende i produksjonsparadigme enn i evalueringsparadigme ble det ikke funnet støtte for. Ved å legge sammen belief-skårene for nøytralt og positivt innhold under hver av paradigmene viser det seg at effekten av belief-bias er større for evaluering (7.16) enn for produksjon (6.70). Sekundærhypotesen ble gjort på grunnlag av resultatene til Markovits og Nantel (1989) hvor de fant at sammenlagt belief-skåre for evaluering var 5.41 og for produksjon 6.28. Dog hadde de en hypotese om at belief-skårer i produksjon ville være lavere enn i evaluering, og deres hypotese ble dernest motbevist. I sin diskusjon rundt resultatene drøfter de om det kan være at de prosessene som belief-bias opererer i er forskjellige i de to paradigmene. Markovits og Nantel ville motvirke deltakerens mulige fokus på konklusjonen i stedet for premissene, tidligere nevnt selektiv undersøkelsesmodell, og mente at produksjon av konklusjoner ville gjøre nettopp dette sammenlignet med evaluering. Dette er det funnet støtte for i det foreliggende eksperimentet. Der hvor deltakerne er tvunget til å vurdere premissene (i produksjon) mener Markovits og Nantel at

en mulig årsak til belief-bias effekten kan være at deltakerne kommer opp med flere mulige og troverdige konklusjoner, men at under konkurransen om hvem av disse konklusjonene som vinner frem vil den som er mest troverdig seire. Hvorfor resultatene i det foreliggende eksperimentet er annerledes blir til dels spekulasjon, men det kan ha noe med selve replikasjonsprosessen å gjøre. Det er strenge krav til hva som kan kalles replikasjon og ikke, metoden brukt i det foreliggende eksperimentet er annerledes fra Markovits og Nantels ved at materialene i eksperimentet ble oversatt til norsk. Oversettelsen ble gjort nøye slik at det skulle ligne mest mulig på deres eksperiment og slik at begrepene, både de positive og nøytrale, skulle ha samme betydning for deltakeren.

I likhet med resultatene i det foreliggende eksperimentet var belief-skårene for positivt innhold høyere enn for nøytralt i begge paradigmene i eksperimentet til Markovits og Nantel (1989); nøytral evaluering: 2.36, positiv evaluering: 3.03, nøytral produksjon: 2.39, og positiv produksjon: 3.89. I tillegg til at det ble regnet ut belief-skåre ut ifra innhold og paradigme, ble det også delt inn i de fire logiske strukturene syllogismene ble lagd, se tabell 1. Resultatene i det foreliggende eksperimentet viser tydelig at strukturene «A er B, C er B» og «A er B, C er ikke A» var mest preget av belief-bias effekten totalt sett. Disse to strukturene var ubestemte, det kunne ikke trekkes en gyldig konklusjon ut i fra premissene, dette er også et funn som samsvarer med resultatene til Markovits og Nantel og Oakhill og Johnson-Laird (1985). Funnet underbygger en av prediksjonene som blir gjort av misforstått nødvendighetsmodell om at når deltaker blir utsatt for en ubestemt syllogisme vil han eller hun til tross for å raskt forstå at det ikke finnes en gyldig konklusjon, avstå fra å svare nettopp det, til fordel for en læringshistorie av å avgi svar under press. Oakhill og Johnson-Laird fant ett unntak fra denne trenden i data, og det var hva angikk konklusjoner som var definisjonelt falske, da svarte heller deltakerne den logiske korrekte responsen «ingen gyldig konklusjon». Oakhill og Johnson-Laird fant at responser viste høyere

effekt av belief-bias dersom de måtte produsere en empirisk falsk konklusjon enn en definisjonelt falsk, dette tyder på at deltakerne har lettere for å finne omstendigheter (via mentale modeller) hvor et empirisk falskt argument kan passe inn. Misforstått nødvendighetsmodell predikerte at deltaker ikke ville svare at det ikke fantes en gyldig konklusjon når utsatt for en ubestemt syllogisme, og de antok at det var på grunn av en læringshistorie av å avgi svar under press. Selv om resultatene i det foreliggende eksperimentet viser at prediksjonen kan stemme, så vil ikke modellen være passende for å forklare resten av resultatene, nettopp fordi ordet nødvendig ikke forekommer i instruksjonene. Markovits og Nantel formulerte sine instruksjoner på en slik måte som for å unngå det ordet – og det samme ble gjort i det foreliggende eksperimentet. I stedet for å bruke ordet nødvendigvis, omformulerte Markovits og Nantel til at konklusjonen måtte *uten tvil* følge fra premissene.

Tabell 2 som er laget for produksjonsparadigme og viser mer spesifikt forholdet mellom produsering av troverdige og/eller valide konklusjoner eller ingen spesifikk samsvarer til en viss grad med resultatene til Markovits og Nantel (1989). Der det er størst forskjell mellom dataene er graden av respons som indikerer belief-bias effekt på de ubestemte syllogismene. I det foreliggende eksperimentet er belief-skåren for den ubestemte formen «C er B» for nøytralt innhold 0,53 og positivt innhold 0,64, mens det hos Markovits og Nantel er 0,40 for nøytralt innhold og 0,74 for positivt innhold. For de ubestemte formene er logisk korrekt respons ansett for å være at det ikke er noen spesifikk konklusjon. For den ubestemte formen «C er ikke A» er andel skårer for nøytralt innhold 0,56 og for positivt innhold 0,59, mens det hos Markovits og Nantel er 0,29 for nøytralt innhold og 0,49 for positivt innhold. De to bestemte formene hvor det var mulig å produsere en gyldig konklusjon viste størst andel av logisk korrekt respons, ved formen «C er A» ble responsene fordelt med andel 0,91 logisk korrekt respons og andel 0,09 belief-bias respons under nøytralt innhold. Ved positivt innhold ble responsene fordelt med andel

0.82 logisk korrekt respons og 0.19 belief-bias respons. Formen «C er A» er derfor strukturen med høyest grad av logisk korrekte responser i produksjonsparadigme, og sammenlignet med tallene fra evaluering i tabell 1 ser man at den samme strukturen er den med lavest belief-skåre både under nøytralt og positivt innhold.

De abstrakte problemene alle deltakerne måtte svare på før de startet på syllogismene skulle ha den funksjon å starte den logiske resonneringsprosessen, det skulle fungere som en oppvarming. Effekten ønsket av Markovits og Nantel (1989) var at det skulle redusere belief-bias effekten i syllogismene. For de deltakerne som ble kategorisert som kondisjonale (alle riktige) var belief-bias effekten minimal, med 0.80 og 1 i gjennomsnitt belief-skåre. Den kategorien med høyest gjennomsnittlig belief-skåre var gruppen som ble skåret som at de hadde en bidireksjonal forståelse av forholdet mellom XAR og YOF, for nøytralt innhold er gjennomsnittet 3.56 og for positivt innhold 5.14. Dette funnet tyder på at antakelsen om at det er et gjensidig avhengig forhold mellom stimuliene fører til antakelser om syllogismene i tillegg. Trenden i resultatene i det foreliggende eksperimentet er like de resultatene Markovits og Nantel fikk, det var lavest grad av belief-bias effekt for de som ble kategorisert kondisjonale, og høyere for de som ble kategorisert i noen av de andre gruppene.

Belief-bias er et mentalistisk fenomen og har kun blitt undersøkt av den kognitivistiske fagdisiplinen. Ideen om kognitive bias som systematiske feil i resonnering passer ikke med den tradisjonelle økonomiske modellen av menneskelig resonnering som beskriver hvordan vi skal ta optimale valg hele tiden. Prospektteori tar utgangspunkt i at mennesker er begrenset i sin evne til å resonner optimalt, og at valgene vi tar er basert på mulige tap eller vinningsveid opp mot utgangspunktet for valget (Kahneman og Tversky, 1979). I tillegg til prospektteorien var Kahneman og Tversky (1974) også først ute til å snakke om de systematiske feilene vi gjør når vi bruker heuristikker for ofte. I deres utvikling av prospektteorien åpnet de også for måter

mennesker kunne gjøre feilslutninger, og belief-bias, blant andre kognitive bias, ble en del av kognitiv psykologi sin forklaring av menneskelig atferd.

Atferdsanalyse ser på mange kognitive fenomener som hypotetiske konstrukter, ukontrollerbare i en eksperimentsetting, og derfor også utilstrekkelig som eneste forklaring av atferd. Vi går ut ifra at instruksjonene deltakerne ble gitt muntlig og skriftlig før de skulle løse oppgavene fungerte som foranledninger for å utføre atferd. Å kunne lese instruksjoner og forstå det, er en del av verbal atferd. Alle deltakerne i studien var studenter på universitetsnivå, så vi antar at alle har en læringshistorie med en variasjon av verbal atferd, dataene er også en indikasjon på at deltakerne forsto oppgaven. En læringshistorie av regelfølgning er nødvendig for at studentene skal vite hva de skal gjøre i møte med en regel. I likhet med all operant atferd må den ha blitt lært og etablert i et individs atferdsrepertoar med en historie av forsterkning for at atferden skal forekomme under liknende betingelser ved en senere anledning. Neste steg i analysen blir så å forsøke å finne årsaken til at individet avviker fra regelen. Atferdsanalysen inkluderer motivasjon som en del av forklaring av atferd, f.eks. når vi er deprivert for vann vil motivasjonen for å finne vann være høy. Spørsmål kan da stilles til hva som er motivasjonen til deltakerne for å følge regelen, og da må vi også snakke om konsekvenser. Atferdsanalyse baserer forskningen sin på at atferd blir selektert ved sine konsekvenser. Instruksjonene er viktige i forklaring av atferden fordi den signaliserer til deltaker at forsterker er tilgjengelig. Instruksjonene i dette eksperimentet spesifiserer kun hva deltakeren skal gjøre, og ingenting om hva som er konsekvensen av atferden direkte, derfor kan vi anta at konsekvensene som motiverer deltakeren til problemløsningsatferden ikke er i direkte relasjon med selve atferden å bedømme gyldigheten av (eller produsere) en konklusjon. Sporing, når motivasjonen for å følge en regel ligger i konsekvensen av å følge eller ikke følge regelen, kan vurderes som et svar her. Deltakerne i studien var alle studenter ved OsloMet, og de ble rekruttert ved at eksperimentator

kom inn i klasserom etter endt forelesning og spurte om de ville delta i et eksperiment som ville hjelpe eksperimentator å skrive masteroppgaven. Akkurat hva som motiverte de enkelte studentene til å si ja, om det var en læringshistorie av å hjelpe andre, om det var vissheten om at de kanskje ville trenge lignende hjelp i fremtiden, eller om det var en følelse av ansvar ovenfor læringsinsituasjonen, blir spekulasjon uten å snakke med studentene og be dem spesifisere årsaken. Et annet motiv som kan vurderes er presset studentene kan føle av å bli spurt i plenum, det kan være at noen studenter sa ja for ikke å skille seg ut fra mengden, dog skal det bemerkes at det aldri var en hel klasse som sa ja, og at det alltid var flere som sa nei og gikk ut av klasserommet før eksperimentet startet. Sporing er en fullstendig legitim måte å følge en regel på, så hvis vi tar som utgangspunkt at motivasjonen lå til rette for at deltakerne skulle følge regelen, for eksempel at konsekvensen å få anerkjennelsen fra masterstudenten og foreleser for at de sa ja til å bli med var motivasjon nok, så må vi lete etter årsak for den såkalte belief-bias effekten et annet sted. Selve oppgaven, å vurdere premisser for å evaluere eller produsere en gyldig konklusjon, er en oppgave i såkalt deduktivt resonnement. Deltakerne kan ha varierende grad av erfaring med å gjøre dette, og strukturen på syllogismene brukt i dette eksperimentet var designet for å evokere frem feil respons ved at troverdige konklusjoner ble gitt eller evokert frem for syllogismer som var ugyldige, og ikke troverdige konklusjoner ble forsøkt evokert frem når premissene la grunnlag for en gyldig konklusjon. Selv om instruksjonene var klare på at de kun skulle evaluere en konklusjon som gyldig (eller produsere) dersom den fulgte logisk fra premissene, ble belief-bias komponenten lagt til for å se om deltakerne klarte å følge regelen. I instruksjonene ble det også spesifisert at deltakerne kun skulle forholde seg til informasjonen gitt i premissene og gå ut ifra at informasjonen var sann. Å trekke konklusjoner ut ifra informasjon vi blir gitt av andre eller av miljøet rundt oss er noe vi gjør hele tiden, men det er sjeldnere at man opplever å være i en situasjon hvor man blir spurt om å løse spesifikke oppgaver i det. Fra et

atferdsanalytisk perspektiv blir belief-bias effekten et avvik fra regelfølgning, og vi snakker ikke om mentale prosesser slik som kognitiv psykologi gjør.

Et fenomen som er mentalistisk kan ikke forventes å kunne bli fullstendig forklart av atferdsanalysen. Palmer (1991) skrev at det beste som kan oppnås når fenomener er for store eller vage til å kontrolleres i eksperimentsetting er fortolkning. Fortolkningen som er gjort her er basert på vitenskapelige prinsipper atferdsanalysen står for. Hadde en atferdsanalytiker skulle undersøkt belief-bias effekten måtte man forsøkt å isolere øyeblikket man vet at belief-bias oppstår hos individet. Feilen i regelfølgning måtte ha blitt observert over tid og det måtte ha blitt forsøkt analysert og identifisert funksjonelle relasjoner mellom atferden og andre variabler utenfor individet. Belief-bias måtte blitt forsøkt definert som en uavhengig variabel som vi måtte ha fått kontroll over, for å se effekten på den avhengige variabelen; regelfølgningen. Kognitiv psykologi beskriver belief-bias effekten som et delvis ubevisst fenomen, måten de gjør det bevisst på er at de i ulike forklaringsmodeller beskriver hvordan vi under resonnering og søken etter alternative konklusjoner velger å basere avgjørelsen på tro når vi er usikre. Hvis belief-bias til en viss grad er en bevisst handling kan den komme under kontroll av forsterkning. Dog å isolere det øyeblikket hvor individet, etter å ha forsøkt å resonnerer deduktivt uten påvirkning av tro, må ta med tro som en faktor i vurderingen virker vanskelig og muligens umulig.

Atferdsanalytiske prinsipper kan brukes til å fortolke belief-bias effekten til en viss grad, men etterhvert stopper det uten at man har forklart fenomenet ordentlig etter atferdsanalytisk standard. Ved atferdsanalyse tilbyr fortolkning etter atferdsanalytiske prinsipper etablert gjennom vitenskapelig forskning, og kognitiv psykologi retter fokuset mot utenforliggende variabler i søken etter årsakssammenhenger, kan vi muligens oppnå større forståelse for privat, menneskelig atferd.

Referanser

- Baum, W. M. (2005). *Understanding behaviorism: behavior, culture, and evolution* (2 ed.). Oxford: Blackwell publishing.
- Dougher, M. J. (1995). A bigger picture: Cause and cognition in relation to differing scientific frameworks. *Journal of Behavior Therapy & Experimental Psychiatry*, 26(3), 215-219.
- Evans, J., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, 11(3), 295-306.
doi:<http://dx.doi.org/10.3758/BF03196976>
- Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (s. 153-190). New York: Springer.
- Hayes, S. C., & Ju, W. (1998). Rule-governed behavior. In W. O'Donohue (Ed.), *Learning and behavior therapy* (s. 374-391). Boston: Allyn & Bacon.
- Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263-292.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Markovits, H., & Bouffard-Bouchard, T. (1992). The belief-bias effect in reasoning: The development and activation of competence. *British Journal of Developmental Psychology*, 10(3), 269-284. doi:<http://dx.doi.org/10.1111/j.2044-835X.1992.tb00577.x>
- Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & cognition*, 17(1), 11-17.

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, 45(3), 257-284.

Oakhill, J., Johnson-Laird, P., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31(2), 117-140.

Oakhill, J. V., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The Quarterly Journal of Experimental Psychology*, 37(4), 553-569.

Palmer, D. C. (1991). A behavioral interpretation of memory. In J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (s. 261-279). Reno, NV: Context press.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69, 99-118.

Simon, H. A. (1983). *Reason in human affairs*. California: Stanford University Press

Skinner, B. F. (1976). *About behaviorism* New York City: Vintage books / Random house.

Skinner, B. F. (2014). *Verbal behavior*. Cambridge, USA: BF Skinner Foundation.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: toward a test of rational thinking*: MIT Press.

Toplak, M. E., Stanovich, K. E., & West, R. F. (2012). Intelligence and rationality. In R. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence (3rd Edition)* (s. 784-826). Cambridge UK: Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124-1131.

Tabell 1

De fire logiske formene

Logisk form	Evaluering		Produksjon	
	Nøytral	Positiv	Nøytral	Positiv
A er B, C er A	0,40	0,86	0,19	0,36
A er B, C er B	0,80	1,71	1,06	1,36
A er B, C er ikke A	0,87	0,86	1,13	1,18
A er B, C er ikke B	0,67	1,00	0,50	0,91
Sum	2,73	4,43	2,88	3,82

Notat. Tabellen viser gjennomsnittlige belief-skårer (maksimum 2 poeng) for de fire logiske formene som en funksjon av type spørreskjema (evaluering eller produksjon) og innhold (positiv eller nøytral). Første kolonne viser de fire logiske formene premissene ble konstruert som. Andre kolonne viser gjennomsnittlig belief-skåre for de fire logiske formene for evalueringsskjema med nøytralt innhold. Tredje kolonne viser gjennomsnittlig belief-skåre for de fire logiske formene for evalueringsskjema med positivt innhold. Fjerde kolonne viser gjennomsnittlig belief-skåre for de fire logiske formene for produksjonsskjema med nøytralt innhold. Femte kolonne viser gjennomsnittlig belief-skåre for de fire logiske formene for produksjonsskjema med positivt innhold. Nederste rad viser summen av gjennomsnittlig belief-skåre for hvert av de fire skjemaene.

Tabell 2

Mulige responser i produksjonsskjemaene

Premisser	Respons	Innhold	
		Nøytral	Positiv
A er B, C er A	*C er B	0,91	0,82
	^C er ikke B	0	0,05
	^ingen spesifikk konklusjon	0,09	0,14
A er B, C er B	^C er A	0,53	0,64
	C er ikke A	0	0
	*ingen spesifikk konklusjon	0,47	0,36
A er B, C er ikke A	C er B	0	0
	^C er ikke B	0,56	0,59
	*ingen spesifikk konklusjon	0,44	0,41
A er B, C er ikke B	^C er A	0	0,09
	*C er ikke A	0,78	0,55
	^ingen spesifikk konklusjon	0,22	0,36

* det logisk korrekte svaret ^ svar forbundet med belief-bias

Notat. Tabellen viser andel av observerte responser for de fire logiske formene fordelt på mulige responser i produksjonsskjemaene, som en funksjon av nøytralt og positivt innhold. Første kolonne viser de fire logiske formene premissene ble konstruert som. Andre kolonne viser de mulige responsene deltakerne kunne avgi for hver av de logiske formene, legg merke til at svarene som er logisk korrekt er markert med *, og svarene forbundet med belief-bias er markert med ^. Tredje kolonne viser andel (av 1,00) responser på hvert av alternativene for de fire logiske formene i nøytralt produksjonsskjema. Fjerde kolonne viser andel (av 1,00) responser på hvert av alternativene for de fire logiske formene i positivt produksjonsskjema.

Tabell 3

Abstrakte problemer

Respons	Nøytralt innhold	Positivt innhold
Kondisjonal	0,80	1,00
Mellomliggende	2,94	3,93
Bidireksjonal	3,56	5,14
Usikker	0	4,00

Notat. Tabellen viser gjennomsnittlige belief-skårer (skala fra 0 til 8, for evaluerings- og produksjonsskjema for nøytralt og positivt innhold som en funksjon av deltakernes responser til de abstrakte problemene. Første kolonne viser de ulike kategoriene deltakerne ble plassert i avhengig av respons. Andre kolonne viser gjennomsnittlig belief-skåre for hver av kategoriene for nøytralt innhold i spørreskjema. Tredje kolonne viser gjennomsnittlig belief-skåre for hver av kategoriene for positivt innhold i spørreskjema.