

Respondering i henhold til stimulusekvivalens som en funksjon av antall trials vs. antall programmerte konsekvenser i etablering av baselinerelasjonene

Erik Arntzen¹, Lars Rune Halvorsen^{1,2} og Christoffer Eilifsen¹

¹OsloMet– storbyuniversitetet og ²Høgskolen i Østfold

To eksperimenter undersøkte om antall trials under trening med en lineær serie (LS) treningsstruktur påvirket respondering i henhold til stimulusekvivalens. I Eksperiment 1 var 30 voksne forsøkspersoner fordelt i tre grupper, med 10 i hver gruppe. Alle treningsblokkene inneholdt henholdsvis 18 trials¹ (Gruppe 1), 36 trials (Gruppe 2), og 54 trials (Gruppe 3). I Eksperiment 2 ble 24 forsøkspersoner fordelt i grupper som lignet de i Eksperiment 1, men her var økningen i antall trials introdusert i en treningsblokk uten programmerte konsekvenser. For begge eksperimentene inneholdt testblokken testtrials og baselinetrials uten programmerte konsekvenser. Resultatene viste at antall forsøkspersoner som responderte i henhold til stimulusekvivalens var to i Gruppe 1, fire i Gruppe 2, og fem i Gruppe 3 for Eksperiment 1 og to, tre og fem for de tilsvarende gruppene i Eksperiment 2. Resultatene fra Eksperiment 1 og 2 antyder at den avgjørende variabelen for emergens av ekvivalensklasser er antall trials og ikke antall programmerte konsekvenser. Reaksjonstidsdataene er vist kun fra Eksperiment 1, og disse reflekterte ikke variasjoner i antallet av baselinetrials i treningsblokkene.

Nøkkelord: Stimulusekvivalens, antall trials, antall programmerte konsekvenser, LS treningsstruktur, studenter

Formation of Stimulus Equivalence as a Function of Number of Trials vs. Number of Programmed Consequences During Training of Baseline Relations

Two experiments investigated whether the number of trials during training with a linear series (LS) training structure affected responding in accordance with stimulus equivalence. In Experiment 1, 30 adult participants were divided into three groups, with 10 in each group. All training blocks contained 18 trials (Group 1), 36 trials (Group 2), and 54 trials (Group 3). In Experiment 2, 24 participants were divided into three groups as in Experiment 1 but the increase in the number of trials was only employed in the training block without programmed consequences. In both experiments, the test block contained test trials and baseline trials without programmed consequences. The results showed that the number of participants who responded according to stimulus equivalence was two in Group 1, four in Group 2, and five in Group 3 for Experiment 1 and two, three and five for the corresponding groups in Experiment 2. Experiments 1 and 2 suggest that the influential variable in producing equivalence classes, is the number of training trials and not the number of programmed consequences. The reaction time data are shown only for Experiment 1, and these data did not reflect variations in the number of baseline trials in the training blocks.

Key words: Stimulus equivalence, number of trials, number of programmed consequences, LS training structure, college students

¹Vi bruker trials i artikkelen ettersom det ikke finnes et dekkende norsk ord.

Forfatterne erklærer at det er ingen interessekonflikter. Alle prosedyrer er utført i henhold til etiske retningslinjer og er i overensstemmelse med 1964 Helsinki Deklarasjonen og dens senere tilføyelser. Dataene i Eksperiment 1 er samlet inn i forbindelse med andreforfatter sin masteroppgave ved Masterprogram for Læring i Komplekse Systemer (nå Masterprogram i atferdsvitenskap), HiAK (nå OsloMet). Forfatterne takker Andreas Nygård for innsamling av mesteparten av dataene i Eksperiment 2. All korrespondanse i forbindelse med manuskriptet adresseres til erik.arntzen@equivalence.net

Matching-to-sample (MTS) arrangement er utbredt brukt for å undersøke forskningsproblemer innen stimuluskontroll, og da også stimulusekvivalens. Stimulus-ekvivalens innebærer at alle stimuli i samme stimulusklasse er gjensidig utskiftbare (e.g., Green & Saunders, 1998). Egenskaper som blir brukt for å definere stimulus-ekvivalens-relasjoner er hentet fra matematisk mengde-teori (Hrbacek & Jech, 1999, pp 29–32), og disse egenskapene er refleksivitet, symmetri og transitivitet (e.g., Sidman & Tailby, 1982). Som regel så trenes og testes de betingede diskriminasjonene i et matching-to-sample (MTS) arrangement (se Arntzen, 2010, for detaljer). Ved for eksempel å trene de betingede diskriminasjonene AB og BC til et gitt mestringskriterium, kan det testes for refleksivitet som innebærer at A er relatert til B, B er relatert til C og C er relatert til A¹. For symmetri kan det testes om B er relatert til A og C er relatert til B. For transitivitet kan det testes om A er relatert til C. I tillegg er det en kombinert test hvor det testes om C er relatert til A (Sidman & Tailby, 1982). Stimulus-ekvivalens kan enten framkomme umiddelbart eller utsatt. Det sistnevnte omtales som *delayed emergence*.

Målinger av emergens av ekvivalensklasser har i hovedsak vært gjort ved prosent korrekt i testene og det er som ofte rapportert som antall forsøkspersoner som responderer i henhold til stimulus-ekvivalens (se Fields et al., 2020, for en diskusjon). Det er også diskutert andre tilleggsmål som reaksjonstid (Dymond & Rehfeldt, 2001). En rekke eksperimenter har vist hvorledes mønstre for reaksjonstid eller hastighet har endret seg fra baselinetrials til symmetri- og ekvivalens-trials (e.g., Arntzen, 2004; Bentall et al., 1993; Spencer & Chase, 1996). Et forhold som bør undersøkes videre er hvorvidt reaksjonstidsmønstre korrelerer med hvor godt baselinere-lasjonene er etablert.

En rekke variabler har vist seg å påvirke

om ekvivalensklasser framkommer eller ikke, og eventuelt hvorfor (se en oversikt i Arntzen, 2012). En slik variabel er trening og testprotokoller, og som innebærer forskjellige arrangementer for presentasjon og rekkefølgen av trials. Det er tre ulike trening og testprotokoller som har vært brukt i forskning på emergente relasjoner; *simultaneous* (SIM) protokoll, *simple-to-complex* (STC) protokoll og *complex-to-simple* (CTS) protokoll (se detaljerte beskrivelser i Buffington et al., 1997; Imam, 2006). SIM protokollen innebærer at alle baselinere-lasjonene blir presentert i en treningsblokk og disse relasjonene trenes til et visst mestringskriterium. Deretter introduseres alle testtrials i en eller flere testblokker. I de foreliggende eksperimentene brukes denne protokollen fordi den har vist seg å være sensitiv for å studere parametere som påvirker emergens av ekvivalensklasser.

En annen variabel er treningsstrukturer, som dreier seg om retningen av relasjonene i trening og testing. *Linear series* (LS), *One-to-Many* (OTM) og *Many-to-One* (MTO) er tre ulike treningsstrukturer som anvendes i forskning på emergente relasjoner (e.g., Saunders et al., 1993). Utgangspunktet er et minimum på tre medlemmer (ABC) i klassene. I LS vil AB og BC relasjoner bli trent, mens i OTM blir AB og AC relasjoner trent og i MTO blir AC og BC relasjoner trent. I LS vil B være noden (den stimulusen som binder to andre stimuli sammen), mens A vil være noden i OTM og C i MTO. Det er flere ulikheter mellom disse treningsstrukturene som gir seg utslag i forskjeller i antall forsøkspersoner som har respondert i henhold til stimulus-ekvivalens avhengig av hvilken treningsstruktur som anvendes (se en oversikt i Arntzen, 2012). En av disse ulikhetene er antall simultane og suksessive enkle diskriminasjoner i treningen. Saunders og Green (1999) sin diskriminasjonsanalyse tar utgangspunkt i hvilke og hvor mange simultane og suksessive diskriminasjoner som blir trent i henholdsvis LS, MTO og OTM. Antakelsen er at jo flere stimuli som blir diskriminert fra hverandre under etable-

¹Refleksivitet testes som regel ikke ettersom den egenskapen ikke er sett på som nødvendig for å vurdere emergens av de andre egenskapene.

ringen av de betingede diskriminasjonene, jo større sjansje er det for at ekvivalensklassene framkommer i testen. Ifølge denne diskriminasjonsanalysen er det kun MTO hvor alle nødvendige enkle diskriminasjoner blir trent. Kritiske blikk på diskriminasjonsanalysen har blitt fremmet av flere (for detaljer se for eksempel Ayres-Pereira & Arntzen, 2019; Sadeghi & Arntzen, 2018) og flere eksperimenter hvor det er en sammenlikning av alle tre treningsstrukturene har ikke uten videre støttet prediksjonene fra diskriminasjonsanalysen om at MTO skulle være overlegen OTM og LS i å produsere ekvivalensklasser (Arntzen & Hansen, 2011; Arntzen & Holth, 1997, 2000). Uansett, det er en enighet om at LS gir det laveste antallet av forsøkspersoner som responderer i henhold til stimulusekvivalens (Arntzen, 2012) og derfor egnet for å studere variabler som påvirker emergens av ekvivalensklasser.

En tredje variabel som kan påvirke resultatene på ekvivalenstestene er hvordan de tre treningsstrukturene er satt sammen. LS er forskjellig fra MTO og OTM blant annet med hensyn til antall stimuli med doble funksjoner. Under etableringen av betingede diskriminasjoner er det *kun* i LS at en stimulus eller flere stimuli har doble funksjoner. For eksempel ved trening med tre medlemmer A-->B-->C, så vil denne doble funksjonen innebærer at B fungerer som sammenlikningsstimulus når AB relasjonen trenes og som utvalgsstimulus når BC relasjonen trenes. Antall stimuli som har slike doble funksjoner i en LS treningsstruktur øker med antall medlemmer i klassene; med fire medlemmer er det to, med fem medlemmer er det tre osv.—alltid antall medlemmer minus to. Videre omtales en stimulus som innehar doble funksjoner som en node i en LS treningsstruktur. En node i MTO og OTM har ikke doble funksjoner, den er henholdsvis sammenlikningsstimulus i MTO og utvalgsstimulus i OTM. Disse strukturelle forskjellene er med på å understreke hvorfor LS i kombinasjon med en SIM protokoll sammenliknet med kombi-

nasjon av andre treningsstrukturer og test og treningsprotokoller gir det laveste antall forsøkspersoner som responderer i henhold til stimulusekvivalens. Dette underbygger også at LS er velegnet for å studere hvilke variabler som kan influere på testresultatene og derfor bakgrunnen for at LS er brukt i det foreliggende eksperimentet.

En fjerde variabel er hvorvidt baselinetrials er til stede i testblokkene eller ikke. For eksempel, Arntzen og Holth (1997) arrangerte et eksperiment med 10 forsøkspersoner i hver gruppe hvor det ikke var baselinetrials i testfasen. De fant at alle 10 i andre test halvdel i OTM gruppen responderte i henhold til stimulusekvivalens, mot sju i andre testhalvdel i MTO gruppen og tre i andre testhalvdel i LS gruppen. Gjennomsnittlig antall trials i MTO gruppen var høyere sammenliknet med OTM gruppen. Saunders et al. (1999) argumenterte for at en mulig forklaring på disse forskjellene i Arntzen og Holth (1997) eksperimentet var mangelen av baseline relasjonene i testen. Derfor arrangerte Arntzen og Holth (2000) et eksperiment som en *single case research design* hvor hver forsøkspersonene ble utsatt for alle tre treningsstrukturene i ulike rekkefølger. Baselinereelasjoner var inkludert i testblokken og resultatene viste at for noen av forsøkspersonene var disse relasjonene intakt til tross for at de ikke framviste ekvivalensklasser. Eilifsen og Arntzen (2009) undersøkte videre betydningen av om baselinereelasjoner var intakt eller ikke i testen i et eksperiment med 20 forsøkspersoner som ble trent med SIM protokollen og LS. Hovedfunnene var at på tross av at baselinereelasjonene var intakt under test, så responderte ni av forsøkspersonene ikke i henhold til stimulusekvivalens; kun tre forsøkspersoner responderte i henhold til stimulusekvivalens under første test. Sytten forsøkspersonene ble deretter eksponert for en ny runde med trening og test, og syv nye forsøkspersoner responderte her i henhold til stimulusekvivalens. Av de øvrige forsøkspersonene hadde syv av dem baselinereelasjonene intakt under test. Disse

resultatene svekker argumentet fra Saunders et al. (1999) om at baselinereelasjoner som bryter sammen under test, kan være en grunn til at forsøkspersoner ikke responderer i henhold til stimulusekvivalens.

En femte variabel som kan påvirke emergens av ekvivalensklasser er hvor godt baseline relasjonene er etablert, altså antall trials i treningen. *Overtraining*² er studert forholdsvis lite innen forskning på stimulus-ekvivalens (Minster et al., 2011). I Minster et al. eksperimentet ble ikke overtraining direkte manipulert. Derimot gjennomførte Bortoloti et al. (2013) et eksperiment hvor de trente en blanding av MTO (AB/AC) og LS (CD/DE) treningsstrukturer, hvor A stimuliene var ansikter som viste ulike emosjoner (sint, nøytral og glad), mens de andre stimuliene var abstrakte former. Treningen var arrangert som en variant av SIM protokollen, men en sekvensiell introduksjon av baseline trials; AB trent først deretter AC osv. En gruppe mottok 48 trials og den andre gruppen 96 trials i den miksete treningsblokken (som inneholdt alle trente relasjoner; AB, AC, CD og DE relasjoner) før testen. Testblokken inneholdt kun 24 trials med BE og EB relasjoner. Hovedresultatene viste at 11 av 17 med 48 i trials miksblokken og 10 av 17 med 96 trials i miksblokken responderte i henhold til ekvivalens. For forsøkspersonene som hadde respondert i henhold til ekvivalens brukte de en *semantic differential scale*³ for å se om det var en overføring av emosjonene fra A til D stimuli. De fant at det var større grad av overføring for gruppen som mottok flest trials. At Bortoloti et al. ikke fant forskjeller mellom gruppene (48 vs. 96 trials) i henhold til emergens av ekvivalensklasser kan være fordi testen ikke inneholdt alle testtrials.

I den foreliggende studien med to ekspe-

²Vi har valgt å bruke det engelske begrepet. Begrepet og betydningen av det er omtalt senere i artikkelen.

³Består av en intervallskala med polariserende ytterpunkter, hvor forsøkspersonen skal skåre på et kontinuum for eksempel for ordparet ondskapsfull/snill. Skalaen kan gå fra 1 til 5 eller 1 til 7. Dette er forskjellig fra en Likert skala hvor forsøkspersonen skal vurdere om han er uenig eller enig forskjellig utsagn.

rimerter ville vi undersøke effektene av antall trials i etableringen av de betingede diskriminasjonene (baseline) med bruk av LS og SIM protokollen og med abstrakte stimuli. Konkret var målet med disse to eksperimentene å se om robuste (antall trials mer enn minimum) etablerte betingede diskriminasjoner hadde større effekt på antall forsøkspersoner som formet ekvivalensklasser enn mindre robuste (bare et minimum antall trials) etablerte betingede diskriminasjoner. Derfor undersøkte vi effekten av et økende antall trials (18, 36 og 54) enten i alle treningsblokkene (Eksperiment 1) eller i treningsblokken uten programmerte konsekvenser (Eksperiment 2) før test av emergente relasjoner.

EKSPERIMENT 1

Basert på analysene av stimuluskontroll-topografi har Dube og McIlvane (1996) argumentert med at emergens av stimulus-ekvivalens kan bli påvirket av (1) raten av forsterkere i baseline og (2) *overtraining* under etablering av betingede diskriminasjoner i baseline. Forskningsspørsmålet i dette eksperimentet var om antall forsøkspersoner som responderte i henhold til stimulusekvivalens økte som en funksjon av antall trials i hver treningsblokk. I tillegg ønsket vi å undersøke om et høyere antall trials i treningsblokkene vil gi forskjeller i reaksjonstid til sammenlikningsstimulus på slutten av treningen og testen for emergente relasjoner enn et lavere antall trials i en treningsblokk.

Metode

Forsøkspersoner

Tretti voksne personer (18–57 år), 25 kvinner og menn, deltok i eksperimentet. Disse var studenter på høgsolenivå eller helse- og sosialarbeidere. Forsøkspersonene ble tilfeldig fordelt i tre eksperimentgrupper. Ingen av forsøkspersonene hadde på forhånd kjennskap til forskningsfeltet stimulusekvivalens, stimuliene som ble

brukt i eksperimentet eller måten de ble arrangert på. De ble informert om at eksperimentet hadde tilknytning til læringspsykologi, og at det krevdes ingen forkunnskaper med unntak av å kunne klikke med musepekeren på datamaskinen. Videre ble de muntlig og skriftlig informert om at de kunne trekke seg på et hvilket som helst tidspunkt, at eksperimentet ikke var forbundet med noen form for ubehag, og at resultatene kunne bli brukt i publisasjonsøyemed. Forsøkspersonene ble også informert om at tiden for gjennomføring av eksperimentet var ca. 90 minutter. Alle forsøkspersonene skrev under på et samtykkeskjema, og debrifing ble gitt til samtlige etter eksperimentet var ferdig.

Setting, apparatur og programvare

Det ble benyttet tre rom i eksperimentet. Rommene var arrangert som lukkede testrom for å minimere støy utenfra, samt at alle former forstyrrelser var forhindrede. Testrommene var henholdsvis 3X4 meter, 4X4 meter, og 6X6 meter. Felles for alle rommene var at forsøkspersonene var omgitt av enten vegger eller romdelere på tre kanter, og at testmaskinen var plassert inn mot en av disse. To av testrommene hadde vinduer.

To datamaskiner ble benyttet i eksperimentet. Begge datamaskinene kjørte Windows 7 og hadde 15.4-tommer skjermer. Programvaren som ble benyttet for å gjennomføre eksperimentet var et spesiallaget MTS program, PsychFusion. Rundt alle stimuliene var det en trykksensitiv sone, 3,4 cm i høyden og 9,4 cm i bredden, hvor forsøkspersonen kunne velge stimuliene med musepekeren. Avstand fra den trykksensitive sonen fra utvalgsstimulus til sammenligningsstimuliene var på 4,3 cm. Det var 10,95 cm i bredden mellom utvalgsstimuliene, og 11,8 cm i høyden mellom utvalgsstimuliene. Målene var like for begge skjermene. Stimuliene var visuelle og inkluderte greske, japanske, arabiske, og kyrilliske symboler som vist i Figur 1.

	1	2	3
A	𐀀	カ	ラ
B	𐀁	キ	𐀂
C	𐀃	𐀄	𐀅

Figur 1. Figuren viser stimuliene som ble benyttet i dette eksperimentet, og stimuliene var like for alle tre gruppene i Eksperiment 1 og 2.

Design

Det ble benyttet en mellom-gruppe design. Tabell 1 viser forskjellen mellom de tre eksperimentgruppene; I Gruppe 1 blir hver trial type presentert tre ganger per blokk; 18 trials per blokk. En trial type kan illustreres som A1B1B2B3, hvor første bokstav-tallkombinasjonen illustrerer utvalgsstimulus (A1) og de neste bokstav-tallkombinasjonene illustrer sammenligningsstimuliene der den eksperimentatordefinerte korrekte sammenligningsstimulusen (B1) er understreket. For Gruppe 2 og Gruppe 3 ble hver trial type presentert henholdsvis seks og ni ganger per blokk, 36 og 54 trials per blokk. Testbetingelsen var lik for alle de tre eksperimentgruppene.

Prosedyre

Instruksjoner

Før forsøkspersonene kunne starte selve testprogrammet, ble følgende instruksjon vist på PC skjermen;

Det vil komme en stimulus midt på skjermen. Du skal klikke på denne med musen. Tre andre stimuli vil komme til syne. Velg en av disse ved å klikke med musen. Hvis du velger den vi har definert som korrekt vil det stå bra, supert, osv. på skjermen. Hvis du trykker feil, så vil det stå feil på skjermen. Nederst på skjermen vil det også telles opp antall korrekte responser. I løpet av eksperimentet vil datamaskinen ikke gi tilbakemelding på om valgene er riktig eller feil, men ut fra det du har lært kan du få alle oppgavene riktig. Gjør så godt du kan for å få mest mulig riktig.

Lykke til!

Nederst på samme skjermbilde sto følgende instruks; Trykk her for å starte.

Trening

En SIM protokoll og en LS treningsstruktur (AB og BC) ble brukt i et MTS format. Alle forsøkspersonene gjennomgikk en treningsblokk hvor de betingede diskriminasjonene ble etablert etterfulgt av en test hvor det ble testet for respondering i henhold til stimulusekvivalens (se detaljer i Tabell 1).

Utvalgsstimulusen ble alltid presentert midt på skjermen, og etter at forsøkspersonene klikket på den med musepekeren ble de tre sammenligningsstimulene presentert randomisert i hjørnene på tvers av trials. Både utvalgsstimulus og sammenlikningstimulene forble på skjermen inntil forsøkspersonen trykket på en av sammenlikningsstimulene; simultan matching.

Avhengig av om forsøkspersonen valgte den sammenlikningsstimulusen som var eksperimentatordefinert som korrekt når en spesiell utvalgsstimulus var til stede, ble ord som for eksempel «Korrekt», «Bra», «Riktig», «Supert», eller «Flott» vist på skjermen. Dersom responsen ikke var i overensstemmelse med den eksperimentatordefinerte relasjonen ble teksten «Feil» presentert på skjermen. Varigheten av de programmerte konsekvensene var 1,5 sek. Mellom hver trial ble musepekeren automatisk resatt til en fast posisjon etter et *inter-trial-interval* på 1 sek.

Tabell 1. Tabellen viser de ulike typer relasjoner som ble trent og testet og antall trials i de forskjellige blokkene for Eksperiment 1 og 2.

Betingelse	Type relasjon	Antall trials
Trening	A1B1B2B3, A2B1B2B3, B1B2B3, B1C1C2C3, B2C1C2C3, B3C1C2C3	18 trials i Gruppe 1 36 trials i Gruppe 2 54 trials i Gruppe 3 per blokk
Testfase	A1B1B2B3, A2B1B2B3, A3B1B2B3, B1C1C2C3, B2C1C2C3, B2C1C2C3, B1A1A2A3, B2A1A2A3, B3A1A2A3, C1B1B2B3, C2B1B2B3, C3B1B2B3, A1C1C2C3, A2C1C2C3, A3C1C2C3, C1A1A2A3, C2A1A2A3, C3A1A2A3	54 trials i en testblokk; 18 BL, 18 SYM, 9 TRA og 9 EQ trials

Merknad: I Eksperiment 1 er forskjellene i antall trials for alle treningsblokkene, mens for Eksperiment 2 er forskjellene i antall treningstrials kun i blokken med uten for programmerte konsekvenser. BL=baseline; SYM=symmetri; TRA=transitivitet; EQ=ekvivalens.

Sannsynligheten for presentasjon av programmerte konsekvenser ble gradvis redusert under treningen gjennom fem treningsblokker. Den første treningsblokken hadde 100 %, dernest 75 %, 50 %, 25 % og til slutt 0 % sannsynlighet for programmerte konsekvenser. Kriteriet for å gå videre fra en treningsblokk var minimum 90% korrekt i henhold til de eksperimentatordefinerte relasjonene. Dersom kriteriet ikke ble oppnådd ble den aktuelle treningsblokken gjentatt helt til kriteriet ble møtt.

Test

Testen fulgte direkte etter den femte treningsblokken, og bestod av 54 trials som ble randomisert presentert og inneholdt av testing av baselinerelasjoner (18 trials), symmetrirelasjoner (18 trials), og transitivitet/ekvivalensrelasjoner (18 trials). Det ble ikke presentert programmerte konsekvenser under testen.

Kriteriet for respondering i henhold til stimulusekvivalens, ble satt til minimum 17 av 18 korrekte responser (over 94,4% korrekte), for baseline-, symmetri-, transitivitet- og ekvivalensrelasjoner.

Avhengige variabler

Antall trials. For hver treningsblokk ble det registrert følgende; totale antall trials (korrekte og feil), og hvor mange treningsblokker eventuelt måtte presenteres på nytt på grunn av at mestringskriteriet for en aktuell blokk ikke ble nådd.

Reaksjonstid. Den ble målt fra presentasjonen av utvalgsstimulus til en av sammenligningsstimuliene ble trykket på via musepekeren. Gjennomsnitt ble utregnet av de individuelle reaksjonstidene, og deretter gjennomsnitt for hver gruppe. Reaksjonstidene som ble målt var de fem siste trials under siste treningsblokk og de fem første og fem siste testtrials.

Statistiske analyser

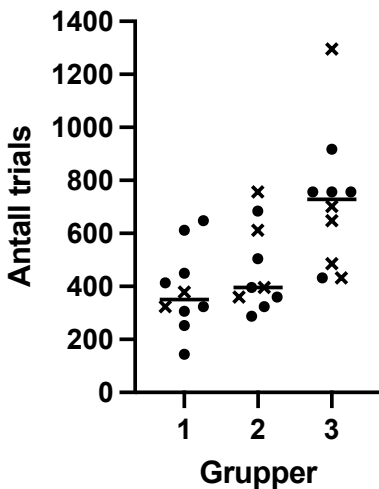
Det er ikke gjort statistiske analyser av antall trials i baseline, etableringen av de betingede diskriminasjonene, ettersom antallet vil være forbundet med manipuleringen i eksperimentet. Vi har benyttet oss av Fisher Exact Test (FET) for resultatene for testene for emergente relasjoner. FET er benyttet ettersom disse dataene er ikke-parametriske eller kategoriske data, i dette tilfelle innebærer det er en binær variabel (respondering i henhold til emergente relasjoner eller ikke) og fordi analysene er gjort på små

utvalg. Det er ikke gjort statistiske analyser for reaksjonstidsdataene mellom gruppene ettersom mønstre er svært likt. Derimot er det gjort statistiske analyser av forløpet fra baseline til testing innad i gruppene og da er det brukt ANOVA, gjentatte målinger, og t-tester.

Resultater og diskusjon

Antall trials under etablering av betingede diskriminasjoner

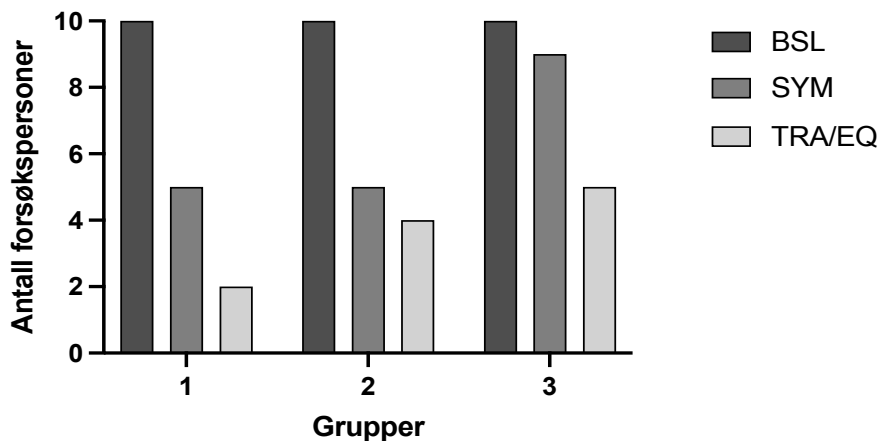
Figur 2 viser det totale antall trials under etableringen av baselinereelasjonene for de tre gruppene. For de to første gruppene er det veldig liten forskjell i gjennomsnittlig antall trials for å nå mestringskriteriet; gjennomsnittet er 351 for Gruppe 1 og 396 for Gruppe 2. For Gruppe 3 er det en betraktelig økning, 729, som også er den gruppen som har det høyeste antall trials i hver treningsblokk. Variasjonen i Gruppe 1 og 2 er omtrent det samme, men større i Gruppe 3.



Figur 2. Antall trials er vist i figuren for hver enkelt forsøksperson i de tre gruppene i Eksperiment. *Merknad:* De vertikale strekene viser gjennomsnittet i hver gruppe. Hvert datapunkt illustrerer antall trials for den enkelte forsøksperson. De datapunktene som er markert med kryss viser de forsøkspersonene som responderte i henhold til stimulusekvivalens.

Analyser av sammenheng mellom antall trials og respondering i henhold til stimulusekvivalens

I Gruppe 1 hadde to av de forsøkspersonene som responderte i henhold til stimulusekvivalens enten over eller under gjennomsnittet av antall treningstrials, mens de åtte andre forsøkspersonene hadde fire under og fire over gjennomsnittet av antall trials (se Figur 2). Videre viser figuren at i Gruppe 2 hadde to av fire forsøkspersoner som responderte i henhold til stimulusekvivalens antall trials under eller som gjennomsnittet, mens to forsøkspersoner hadde over gjennomsnittet av antall trials. De resterende seks forsøkspersonene som ikke responderte i henhold til stimulusekvivalens, var antall trials for fire forsøkspersoner under eller som gjennomsnittet og to forsøkspersoner lå over gjennomsnittet. I Gruppe 3 hadde fire av fem forsøkspersoner som formet ekvivalensklasser et lavere antall trials enn gjennomsnittet, mens fire av fem som ikke responderte i henhold til stimulusekvivalens hadde antall trials som lå over gjennomsnittet. Ingen av



Figur 3. De tre stolpene for hver gruppe viser antall forsøkspersoner som responderte i henhold til de eksperimentator-definerte kriteriene for henholdsvis baseline (BSL), symmetri (SYM) og transitivitet/ekvivalens (TRA/EQ) i Eksperiment 1.

FET testene som ble brukt viste signifikante resultater (se Tabell 2).

Respondering i henhold til stimulusekvivalens

Figur 3 viser antall forsøkspersoner som responderte i henhold til stimulusekvivalens (baselinetrials måtte være intakt og symmetri og transitivitet/ekvivalensstrials måtte være over 90% korrekt) for de tre gruppene. Fra Gruppe 1 til Gruppe 2 var det en liten økning i antall forsøkspersoner som responderte i henhold til stimulusekvivalens, fra to til fire, og en ytterligere økning i Gruppe 3 hvor det var fem forsøkspersoner som responderte i henhold til stimulusekvivalens.

Resultatene for de forsøkspersonene som ikke møtte kriteriet i testblokken ble videre analysert for å se om det var tendenser til at det var forsinket emergens av stimulus-ekvivalens. Disse resultatene ble analysert i forhold til om det var mange feilresponser i første halvdel av testen i motsetning til siste testhalvdel (*delayed emergence*). Ingen av forsøkspersonene viste slike responsmønstre. Ekvivalenstesten bestod av 18 trials og inneholdt ni transitivitetstrials og ni ekvivalensstrials. Testen ble splittet opp for å

undersøke om de enkelte forsøkspersonene responderte i henhold til transitivitet, men ikke i henhold til ekvivalens. Resultatet viste ingen tegn til slike tendenser.

Testene for respondering i henhold til symmetrirelasjonene viste at sammenliknet med transitivitet/ekvivalensstrials var det ytterligere tre forsøkspersoner fra Gruppe 1 og ytterligere en forsøksperson fra Gruppe 2 responderte korrekt i henhold til symmetri. Den største økningen kom imidlertid i Gruppe 3, hvor ytterligere fire av forsøkspersonene som gjorde dette i forhold til symmetrirelasjonene.

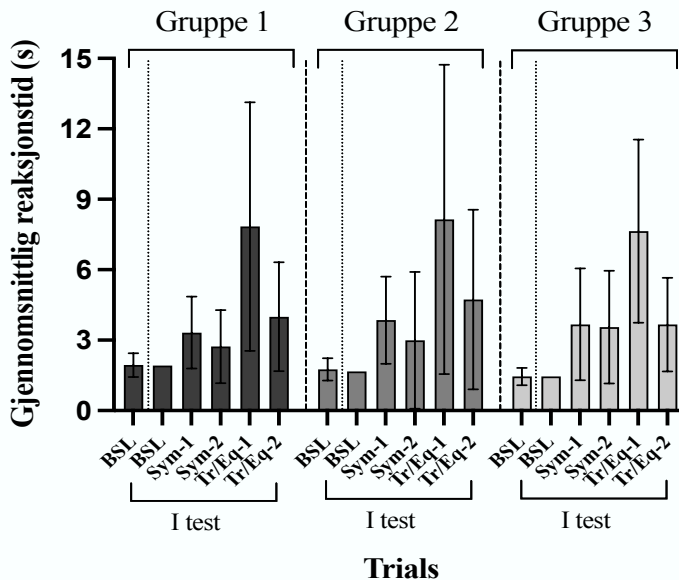
Reaksjonstid

Figur 4 viser at det samme karakteristiske reaksjonstidsmønsteret framkom for alle gruppene—reaksjonstiden til sammenlikningsstimuli for de fem første trials under test for symmetri og transitivitet/ekvivalens økte sammenliknet med de fem siste trials under baselinetrening (reaksjonstiden for transitivitet og ekvivalensstrials er slått sammen ettersom det ikke var forskjeller mellom disse). Økningen er betraktelig større for transitivitet/ekvivalensstrials sammenliknet med symmetristrials. Det er også nedgang fra første til andre testhalvdel

Tabell 2. Ulike resultater fra FET tester.

Tester	Relasjon	p verdi
Gruppe 1 vs. Gruppe 2	Symmetri	i/a
Gruppe 1 vs. Gruppe 2	Transitivitet/ekvivalens	0,1749
Gruppe 2 vs. Gruppe 3	Symmetri	0,0704
Gruppe 2 vs. Gruppe 3	Transitivitet/ekvivalens	0.3250
Gruppe 1 vs. Gruppe 3	Symmetri	0.0704
Gruppe 1 vs. Gruppe 3	Transitivitet/ekvivalens	0.1749

Merknad. i/a: ikke anvendbar ettersom verdiene er like.



Figur 4. Figuren viser gjennomsnittlig reaksjonstid for de ulike relasjoner i hver gruppe i Eksperiment. *Merknad:* Hvert punkt viser median av de fem siste trials i hver relasjon for hver av gruppene. BSL = baseline; 0-node 1 = første halvdel med symmetri trials; 1-node 1 = første halvdel med transitivitet/ekvivalens trials. 2 = andre halvdel.

for både symmetri trials og transitivitet/ekvivalens trials. Tilsvarende økning var enda mer markant for de fem første øktene med test for transitivitet/ekvivalens trials sammenlignet med de fem siste øktene under baselinetrening.

En enveis ANOVA med gjentatte målinger viste en signifikant effekt av relasjonstype for alle tre gruppene. Det innebærer altså at reaksjonstiden økte fra baselinetrials via symmetri trials og til transitivitet/ekvivalens trials. For Gruppe 1: $F(2, 8) = 8,68, p = 0.015$, for

Gruppe 2: $F(2, 9) = 10.25, p = 0.0095$ og for Gruppe 3: $F(2, 9) = 20.00, p = 0.0003$.

For transitivitet/ekvivalens trials var det en signifikant nedgang i reaksjonstider fra første til andre testhalvdel for alle gruppene. For Gruppe 1 $t(8) = 2,37, p < 0.05$, for Gruppe 2 $t(9) = 2,78, p < 0.05$ og for Gruppe 3 $t(9) = 3,89, p < 0.05$.

Oppsummering

Det var flest personer som responderte i henhold til stimulusekvivalens under betin-

gelsen med 54 trials (Gruppe 3) i treningsblokkene, sammenliknet med de andre gruppene (antall trials og ekvivalensklasser blir nærmere diskutert under Generell diskusjon). Forskjellen mellom gruppene var størst for symmetrials. Reaksjonstiden viste en økning for alle deltakerne uavhengig av gruppe fra trening til tidlig i testen og en nedgang fra tidlig til sent i testen.

Forskjellene mellom gruppenes reaksjonstidsmønstre for de ulike relasjonene er minimal. Reaksjonstidsdataene replikerer tidligere funn idet de viser at reaksjonstiden øker fra de siste fem trials under trening til de fem første testtrials for henholdsvis symmetri og transitivitetsekvivalens (Camargo & Haydu, 2015), og at reaksjonstiden for symmetri og transitivitetsekvivalens avtar fra tidlig til sent i testen (Arntzen et al., 2007; Arntzen & Holth, 1997, 2000; Eilifsen & Arntzen, 2009; Holth & Arntzen, 2000; Spencer & Chase, 1996).

Resultatet fra dette eksperimentet og tidligere funn som viser forskjell reaksjonstid for de ulike relasjonene er ikke uten videre i samsvar med Sidman (2000); prediksjoner basert på Sidman sin «bag» teori. Det er flere mulige forklaringer på dette. En mulig forklaring har vært at det foregår former verbal mediering under etablering av de betingede diskriminasjonene (e.g., Sundberg et al., 2018). Ved presentasjon av testtrials vil ikke denne medieringen uten videre være mulig ettersom disse relasjonene er nye for forsøkspersonen. Av den grunn vil reaksjonstiden øke, og den vil øke spesielt for relasjoner som inneholder en node eller flere noder. Denne økningen i reaksjonstid som en funksjon av antall noder er funnet i noen eksperimenter (Bentall et al., 1998), men ikke i andre eksperimenter (e.g., Arntzen & Hansen, 2011; Imam, 2006). Fields og medarbeider (e.g., Fields & Watanabe-Rose, 2008) har lansert en teori om *nodal distance*, som innebærer at antall forsøkspersoner som responderer i henhold til stimulusekvivalens avtar og reaksjonstid øker som en funksjon

av antall noder. (Det er å foretrekke å bruke antall noder framfor *nodal distance* ettersom den sistnevnte begrepet kan ansees som en hypotetisk struktur (Sidman, 1994)). Reaksjonstidsdataene i dette eksperimentet støtter at reaksjonstiden skulle øke som en funksjon av antall noder, men det er viktig å ta i betraktning at i dette eksperimentet er det ikke trials som inneholder mer enn en node. Derimot er en mer sannsynlig grunn for økningen i det foreliggende eksperimentet er at forsøkspersonen har tidligere sett stimulusparene (utvalgsstimulus og sammenlikningsstimulus) sammen i baseline og som presenteres i en symmetrial, derimot har forsøkspersonene ikke sett stimulusparene som presenteres i en transitivitetse- eller ekvivalens trial tidligere.

Antakelsen om at en økning av antall trials i treningsblokken skulle gi seg utslag i reaksjonstid under testen viste seg ikke å stemme. Altså at mer robuste klasser skulle gi seg utslag i kortere reaksjonstid. Det er mulig at slike effekter ikke framkom fordi det ikke var stor nok økning av antall trials i treningsblokkene, noe som bør undersøkes videre. En annen variabel som kan ha hatt innvirkning og som bør undersøkes nærmere er om det vil være forskjeller dersom det ble trent flere medlemmer i klassene.

EKSPERIMENT 2

Forskjellene i respondering i henhold til stimulusekvivalens i Eksperiment 1 kan være en effekt av både antall trials og antall programmerte konsekvenser ettersom antall trials for de tre gruppene var gjennomgående for alle treningsblokkene, med og uten programmerte konsekvenser. Hovedformålet med Eksperiment 2 var derfor å introdusere en økning av antall trials per treningsblokk slik at det ikke direkte påvirket antall programmerte konsekvenser. Av den grunn ble henholdsvis 18, 36 og 54 trials ekstra introdusert i treningsblokken uten programmerte konsekvenser.

Metode

Forsøkspersoner

Tjuefire forsøkspersoner (20–57 år; gjennomsnitt på 29 år), 16 kvinner og åtte menn, deltok i eksperimentet. Rekrutteringen og debriefingen foregikk på samme måte som i Eksperiment 1.

Setting, apparatur og programvare

Det ble benyttet eksperimentavlukker som beskrevet i Eilifsen og Arntzen (2011). Datamaskinene, programvaren og stimulien som ble benyttet i dette eksperimentet var som i Eksperiment 1.

Design

Den samme eksperimentelle designen ble benyttet som i Eksperiment 1, bortsett fra at økningen av antall trials ble introdusert i den siste blokken med tynning av programmerte konsekvenser; treningsblokken uten programmerte konsekvenser.

Prosedyre og avhengig variabel

Prosedyren og avhengig variabel er som beskrevet under trening og test i Eksperiment 1, bortsett fra antall trials i hver treningsblokk. Forsøkspersonene i alle tre gruppene hadde 18 trials i hver blokk til og med blokken med 25% sannsynlighet for programmerte konsekvenser. I blokken med uten programmerte konsekvenser var antall trials 18, 36 og 54 trials, henholdsvis for Gruppe 1, 2 og 3. Ettersom reaksjonstidmålene ikke var forskjellig mellom gruppene i Eksperiment 1 er disse ikke tatt med som en del av resultatene i Eksperiment 2.

Resultater og diskusjon

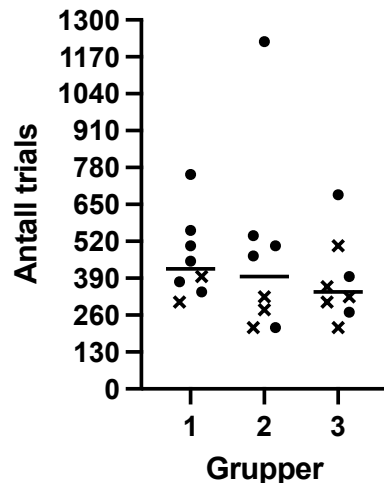
Antall trials etablering av betingede diskriminasjoner

Figur 5 viser antall trials til mestringskriterium for de tre gruppene. Det er spesielt stor variasjon innad i Gruppe 2. Det er en nedadgående tendens av gjennomsnittet fra

Gruppe 1 til Gruppe 3. Gjennomsnittet for Gruppe 1 er 423, 396 for Gruppe 2 og 342 for Gruppe 3.

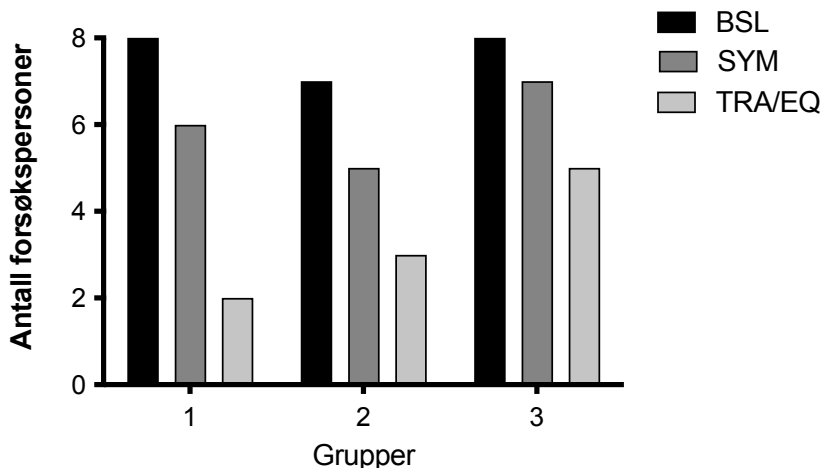
Analyser av sammenheng mellom antall trials og respondering i henhold til stimulusekvivalens

For Gruppe 1 hadde begge forsøkspersonene som responderte i henhold til stimulusekvivalens færre antall trials enn gjennomsnittet, mens for de andre seks forsøkspersonene hadde to under gjennomsnittet og fire over gjennomsnittet. For Gruppe 2 hadde alle tre forsøkspersonene som responderte i henhold til stimulusekvivalens antall trials under gjennomsnittet, mens for de fem andre



Figur 5. Antall trials er vist i figuren for hver enkelt forsøksperson i de tre gruppene i Eksperiment. *Merknad:* De vertikale strekene viser gjennomsnittet i hver gruppe. Hvert datapunkt illustrerer antall trials for den enkelte forsøksperson. De datapunktene som er markert med kryss viser de forsøkspersonene som responderte i henhold til stimulusekvivalens.

forsøkspersonene var det en under og fire over gjennomsnittet. For Gruppe 3 hadde de tre forsøkspersonene som responderte i henhold til stimulusekvivalens antall trials under gjennomsnittet, mens to hadde over gjennomsnittet. De tre forsøkspersonene



Figur 6. De tre stolpene for hver gruppe viser antall forsøkspersoner som responderte i henhold til de eksperimentator-definerte kriteriene for henholdsvis baseline (BSL), symmetri (SYM) og transitivitet/ekvivalens (TRA/EQ) i Eksperiment 2.

Tabell 3. Ulike resultater fra FET tester.

Tester	Relasjon	p verdi
Gruppe 1 vs. Gruppe 2	Symmetri	i/a
Gruppe 1 vs. Gruppe 2	Transitivitet/ekvivalens	0,1749
Gruppe 2 vs. Gruppe 3	Symmetri	0,3250
Gruppe 2 vs. Gruppe 3	Transitivitet/ekvivalens	0,3250
Gruppe 1 vs. Gruppe 3	Symmetri	i/a
Gruppe 1 vs. Gruppe 3	Transitivitet/ekvivalens	0,0349

Merknad: i/a: ikke anvendbar ettersom verdiene er like.

som ikke responderte i henhold til stimulus-ekvivalens hadde en under gjennomsnittet, mens de to siste hadde over gjennomsnittet.

Respondering i henhold til stimulus-ekvivalens

Tjue-tre av 24 forsøkspersoner på tvers av alle tre gruppene responderte korrekt på alle baselinetrials (se Figur 6). Den siste forsøkspersonen fra Gruppe 2 hadde 15 av 18 korrekte baselinetrials i testen. Seks, fem og syv av åtte forsøkspersoner, i henholdsvis Gruppe 1, 2 og 3, responderte korrekt på symmetrials. For transitivitet/ekvivalens var det en økende tendens av forsøkspersoner som responderte i henhold til eksperimentatordefinerte korrekte trials,

henholdsvis to, tre og fem for Gruppe 1, 2 og 3. Ingen av FET testene som ble brukt viste signifikante resultater bortsett fra Gruppe 1 vs. 3 i forhold til transitivitet/ekvivalens (se Tabell 3).

Oppsummering

Resultatene viser individuell variasjon i antall trials til mestringskriteriet er nådd. Det er verdt å merke seg at på tross av at Gruppe 3 hadde flest trials i treningsblokken (den uten programmerte konsekvenser) var det den gruppe som gjennomsnittlig nådde mestringskriteriet med færrest antall trials. Videre at det høyere antall trials i Gruppe 3 sammenliknet med Gruppe 1 og 2 førte til den største økningen av antall forsøks-

personer som responderte i henhold til stimulusekvivalens.

Generell diskusjon

Formålet med de to eksperimentene var å undersøke effektene av et ulikt antall trials under etableringen av de betingede diskriminasjonene på respondering i henhold til stimulusekvivalens. Eksperimentene ble arrangert i en LS treningsstruktur med en SIM protokoll. I Eksperiment 1 bestod alle treningsblokkene av henholdsvis 18, 36 og 54 trials, mens i Eksperiment 2 var det tilsvarende antall trials kun i blokken uten programmerte konsekvenser. Resultatene viste at det var gruppen som hadde 54 trials i en blokk, enten den var i alle treningsblokker (Eksperiment 1) eller kun i den blokken uten programmerte konsekvenser (Eksperiment 2), som hadde det høyeste antall forsøkspersoner som responderte i henhold til stimulusekvivalens. Reaksjonstidsdataene fra Eksperiment 1 viste et likt mønster for alle gruppene, som innebar en økning fra baseline til trials i starten på testen og at økning var betraktelig høyere for transitivitetsekvivalens-trials enn symmetri-trials og en reduksjon i løpet av testen.

Antall trials under etablering av baseline

Antall trials ble manipulert i begge eksperimentene og vil kunne direkte påvirke antall trials til mestringskriteriet er oppnådd, og spesielt i Eksperiment 1. Resultatene viser også at det er tilfelle i Eksperiment 1, hvor det er gradvis økning i gjennomsnittlig antall trials til mestringskriterium fra Gruppe 1 via Gruppe 2 til Gruppe 3, og også mer variasjon i den siste gruppen. De tilsvarende datapunktene i Eksperiment 2 viser at gjennomsnittlig antall trials til mestringskriterium er lavest for Gruppe 3 på tross av at det var 54 trials i motsetning til 36 og 18 i de to andre gruppene. Et resultat som er noe uventet selv om den uavhengige variabelen som ble manipulert var kun til stede i en treningsblokk.

De visuelle analysene som ble gjort av antall trials til mestringskriterium og hvorvidt forsøkspersonene responderte i henhold til stimulusekvivalens eller ikke, viser ikke et klart bilde for alle gruppene. Gruppe 1 i Eksperiment 1 og de tre gruppene i Eksperiment 2 viser en viss tendens for at de forsøkspersonene som formet ekvivalensklasser hadde færrest antall trials. Andre eksperimenter har vist en korrelasjon mellom antall trials under etablering av baseline og respondering i henhold til stimulusekvivalens i den forstand at forsøkspersoner med færre antall trials responderte i henhold til stimulusekvivalens (Mensah & Arntzen, 2017; Narrey et al., 2015). Antall trials til mestringskriterium og resultatet på tester for emergente relasjoner bør undersøkes nærmere ettersom resultatene i de to eksperimentene i den foreliggende studien ikke er entydige og eksperimentene er gjennomført med et forholdsvis lite antall forsøkspersoner. Videre så bør det undersøkes hvordan antall trials og feil i etableringen av de betingede diskriminasjonene henger sammen. Arrangementet i de fleste eksperimenter innen denne forskningsgrenen er slik at en eller to feil kan føre til at hele treningsblokken gjentas. Hvis det skjer flere ganger og særlig dersom det skjer mot slutten av en treningsblokk, så vil antall trials totalt sett bli høyt mens antall feil nødvendigvis ikke vil være vesentlig høyt.

Respondering i henhold til stimulusekvivalens

Resultater fra dette eksperimentet er i samsvar andre eksperimenter som har brukt abstrakte stimuli, LS og SIM protokollen og som viser et forholdsvis lavt antall forsøkspersoner som responderer i henhold til stimulusekvivalens (Arntzen, 2010; Arntzen & Hansen, 2011; Arntzen & Holth, 1997, 2000; Eilifsen & Arntzen, 2009; Saunders et al., 1993).

Som tidligere nevnt er det forholdsvis lave antall forsøkspersoner som responderer i henhold til stimulusekvivalens noe av grunnen til at LS strukturen er ideell for å studere

variabler som kan påvirke emergens av ekvivalensklasser, og da effektene av antall trials under baseline. Begge eksperimentene viste at økning av antall trials i en treningsblokk økte antall forsøkspersoner som responderte i henhold til stimulusekvivalens, og da spesielt 54 trials i en blokk. Det er viktig å ta i betraktning at for eksempel for Gruppe 3 utgjorde det en økning på 54 trials per treningsblokk over 5 treningsblokker i Eksperiment 1, mens det kun var i siste treningsblokk i Eksperiment 2. På tross av dette så var effektene at fem av 10 forsøkspersoner i Eksperiment 1 vs. fem av åtte forsøkspersoner i Eksperiment 2 responderte i henhold til stimulusekvivalens. Funnene fra de foreliggende eksperimentene støtter Saunders et al. (1999) sin antakelse om at antall trials som benyttes i etablering av de betingede diskriminasjonene kan påvirke emergens av ekvivalensklasser.

Når vi sammenlikner resultatene fra Eksperiment 1 og 2, så kan det ikke være antall programmerte konsekvenser som er avgjørende for økningen av antallet forsøkspersoner som responderer i henhold til stimulusekvivalens. Derimot er det antall trials som er introdusert i tillegg det ordinære antall i treningsblokken som ser ut til å være den avgjørende faktoren ettersom økningen av antall forsøkspersoner som responderer i henhold til stimulusekvivalens er like stor når antall trials er økt i treningsblokken uten programmerte forsterkere som i alle treningsblokkene. Flere eksperimenter bør gjøres for å kunne underbygge hvor gyldige disse resultatene er. Samtidig bør det introduseres et større antall trials i treningsblokkene for å se om forskjellene mellom et minimum antall treningstrials og et større antall treningstrials vil føre til et høyere antall forsøkspersoner som responderer i henhold til stimulusekvivalens.

Betydningen av at baselinerelasjoner er intakt i testen

Det er selvsagt en forutsetning at de betingede diskriminasjonene er etablert (baseline) før testen for emergente relasjoner iverksettes (se forøvrig Arntzen, 2012, for en

diskusjon mestringskriterium for baselinerelasjoner). Betydningen av hvor solid baseline er for emergens av ekvivalensklasser har vært hovedspørsmålet i de foreliggende eksperimentene. Videre har det vært diskutert om de tilfellene hvor ekvivalensklasser ikke har framkommet under testen om det det kan skyldes at baselinerelasjonene ikke er i intakt lengre (se for eksempel Saunders et al., 1999). En måte å undersøke hvorvidt betydningen av om baselinerelasjonene er intakt eller ikke, er som i de foreliggende eksperimentene hvor det er presentert resultater som i Figur 3 og 6. I Eksperiment 1 hadde alle forsøkspersonene i alle gruppene baselinerelasjonene intakt selv om de ikke responderte i henhold til stimulusekvivalens, mens i Eksperiment 2 hadde 13 av 14 forsøkspersoner baselinerelasjonene intakt selv om de ikke responderte i henhold til stimulusekvivalens. Dette svekker antakelsen til Saunders et al. (1999) om at en generell årsak til at forsøkspersoner ikke responderer i henhold til stimulusekvivalens er at baselinerelasjoner ikke er intakte under test.

Hvordan testblokker arrangeres for å teste for emergens avhenger av flere forhold som omtalt i innledningen og da spesielt avhengig av hvilken trening- og testprotokoll som benyttes. Et annet forhold som er diskutert ovenfor er om baseline trials er til stede i testblokken mikset med testtrials. I eksperimenter innen forskning på stimulusekvivalens er en miks av baseline- og testtrials er veldig vanlig (Saunders et al., 1999), men dette er ikke inngående undersøkt. En systematisk oversikt over publiserte studier med hensyn til om baseline trials er til stede i testblokker eller ikke vil være et viktig bidrag i utvidelsen av forståelsen av emergens ekvivalensklasser.

Begrepet overtraining

Begrepet overtraining har vært brukt i litt forskjellig betydning, både som økning av antall trials i etableringen av betingede diskriminasjoner og i en mer spesifikk betydning hvor det er en økning av antall trials etter at

mestringskriterium er nådd. Den sistnevnte bruken av begrepet er omtalt i Driskell et al. (1992) og Minster et al. (2011).

Det er utført studier med variabelen *overtraining* innenfor andre områder enn stimulusekvivalens, slik som *behavioral contrast* og *peak-shift* (Dukhayil & Lyons, 1973), *work-ethic effect* (Clement et al., 2000; Vasconcelos et al., 2007), *reversal training* (Beale, 1970; Ludvigson & Caul, 1964), *shift learning* (e.g., Nakagawa, 2001), og *retention* (Driskell et al., 1992; Rohrer & Pashler, 2007).

Overtraining eller *overlearning* har vært omtalt som en variabel som kan påvirke emergens av ekvivalensklasser (e.g., Sidman et al., 1985). Som nevnt i innledningen er det noen studier innen stimulusekvivalens som har undersøkt effektene av *overtraining* (Bortoloti et al., 2013; Minster et al., 2011; Travis et al., 2014). I studien til Bortoloti et al. (2013) er det ikke helt åpenbart hvordan arrangementet med flere trials er administrert, men beskrivelsene tyder på at et visst antall trials er lagt til for en eller flere betingelser eller grupper uten at nødvendigvis mestringskriteriet er oppnådd. Det samme er gjort i Travis et al. (2014) hvor noen av gruppene med enkel diskriminasjon i en fase med fortrening fikk henholdsvis 100 og 500 ekstra trials. I de foreliggende eksperimentene er antall trials presentert som i eksperimentene til Bortoloti et al. (2013) og Travis et al. (2014).

Arrangementet med økningen av baselinetrials i de to eksperimentene kan være problematisk selv om det som vist ovenfor ikke er uvanlig å legge til antall trials uten å sikre spesifikt at et mestringskriterium er oppnådd først. Vi vet altså ikke med sikkerhet i de to foreliggende eksperimentene om antall trials nødvendigvis økte etter at forsøkspersonene hadde oppnådd mestring eller om mestringskriterium ble oppnådd i løpet av de ekstra trials. Framfor et arrangement slik det var i disse eksperimentene bør framtidige eksperimenter trene forsøkspersonene til et mestringskriterium på 90% eller høyere for et visst antall trials i en blokk (antall trials bestemmes av antall betingede diskrimina-

sjoner som må etableres og antall ganger hver trial type presenteres). Deretter bør grupper av forsøkspersoner eller forsøkspersoner i ulike betingelser bli eksponert for blokker med et økende antall trials i tillegg. Dette vil være i tråd med *overtraining* som definert ovenfor.

Relevansen for studier innen anvendte atferdsanalyse

I en anvendt setting hvor en LS treningsstruktur skal benyttes er det viktig å tilrettelegge slik at flere forsøkspersoner framviser ekvivalensklasser. For det første vil det å øke antall trials in en treningsblokk uten programmerte konsekvenser være gunstig; mye mer tidsbesparende enn at alle treningsblokkene må ha et økt antall treningstrials. For det andre har en rekke tidligere eksperimenter vist til positive effekter av å inkludere minst en meningsfull stimulus og da som node i treningsstrukturen (e.g., Mensah & Arntzen, 2021). Et slikt arrangement vil gjøre at et betraktelig høyere antall forsøkspersoner vil respondere i henhold til stimulusekvivalens enn dersom alle stimuliene er abstrakte som i det foreliggende eksperimentet. For eksempel viste resultatene fra Mensah og Arntzen viste at antall forsøkspersoner som responderte i henhold til stimulusekvivalens økte fra cirka 10% til nærmere 80%.

Begrensninger og framtidige studier

Et eksperiment som kan avledes fra de to foreliggende eksperimentene er effektene av antall baselinetrials utover mestringskriterium og hvordan dette kan påvirke det som er betegnet som utsatt emergens (*delayed emergence*; DE) av ekvivalensklasser. DE vil være tilfeller hvor ekvivalens emergerer etter flere testtrials (e.g., Sidman et al., 1985). Flere eksperimenter har rapport om DE (en detaljert analyse av *delayed emergence* finnes i Arntzen & Mensah, 2020). Det er ulike teoretiske betraktninger hvorfor DE forekommer (se for eksempel Sidman, 1994), og en slik teori er framsatt av Spradlin et al. (1992): "As long as some key relations within the class have been tested and demon-

strated and as long as the “to be recovered” relations are presented in sessions including these intact relations, it is easy to see how recovery might occur across repeated tests under extinction conditions.” (s. 34). I de foreliggende eksperimentene så vi ikke noen resultater som pekte i retning av DE, og det er sikkert mange grunner til det, men av dem er at vi hadde arrangert et begrenset antall test trials. Altså, det bør undersøkes om et økt antall treningstrials utover mestringskriteriet vil øke tilfeller av DE.

Et annet eksperiment for å utvide resultatene fra de foreliggende eksperimentene vil være å inkludere tester for å sjekke om ekvivalensklassene er opprettholdt over tid etter at det er gjennomført en økning av trials når mestringskriterium er oppnådd. Prediksjonen vil være at der hvor antall trials i treningsblokkene er høyest vil være de tilfellene hvor ekvivalensklassene blir opprettholdt lengst. Et slikt resultat vil være i tråd med Spradlin et al. (1992) sin antakelse om at *overtraining* kan føre til en større stabilitet av ekvivalensklasser.

Oppsummering og konklusjon

Få tidligere studier har hatt som hovedfokus å undersøke hvilke effekter antall trials under etablering av baselinereelasjoner har på forsøkspersonenes respondering i henhold til stimulusekvivalens med en LS treningsstruktur og SIM protokoll. Resultatene fra de to foreliggende eksperimentene har vist at antall trials i baseline kan være en virksom variabel for emergens av ekvivalensklasser. Resultatene fra Eksperiment 1 viser at et økt antall trials, forsøkspersoner med 56 trials treningsblokkene, viste det høyeste antall forsøkspersoner som responderte i henhold til stimulusekvivalens sammenliknet med de gruppene med 36 og 18 trials i treningsblokkene. Resultatene fra Eksperiment 2 viser at det er antall trials i seg selv og ikke forskjeller i antall programmerte konsekvenser som synes å være den viktigste variabelen. Men, det kreves flere forsøkspersoner i framtidige eksperimenter før man kan gjøre sikrere

konklusjoner. Reaksjonstidsdataene fra Eksperiment 1 replikerte tidligere funn hvor den største økningen var fra baselinetrials til ekvivalensstrials. Videre så var det ingen forskjeller i reaksjonstidsmønstre for de tre gruppene, så det innebærer at effekten av antall treningstrials ikke influerte på forskjeller i reaksjonstid.

Referanser

- Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record*, 54(2), 275–291. <https://doi.org/10.1007/BF03395474>
- Arntzen, E. (2010). Om stimulusekvivalens. Teoretiske betraktninger og noen praktiske implikasjoner. I F. Svartdal & S. Eikeseth (Eds.), *Anvendt atferdsanalyse* (2 ed., s. 100–138). Gyldendal Akademika.
- Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis*, 13(1), 123–135. <https://doi.org/10.1080/15021149.2012.11434412>
- Arntzen, E., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivalence. *European Journal of Behavior Analysis*, 8(2), 177–191. <https://doi.org/10.1080/15021149.2007.11434281>
- Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis*, 12(2), 483–503. <https://doi.org/10.1080/15021149.2011.11434397>
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, 47(2), 309–320. <https://doi.org/10.1007/BF03395227>
- Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record*, 50(4), 603–628. <https://doi.org/10.1007/BF03395227>

- [org/10.1007/BF03395374](https://doi.org/10.1007/BF03395374)
- Arntzen, E., & Mensah, J. (2020). On the effectiveness of including meaningful pictures in the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *113*(2), 305–321. <https://doi.org/10.1002/jeab.579>
- Ayres-Pereira, V., & Arntzen, E. (2019). Emergence of large equivalence classes as function of training structures. *Mexican Journal of Behavior Analysis*, *45*(1), 20–47. <https://doi.org/dx.doi.org/10.5514/rmac.v45.i1.70864>
- Beale, I. L. (1970). The effects of amount of training per reversal on successive reversals of a color discrimination. *Journal of the Experimental Analysis of Behavior*, *14*(3), 345–352. <https://doi.org/10.1901/jeab.1970.14-345>
- Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *46B*(2), 187–214. <https://doi.org/10.1080/14640749308401085>
- Bentall, R. P., Jones, R. M., & Dickins, D. W. (1998). Errors and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record*, *49*(1), 93–115. <http://thepsychologicalrecord.siuc.edu/index.html>
- Bortoloti, R., Rodrigues, N. C., Cortez, M. D., Pimentel, N., & de Rose, J. C. (2013). Overtraining increases the strength of equivalence relations. *Psychology & Neuroscience*, *6*(3), 357–364. <https://doi.org/10.3922/j.psns.2013.3.13>
- Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *The Psychological Record*, *47*(1), 69–96. <https://doi.org/10.1007/BF03395213>
- Camargo, J., & Haydu, V. B. (2015). Class size effects on the establishing, maintenance, and recovery of equivalence relations. *Behavior Analysis: Research and Practice*, *15*(1), 36–49. <https://doi.org/10.1037/h0101068>
- Clement, T. S., Feltus, J. R., Kaiser, D. H., & Zentall, T. R. (2000). “Work ethic” in pigeons: Reward value is directly related to the effort or time required to obtain the reward. *Psychonomic Bulletin & Review*, *7*(1), 100–106. <https://doi.org/10.3758/BF03210727>
- Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*(5), 615–622. <https://doi.org/10.1037//0021-9010.77.5.615>
- Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis. I T. R. Zentall & P. M. Smeets (Eds.), *Stimulus Class Formation in Humans and Animals* (s. 197–218). Elsevier.
- Dukhayyil, A., & Lyons, J. E. (1973). The effect of overtraining on behavioral contrast and peak shift. *Journal of the Experimental Analysis of Behavior*, *20*(2), 253–263. <https://doi.org/10.1901/jeab.1973.20-253>
- Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, *19*(1), 8–12. <https://www.eahb.org/bulletin2001>
- Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis*, *10*(2), 187–202. <https://doi.org/10.1080/15021149.2009.11434318>
- Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to-sample procedures. *European Journal of Behavior Analysis*, *12*(1), 152–172. <https://doi.org/10.1080/15021149.2011.11434361>
- Fields, L., Arntzen, E., & Doran, E. (2020). Yield as an essential measure of equivalence class formation, other measures, and new determinants. *The Psychological Record*, *70*(2), 175–186. <https://doi.org/10.1007/s40732-020-00377-3>

- Fields, L., & Watanabe-Rose, M. (2008). Nodal structure and the partitioning of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 89(3), 359–381. <https://doi.org/10.1901/jeab.2008-89-359>
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. I K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (s. 229–262). Plenum Press.
- Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurent responding? *The Psychological Record*, 50(3), 305–338. <http://the-psychological-record.siuc.edu/index.html>
- Hrbacek, K., & Jech, T. (1999). *Introduction to mathematical set theory* (3rd ed.). Marcel Decker, Inc.
- Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior*, 85(1), 107–124. <https://doi.org/10.1901/jeab.2006.58-04>
- Ludvigson, H. W., & Caul, W. F. (1964). Relative effect of overlearning on reversal and nonreversal shifts with two and four sorting categories. *Journal of Experimental Psychology*, 68, 301–306. <https://doi.org/10.1037/h0042254>
- Mensah, J., & Arntzen, E. (2017). Effects of meaningful stimuli contained in different numbers of classes on equivalence class formation. *The Psychological Record*, 67(3), 325–336. <https://doi.org/10.1007/s40732-016-0215-y>
- Mensah, J., & Arntzen, E. (2021). The Role of Meaningful Stimuli in Large Stimulus Classes. *The Psychological Record*. <https://doi.org/10.1007/s40732-021-00483-w>
- Minster, T. S., Elliffe, D., & Muthukumaraswamy, S. D. (2011). Emergent stimulus relations depend on stimulus correlations and not on reinforcement contingencies. *Journal of the Experimental Analysis of Behavior*, 95(3), 327–342. <https://doi.org/10.1901/jeab.2011.95-327>
- Nakagawa, E. (2001). Effects of overtraining on shift learning in matching-(or nonmatching)-to-sample discrimination in rats. *The Psychological Record*, 51(4), 473–493. <http://the-psychological-record.siuc.edu/index.html>
- Nartey, R. K., Arntzen, E., & Fields, L. (2015). Enhancement of equivalence class formation by pretraining discriminative functions. *Learning & Behavior*, 43(1), 20–31. <https://doi.org/10.3758/s13420-014-0158-6>
- Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, 16(4), 183–186.
- Sadeghi, P., & Arntzen, E. (2018). Eye-movements, training structures, and equivalence class formation. *The Psychological Record*, 68(4), 461–476. <https://doi.org/10.1007/s40732-018-0290-3>
- Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record*, 43(4), 725–744. <https://doi.org/https://doi.org/10.1007/BF03395909>
- Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior*, 71(2), 195–214. <https://doi.org/10.1901/jeab.1999.71-195>
- Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior*, 72(1), 117–137. <https://doi.org/10.1901/jeab.1999.72-117>
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Authors Cooperative.

- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74(1), 127–146. <https://doi.org/10.1901/jeab.2000.74-127>
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-members stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 43(1), 21–42. <https://doi.org/10.1901/jeab.1985.43-21>
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5–22. <https://doi.org/10.1901/jeab.1982.37-5>
- Spencer, T. J., & Chase, P. N. (1996). Speed analysis of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 65(3), 643–659. <https://doi.org/10.1901/jeab.1996.65-643>
- Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. I S. C. Hayes & L. J. Hayes (Eds.), *Understanding Verbal Relations* (s. 29–42). Context Press.
- Sundberg, C. T., Sundberg, M. L., & Michael, J. (2018). Covert verbal mediation in arbitrary matching to sample. *Journal of the Experimental Analysis of Behavior*, 109(3), 600–623. <https://doi.org/10.1002/jeab.434>
- Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of the Experimental Analysis of Behavior*, 102(1), 47–65. <https://doi.org/10.1002/jeab.91>
- Vasconcelos, M., Urcuioli, P. J., & Lionello-DeNolf, K. M. (2007). Failure to replicate the “work ethic” effect in pigeons. *Journal of the Experimental Analysis of Behavior*, 87(3), 383–399. <https://doi.org/10.1901/jeab.2007.68-06>
-