



LUNDS
UNIVERSITET

INSTITUTIONEN FÖR PSYKOLOGI

*Korrektion av implicit ingrupsfavorisering och
utgrupphostilitet*

Anna Jansson
Jonna Sjölin

Lunds Universitet Kandidatuppsats ht 2010

Handledare: Martin Bäckström
Öyvind Jørgensen

Abstract

The aim of the study was to investigate correction regarding ingroup favoritism and outgroup hostility on an implicit level. Current research revealed that people tend to make amendments for ingroup bias in the presence of an outgroup member. In order to test correction of bias implicitly, the participants received false feedback regarding their test results on an implicit test that measured racial bias. One group received false feedback about showing ingroup favoritism towards Swedes, and one group received false feedback about showing outgroup hostility towards Arabs. The reason for using two different kinds of feedback was to investigate if the participants were able to distinguish between ingroup favoritism and outgroup hostility. The predictions were that participants who received false feedback regarding ingroup favoritism should correct for ingroup bias and the group that received feedback regarding outgroup hostility should correct for outgroup bias, but the results did not support the hypotheses. However, the results revealed that individuals that were externally motivated to control for prejudice did correct for outgroup hostility. The results are discussed in the light of previous research.

Keywords: Ingroup, Outgroup, favoritism, correction

Abstract

Syftet med föreliggande studie var att undersöka korrektion av ingrupsfavorisering och utgrupphostilitet på implicit nivå. Forskning har visat att människor korrigerar för ingrupsbias i närvaron av utgruppsmedlemmar. Denna studie testade korrektion av etnisk bias under ett rasrelaterat implicit test genom att ge två experimentgrupper falsk feedback. En av grupperna fick veta att de hade stark automatisk positivitet till svenskar och den andra gruppen fick veta att de hade en stark automatisk negativitet mot araber. Olika former av falsk feedback användes för att undersöka om deltagarna reagerade olika på påstådd ingrupsfavorisering och utgrupphostilitet. Det predicerades att grupperna skulle korrigera för ingrupsbias eller utgrupphostilitet beroende på vilken feedback de fick, men resultaten visade inget stöd för hypoteserna. Däremot så uppstod en korrektionseffekt så att individer med hög extern motivation att kontrollera fördomsfullt beteende fick en mer positiv bias till utgruppen. Resultaten diskuteras i förhållande till tidigare forskningsfynd.

Nyckelord: Ingrupp, utgrupp, favorisering, korrektion

Författarna till denna uppsats vill tacka sina handledare, Martin Bäckström och Öyvind Jörgensen, för deras assistans med kunskap, stöd, tillgänglighet och framför allt tålamod. Deras hjälp har för oss varit ovärderlig.

Korrektion av Implicit Ingrupsfavorisering och Utgrups hostility

Syftet med denna studie var att undersöka hur falsk feedback angående bias mot in- eller utgruppen påverkade resultatet av ett implicit attitydtest. Hypoteserna var att individer som informerades om att de favoriserade sin ingrupp skulle korrigera sin implicita attityd mot ingruppen, och att personerna som informerades om att de uppvisade utgrups negativitet skulle korrigera sin implicita attityd mot utgruppen. Då experimentet utfördes i ett laboratorium och feedbacken inte var sann kan resultaten vara svåra att generalisera utanför denna specifika laboratoriemiljö. Trots detta kan denna typ av forskning ge kunskap om hur individer hanterar eventuell skillnad mellan sina värderingar och sitt beteende, och hur olika sammanhang, exempelvis information om ingrupsfavorisering eller utgrups hostility, påverkar korrigering av attityder.

Inledningsvis kommer relevant forskning angående attitydkorrigering att kartläggas. Det centrala kommer vara skillnaden mellan ingrupsfavorisering och utgrups hostility, och olika faktorer som påverkar försök till attitydkorrigering. Sambandet mellan automatiska och kontrollerade processer kommer också beröras. Efter denna litteraturöversikt kommer en mer detaljerad beskrivning av den aktuella studien.

Ingrupsfavorisering och Social Identity Theory

Brewer (1999) kritiserar idén att ingrupsfavorisering alltid följs av en negativ attityd mot utgruppen, och menar att så inte alltid är fallet. Positiv bias mot ingruppen, exempelvis känslor som stolthet och föreställningar att gruppen är överordnad innebär inte att individer automatiskt har negativa attityder mot andra grupper. Snarare uttrycks en stark preferens för sin egen grupp. Favorisering av sin ingrupp kan dock vara ett stort problem. Bakgrunden till diskriminering kan vara en relativ avsaknad av positiva attityder mot utgruppen, exempelvis beundran. Exempelvis är ett vänligt bemötande av en ingrupsmedlem mer förväntat än vad det är hos andra grupper. Detta fenomen förstärks då individer generellt föredrar det som är bekant sedan tidigare, i detta fall sin ingrupp. Diskriminering kan även grunda sig i motivationen att bibehålla en positiv association till ingruppen, snarare än att vålla utgruppen skada. Dock kan preferenser för ingruppen, exempelvis föreställningar om ingruppen som moraliskt överordnad leda till negativa attityder mot utgruppen (Brewer, 1999). Ingrupsbias kan vinnas genom att favorisera ingruppen eller diskriminera utgruppen (Mummendey & Schreiber, 1983). Brewer (1999) menar att grupper som upplever att de har en hög status är mer fokuserade på att upprätthålla sin självkänsla. Grupper med lägre status är mer fokuserade på att degradera utgruppen, detta för att minska skillnaden mellan in- och utgrupp.

Enligt *The Social Identity Theory* (Reynolds, Turner & Haslam, 2000) kan individens positiva självbild bygga på ingruppens status. Individen kan skapa hög status för sin ingrupp genom att associera sin grupp till något prestigefyllt och unikt. Motivationen till en positiv självbild kan leda till ingrupsfavorisering, då individens drivkraft att framställa ingruppen som positiv kan leda till en orättvis utvärdering av ingruppen i jämförelse andra grupper. En avgörande faktor gällande individens benägenhet till ingrupsbias är hur starkt självbilden är knuten till gruppen. Vid stark association blir individens sociala kategorisering av ingruppstillhörighet viktig och motivationen att ingruppen skall framstå som positiv är större.

Förutom association med en grupp som besitter hög status kan individen skapa en positiv självbild genom att ha värderingar som anses vara av hög moral. Normer kring rättvisa kan resultera i att diskriminering av utgrupper går att undvika så länge det finns andra sätt att upprätthålla ingruppens status och därmed individens välbefinnande. Mummendey och Schreiber (1983) testade hur benägna personer är att undvika favorisering av ingruppen på bekostnad av utgruppen genom att låta försöksdeltagare utvärdera sin egen och andra gruppers prestation på en uppgift. Ingruppens prestation värderades som positiv i alla betingelser. En mindre positiv värdering av utgruppen kunde enbart påvisas om deltagarna var instruerade att utvärdera in- och utgruppens prestationer efter samma kriterier. Mummendey och Schreiber menade att en förklaring till detta var att grupperna var tvingade att utvärdera ingruppen mer positivt än utgruppen, för att bibehålla känslan av unikheter i det fallet utvärderingen grundade sig på samma kriterier. I betingelsen där utvärderingen skedde efter olika kriterier behövde deltagarna inte tävla med de andra grupperna på samma sätt. Då känslan av unikheter inte var hotad valde deltagarna att tillämpa normen om rättvisa, och därmed inte favorisera sin egen grupp i utvärderingen av prestationen.

Attityder som automatiska och kontrollerade processer

Kulturella stereotyper är en gemensam association människor har om en grupp, och associationerna aktiveras per automatik vid attitydobjektets (exempelvis invandraren) närvaro hos både högt och lågt fördomsfulla individer (Devine, 1989). Enligt *Motivation and Opportunity as Determinants* (MODE) teorin (Olson & Fazio, 2009) påverkar attityder vårt beteende genom både automatiska och kontrollerade processer. Skillnaden mellan dessa två processer ligger i om det är en automatiskt aktiverad reaktion till attitydobjektet eller om det är en kognitivt medveten handling. Vid en kognitivt medveten handling överväger individen ett visst beteendes fördelar och nackdelar. Våra beteenden kan även påverkas av våra attityder på en automatisk, implicit nivå. Exempelvis vid social interaktion med attitydobjektet aktiveras attityden automatiskt, exempelvis i form av värderingar och förväntningar. Detta

kan i sig påverka beteendet utan att individen överväger sina handlingar på en medveten nivå. En automatiskt aktiverad attityd kan skapa en förväntning som kan påverka tolkningen av situationen, i det här fallet den sociala interaktionen med attitydobjektet. Vid en sådan situation är beteendet en direkt följd av den automatiskt aktiverade attityden. Olson och Fazio menar att automatiska och kontrollerade processer kombineras i samband med att individen försöker utöva kontroll över en automatiserad attityd.

Enligt MODE-modellen kan individen korrigera ett automatiserat attitydbeteende, förutsatt att möjligheten och motivationen finns. Det är en ren nödvändighet att individen är motiverad, all form av kognitiv bearbetning som tar upp arbetsminneskapacitet kräver att det finns en form av motivation. Motivationen att korrigera sitt attitydbeteende behöver inte vila på samma grund. När det gäller attityder som kan uppfattas som fördomsfulla kan både inre processer, till exempel att individen inte vill vara fördomsfull på grund av inre övertygelse, samt yttre processer till exempel att inte uppfattas som fördomsfull av allmänheten, leda till en korrigeringsprocess (Olson & Fazio, 2009).

Utöver motivation måste individen ha möjlighet, exempelvis tid och resurser, för att kunna korrigera ett automatiserat attitydbeteende (Olson & Fazio, 2009). Detta innebär att om människor är under tidspress eller om de inte vet hur de kan påverka något blir möjligheten att korrigera beteendet begränsat. Därför krävs både möjlighet och motivation för att påverka attitydbeteende.

Fazio, Jackson, Dunton och Williams (1995) undersökte motivationen att kontrollera fördomsfulla reaktioner, samt huruvida automatiska och kontrollerade processer kombinerades vid rasrelaterade attityder. Deltagarna fick fylla i skalan *Motivation to Control Prejudiced Reactions* (Dunton & Fazio, 1997) som mäter motivationen att kontrollera för fördomsfullhet och *Modern Racism Scale* (McConahay, Hardee & Batts, 1981), ett frågeformulär som syftar till att mäta medvetna subtila rasistiska tendenser. De fick även göra ett implicit attitydtest, där de primades med bilder av svarta personers ansikten, för att sedan kategorisera positiva och negativa ord. Detta för att se om primingen påverkade responstiden på så sätt att deltagarna associerade de svarta personernas ansikten med positivt eller negativt knutna ord. Individer med höga poäng på *Motivation to Control Prejudiced Reactions* uppvisade en inkongruens mellan *Modern Racism Scale* och det implicita attitydtestet till skillnad från personer med låga poäng på *Motivation to Control Prejudiced Reactions*, vars resultat på det explicita frågeformuläret *Modern Racism Scale* och det implicita primingtestet var samstämmiga. De högt motiverade individerna rapporterade mindre fördomar på *Modern Racism Scale* än de som uppvisade lite motivation att kontrollera för fördomsfullhet. Fazio et

al. (1995) tolkade dessa resultat som evidens för att motivation är en viktig komponent för att kontrollera fördomsfulla reaktioner. Det uppstod en interaktionseffekt mellan de explicita kontrollerade attityderna och de implicita automatiserade attityderna för individer som uppvisade hög motivation till att kontrollera fördomar. Olson och Fazio (2009) menar att kombinationen av automatiska och kontrollerade processer visar sig ofta i form av interaktion mellan individens implicita och explicita attityder.

Automatisk stereotypaktivering och *cues for control*

Monteith, Ashburn-Nardo, Voils och Czopp (2002) menar att vid stereotypaktivering och användning av stereotyper kan individen uppleva en diskrepans mellan den automatiska stereotypen och medvetna, kontrollerade åsikter. När individen är medveten om denna diskrepans följer känslan av skuld, reflektion och beteendehinhibering av användandet av stereotyper. Detta i sig leder till att individen etablerar *cues for control*, en association mellan den diskrepanta responsen och känslan av skuld. Denna association, eller ledtråd, hjälper individen att fokusera på en icke stereotypkongruent respons genom att hämma den automatiska stereotypaktiveringen.

Monteith et al. (2002) hävdar att utvecklandet av dessa *cues for control* är nödvändigt för att kunna kontrollera de automatiska processer som används vid stereotypaktivering. Vidare menar de att även under en kognitivt krävande uppgift, vilket begränsar individens möjlighet till kontroll, är det möjligt att kontrollera automatisk stereotypaktivering. Detta eftersom *cues for control* kommer att rikta individens uppmärksamhet från den automatiska stereotypen, vilket leder till stereotypinhibering.

För att testa om etablerandet av *cues of control* leder till beteendehinhibering lät Monteith et al. (2002) vita försöksdeltagare utföra ett *Implicit Association Test* (Greenwald, McGhee & Schwartz, 1998), ett implicit attitydtest som mäter associationerna mellan sociala kategorier och valens. Efter testet fick deltagarna veta resultatet, för att sedan skatta hur de emotionellt upplevde sitt resultat på det implicita testet. Sedan utfördes ytterligare ett test där försöksdeltagarna placerade olika ord och namn i kategorierna omtyckt och inte omtyckt. Hypotesen var att personer som uppvisade negativ affekt på grund av resultatet från *Implicit Association Test* skulle uppvisa beteendehinhibering genom att kategorisera namnen associerade till svarta personer mer positivt i jämförelse med personerna som inte uppvisade lika mycket negativ affekt. Resultatet visade att deltagare som hade starkare negativ bias gentemot svarta personer under *Implicit Association Test* uppvisade mer negativ affekt. Det visade sig även finnas ett samband mellan negativ affekt och gillandet av svarta personers namn. Monteith et al. tolkade resultatet som att människor kan lära sig att reglera stereotyp

beteende med hjälp av beteendehibering, som en följd av skuldkänslor (Monteith et al. 2002).

Intern och extern motivation att kontrollera fördomar

Beteendehibering kan vara en produkt av skuldkänslor, det är dock mycket svårt att kontrollera för exempelvis rasstereotyper (Monteith et al., 2002). Enligt *Self-Determination Theory* (Deci & Ryan, 2000) så kan internt reglerade koncept bidra till ett mer konsekvent beteende än de som är externt reglerade (exempelvis social press). Ifall ett koncept är internaliserat bidrar detta till en ökad stabilitet i beteende över tid och olika miljöer. Detta är relevant i relation till MODE-teorin som menar att det är nödvändigt att individen är motiverad (Olson & Fazio, 2009). Enligt *Self-Determination Theory* måste dock motivationen komma från ett internaliserat ideal om beteendet skall vara stabilt (Deci & Ryan, 2000).

Det är en stor utmaning att försöka kartlägga den bakomliggande motivationen till att inte uttrycka fördomar. Exempelvis är det svårt att fastställa om en fördomsfri respons är ett tecken på individens attityder eller ett uttryck för vad som anses vara samhällets normer. Om en person bryter mot vad som anses vara ett korrekt beteende enligt allmänheten finns det en risk för konsekvenser i form av icke godkännande (Plant & Devine, 1998). Då det inte är socialt accepterat att uttrycka exempelvis rasistiska åsikter, kan detta bidra till en subtilare form av rasism (Crosby, Bromley & Saxe, 1980).

Plant och Devine (1998) konstruerade *internal motivation to respond without prejudice scale* (IMS) och *external motivation to respond without prejudice scale* (EMS), vars syfte är att särskilja intern och extern motivation att vara fördomsfri. Intern motivation definieras som att personens drivkraft bottnar i en internalisering av ett fördomsfritt ideal, medan extern motivation att kontrollera fördomar bottnar i en drivkraft att uppvisa ett fördomsfritt beteende inför andra oavsett individens attityder.

Genom att jämföra IMS och EMS med andra fördomsrelaterade självskattningsformulär drog Plant och Devine (1998) slutsatsen att det fanns en tydlig skillnad mellan vad de båda formulären mäter. Personer som uppvisade låg fördomsfullhet på exempelvis *Modern Racism Scale* fick höga poäng på IMS. Till motsats från IMS korrelerade EMS med mer fördomsfullhet, men bara när försöksdeltagarna var anonyma. Skalorna korrelerade även negativt med varandra, vilket också bekräftade idén om att de mäter två olika former av motivation att kontrollera fördomar.

I relation till *Self-Determination Theory* hävdar Butz och Plant (2009) att personer med höga poäng på IMS och låga poäng på EMS borde mer konsekvent kontrollera för fördomsfullt beteende, även i situationer utan social press. Hög poäng på EMS borde i sin tur

bidra till lägre internalisering av ett fördomsfritt ideal i jämförelse med höga poäng på IMS. Detta borde innebära mindre konsekvent kontroll av fördomsfullt beteende.

Plant och Devine (1998) bekräftade detta genom att låta personer med olika resultat på EMS och IMS rapportera karaktärsdrag för svarta personer, antingen anonymt på ett frågeformulär eller till en försöksledare. Vissa karaktärsdrag var stereotypkongruenta (exempelvis atletiska, aggressiva). Personer med höga poäng på IMS rapporterade lite stereotyper i alla betingelser. Personer med höga poäng på EMS och låga på IMS rapporterade mer stereotyper i den privata betingelsen, i jämförelse med den offentliga. Gruppen med låga poäng på både IMS och EMS rapporterade stereotyper i både den offentliga och privata betingelsen.

Då implicita attityder är svåra att kontrollera menar Butz och Plant (2009) att det är omöjligt att kontrollera för detta om individen inte har ett internaliserat ideal som är samstämmigt med attityden. I relation till *Self-Determination Theory* bör detta innebära att personer med hög intern motivation att kontrollera fördomsfullt beteende kan kontrollera implicita och explicita attityder, personer med både hög intern och extern motivation bör ha möjlighet att kontrollera de explicita, men inte de implicita attityderna. Detta för att de inte har internaliserat en fördomsfri attityd i samma utsträckning som den förstnämnda gruppen.

Devine, Plant, Amodio, Harmon-Jones och Vance (2002) visade att personer med höga poäng på IMS och låga på EMS uppvisade minst bias mot utgruppen under *Implicit Association Test* i jämförelse med andra kombinationer av resultat från IMS och EMS, även när författarna manipulerade testet för att göra det mer kognitivt krävande. Författarna spekulerade i att detta kan bero på att de internt motiverade deltagarna hade försvagat sina negativa associationer med utgruppen, och därmed aktiverades inte stereotyperna. Detta är relevant i relation till Monteiths et al. (2002) teori om *cues of control* som hävdar att ledtrådar drar uppmärksamheten ifrån den automatiska stereotypen, vilket leder till stereotypinhibering och därmed en mindre fördomsfull respons.

Plant och Devine (1998) menar att många teorier idag handlar om samspelet mellan automatiska och kontrollerade processer i relation till kontroll av fördomar, och att IMS och EMS då kan vara värdefulla begrepp som handlar om källan till motivationen att kontrollera fördomsfulla responser. Att få förståelse för dessa begrepp kan ha stora teoretiska och empiriska fördelar gällande utvärderingen av hur olika kontrollmekanismer uttrycks beroende på om motivationen är intern eller extern.

Ingrupsfavorisering i kontextuella sammanhang

Fazio et al. (1995) visade i en studie att vita personer tenderar att rapportera mindre fördomsfulla responser på *Modern Racism Scale* när försöksledaren var svart, i jämförelse med när försöksledaren är vit. Björklund, Bäckström och Jörgensen (2010) visade att människor tenderar att skatta sin ingrupp mer positivt när denna skattning ges verbalt till en medlem av deras ingrupp trots att flertalet av deltagarna ansåg att kontexten inte påverkade dem. Ingrupsfavorisering beror inte nödvändigtvis på att människor har negativa attityder till utgruppen, utan kan grunda sig i en stark preferens för ingruppen (Brewer, 1999). I ett experiment har svenskar skattat attribut gällande ingruppen svenskar antingen verbalt eller skriftligt med en svensk eller palestinsk försöksledare. Resultaten visade att svenskar skattade ingruppen svenskar mer positivt i verbal skattning till svensk försöksledare. Samma resultat uppkom vid replikering av experimentet med en iransk försöksledare (Björklund et al., 2010). Björklund et al. tolkade resultatet som att det är möjligt att individerna förändrade sin ingruppskattning beroende på kontexten och att vid interaktion med en utgruppsmedlem skulle individen vara benägen att skatta sin ingrupp mindre positivt och vid interaktion med ingruppsmedlem så skulle detta leda till mer ingrupsfavorisering.

Beroende på om den underliggande motivationen att kontrollera för fördomsfullhet är intern eller extern kan den tillämpas på olika sätt (Butz & Plant, 2009; Plant & Devine, 1998). Björklund et al. (2010) menar att det är möjligt att människor förändrar sin ingruppskattning beroende på kontexten. Med hänvisning till Devine (1998) verkar det rimligt att personer med höga poäng på EMS uppvisar inkonsekvent beteende gällande kontroll av fördomar beroende på vilken miljö de befinner sig i.

Implicit Association Test och Sorting Paired Features Task

I följande stycke kommer *Implicit Association Test* (Greenwald et al., 1998) och *Sorting Paired Features Task* (SPF) (Bar-Anan, Nosek & Vianello, 2009) beskrivas. *Implicit Association Test* mäter den automatiska utvärderingen av stimuli (Greenwald et al., 1998). Försöksdeltagarna blir introducerade till sociala kategorier (exempelvis svarta och vita personer) som är ihop parade med ett attribut (exempelvis positivt eller negativt), för att sedan få instruktioner om att så snabbt som möjligt trycka på en viss tangent när en viss kategori eller ett visst ord (exempelvis en tangent för svarta personer och positiva ord och en tangent för vita personer och negativa ord) visas. Associationen mellan den sociala kategorin (exempelvis svarta) och ett visst attribut (exempelvis positiv) utvärderas genom att mäta responstiden. Om individen exempelvis har en stark positiv association med svarta personer kommer de reagera och trycka snabbt på tangenten för svarta personer och positiv.

SPF (Bar-Anan et al., 2009) mäter också associationer mellan sociala kategorier och valens. Den grundläggande logiken är den samma som för *Implicit Association Test*, ju snabbare kategorisering av två olika stimuli, desto starkare association är det mellan dessa. Till skillnad från *Implicit Association Test* tillåter SPF 4 olika kategoriseringsmöjligheter samtidigt. Varje kategori är ett par som består av en social kategori (exempelvis svarta) och valens (exempelvis positiv). Det finns en tangent för varje kategori, och försöksdeltagarens uppgift är att trycka på den tangent, vars kategoripar korresponderar med det kategoripar som visas på skärmen. Bar-Anan et al. menar att med tanke på att kategorisering av bilder och ord görs samtidigt kan ett slags stroopeffekt uppstå ifall individen inte automatiskt associerar ordet med bilden. Detta skulle innebära att det blir en större kognitiv börda att kategorisera bilden till ordet som individen inte har en automatisk association till och därmed tar det längre tid.

Implicit Association Test (Greenwald et al., 1998) mäter den relativa preferensen genom att associera två objekt (exempelvis svenskar och invandrare) med två attribut (exempelvis ord med positiv eller negativ valens). Om preferensen för svenskar framför invandrare skall mätas så får man ett mått på om svenskar föredras. Till skillnad från *Implicit Association Test* mäter SPF positiva och negativa associationer till objekten isolerat, då det finns fyra kategoriseringsmöjligheter samtidigt (arab/positiv, arab/negativ, svensk/positiv och svensk/negativ). Detta gör det möjligt att uppskatta om en viss social kategori är relaterad till positiv affekt, negativ affekt eller båda (Bar-Anan et al., 2009).

I en nätbaserad studie visade Bar-Anan et al. (2009) att personer tenderade att uppvisa en snabbare kategorisering på SPF för vita personer och positiva ord i jämförelse med vita personer och negativa ord. Försöksdeltagarna kategoriserade även svarta personer och negativa ord snabbare än svarta personer och positiva ord. När man skiljde på svarta och vita försöksdeltagare uppvisade båda grupperna ingrupsfavorisering vilket gav stöd åt *Social Identity Theory*. Gällande självskattningsformulären korrelerade kategorin vita personer och positiva ord positivt med resultaten från frågeformulären gällande preferenser för vita personer, och associationen mellan svarta personer och positiva ord korrelerade negativt med preferenser för vita personer under den explicita självskattningen. Bar-Anan et al. menade att detta var evidens för att SPF korrelerade med de explicita attityder som självskattningstest mäter. Bar-Anan et al. hävdar att SPF kan ge ett nyanserat resultat mellan etnicitet och valens då ingrupsfavorisering och utgrupphostilitet inte behöver vara korrelerat.

Kontroll av automatiska responser

Implicit Association Test mäter människors underliggande attityder, det har länge ansetts att dessa attityder inte varit tillgängliga för medveten kontroll. Blair (2002) visade i en review att det finns mycket som tyder på att implicita responser kan påverkas av bland annat uppmärksamhet, individens motiv, samt hur stimuli presenteras under de implicita testen. Lowery, Hardin och Sinclair (2001) testade automatisk negativitet emot svarta med hjälp av en pappersversion (pen and pencil) av *Implicit Association Test*, där försöksdeltagarna blir ombedda att med en penna markera kategoritillhörighet på ett papper så snabbt som möjligt istället för att använda sig av tangenterna på datorn. Resultatet visade att försöksdeltagarna uppvisade mindre automatisk negativitet till svarta personer om försöksledaren var svart i jämförelse med om försöksledaren var vit. Detta skulle kunna vara en form av social önskvärdhet som gjorde att deltagarna korrigerade sina responser. Lowery et al. tolkade det däremot som att deltagarna uppvisade något som kallas för *social tuning*. Detta innebär att människor försöker hitta något gemensamt med utgruppsmedlemmen genom att justera sitt eget perspektiv så att ens egna attityder korresponderar med de attityder som utgruppsmedlemmen antas ha. Detta är intressant i förhållande till forskningen av Fazio et al. (1995) där individer uppvisade mindre explicit rasism på *Modern Racism Scale* när försöksledaren var svart än när försöksledaren var vit, och Björklund et al. (2010) kom fram till att personer uppvisade mindre ingrupsfavorisering när försöksledaren var en utgruppsmedlem. Om liknande resultat kan uppnås vid implicit testning, kan detta vara en indikation på att det är möjligt att påverka automatiska responser.

Det har även visat sig att så enkla instruktioner som att be deltagare att vara mer fördomsfulla vid rasrelaterat *Implicit Association Test* har lett till ökad ingrupsfavorisering och att be försöksdeltagarna att vara mindre fördomsfulla har lett till minskad ingrupsfavorisering (Wallaert, Ward & Mann, 2010). Eftersom sådana instruktioner har visat sig kunna påverka en automatisk process kan det vara så att automatiska stereotyper delvis kan kontrolleras, och att det som mäts vid implicita attitydtest består av en blandning av automatiska och kontrollerade processer precis som Olson och Fazio (2009) hävdar.

Sammanfattningsvis tenderar människor att favorisera sina ingrupper över utgrupper (Brewer, 1999). Kontexten verkar dock styra i vilken utsträckning individer favoriserar sin ingrupp (Björklund et al., 2010). Om människor upplever negativ affekt vid uppvisandet av rasstereotypiskt beteende (Monteith et al., 2002) kan detta grunda sig i olika former av motivation till att inte uppvisa fördomsfullt beteende (Plant & Devine, 1998). Detta kan

fungera som en drivkraft att korrigera för diskrepansen mellan uppvisat och önskat beteende, förutsatt att möjligheten finns (Fazio et al., 1995).

Trots tendensen till favorisering av ingruppen kan det anses att det finns normer i dagens Sverige mot diskriminering av andra grupper. Därför framställdes hypoteserna att om individer blev medvetna om att de uppvisade ingrupsfavorisering skulle de korrigera för denna bias och om de blev medvetna om att de uppvisade utgruppshostilitet skulle de korrigera för utgruppsnegativitet. Då tidigare forskning har visat en förmåga att korrigera resultaten från *Implicit Association Test* (Lowery et al., 2001; Wallaert et al., 2010) var det rimligt att anta att korrigering även är möjligt för SPF.

För att undersöka om information gällande uppvisande av ingrupsfavorisering eller utgruppshostilitet ledde till korrigering av attityder fick försöksdeltagarna falsk feedback under SPF. Efter feedbacken fick de göra om testet.

Det predicerades att försöksdeltagare som fått feedback om att de uppvisade en stark automatisk positivitet till svenskar skulle korrigera för denna påstådda bias vid andra omgången av SPF genom en långsammare respons till svensk med positiva ord medan övriga responsmönster inte skulle påverkas. Vidare predicerades att deltagare som fått feedback att de uppvisade stark automatisk negativitet mot araber skulle korrigera för denna påstådda bias i andra omgången av SPF genom ett långsammare responsmönster till arab med negativa ord, samt att övriga responsmönster inte skulle påverkas. Det förväntades att kontrollgruppen som inte fick någon feedback angående sitt resultat på testet inte skulle uppvisa någon signifikant skillnad i responstiderna mellan omgångarna.

Anledningen till att den sociala kategorin araber valdes som utgrupp var att det vore intressant att studera svenskars attityder till denna grupp, då araber är stigmatiserade i dagens Sverige (Agerström & Rooth, 2009).

SPF isolerar bias rörande ingrupsfavorisering och utgruppshostilitet (Bar-Anan et al., 2009). Detta kan bidra till att mätinstrumentet kan vara användbart för att undersöka om försöksdeltagarna separerar ingrupsfavorisering från utgruppshostilitet, genom att korrigera för ingrupsbias om de får feedback att de favoriserar ingruppen, och korrigera för utgruppsbias om de får feedbacken att de uppvisar en negativitet mot utgruppen.

Vidare var det av intresse att undersöka hur skillnader i deltagarnas externa och interna motivation att inte uppvisa fördomar påverkade resultatet på SPF. Vi uppställde däremot ingen hypotes angående hur detta skulle kunna påverka reaktionstiderna.

Metod

Försöksdeltagare

Urvalet bestod av 90 försöksdeltagare (50 kvinnor och 40 män) mellan 19-54 år ($M=24,5$, $SD=5,4$). De flesta studerade vid Lunds Universitet på Institutionen för Psykologi eller Institutionen för Teologi. Bland försöksdeltagarna var 12 personer utlandsfödda, 16 personer hade 2 föräldrar som var födda utomlands och 11 personer hade 1 förälder som var utlandsfödd. Försöksdeltagare rekryterades genom att presentera studien på olika psykologiföreläsningar, erbjuda personer på universitetsområdet att delta, samt sätta upp information angående studien på olika fakulteter.

Deltagarna randomiserades in i 3 lika stora grupper i samband med att de kom in i testlokalen. Experimentgrupperna fick falsk feedback och en av grupperna fungerade som kontrollgrupp. För att undvika bias var försöksledarna inte medvetna om vilken grupp som fick vilken feedback. Som tack för sin medverkan erbjöds deltagarna kexchoklad, samt att medverka i utlottning av biobiljetter.

Material

För att testa implicit korrigerad av ingrupsfavorisering och utgruppshostilitet användes *Sorting Paired Features Task* (SPF) (Bar-Anan et al., 2009) som bestod av 16 olika ansiktsbilder, 8 bilder på araber och 8 på svenskar. Bilderna valdes utifrån fysiska attribut som associeras med antingen svensk eller arab. Bilderna var mestadels på lokala fotbollsspelare, detta för att sannolikheten att någon skulle bli igenkänd var liten. Det var en jämn könsuppdelning för både gruppen araber och svenskar och alla bilderna visades lika många gånger i en slumpmässig ordning. För att undvika att attraktivitet och emotionsuttryck hos ansiktena skulle påverka bedömningen var detta balanserat för de olika grupperna.

Både araber och svenskar kunde kategoriseras som positiva och negativa, och detta skedde i lika stor utsträckning för båda de sociala kategorierna. Det fanns 4 positiva ord, exempelvis underbar, och 4 negativa ord, exempelvis avskyvärd, som repeterades hela tiden. Ordningen var slumpmässig. Detta bidrog till 4 olika kategorier sammanlagt (svensk/positiv, svensk/negativ, arab/positiv, arab/negativ). Kategorierna presenterades i de 4 hörnen på skärmen. I vilket hörn en viss kategori visades varierade. Denna variation var slumpmässig. Detta för att undvika att eventuella skillnader i höger och vänster hand, samt andra effekter skulle leda till bias. Bilden och ordet som skulle kategoriseras visades i mitten av skärmen (Se *Figur 1*). Testomgångarna balanserades med hjälp av blockrandomisering.



Figur 1. Bilden är ett exempel skapat av Bar-Anan et al. (2009) och visar hur *Sorting Paired Features Task* kan se ut på skärmen.

För att uppnå manipulationskontroll angående om försöksdeltagarna uppfattade den falska feedbacken under SPF konstruerades följande frågor: *"Fick du någon feedback angående dina attityder under testet?"*. Det fanns två svarsalternativ, *Ja* och *Nej*. Anledningen till att en ledande fråga valdes var antagandet att färre personer hade gett information om feedbacken under SPF om frågan varit öppen. Nästa fråga var; *"Om ja, vad fick du för feedback?"*. Detta för att undersöka om försöksdeltagarna uppfattade distinktionen mellan ingrupsfavorisering och utgruppshostilitet. Tredje frågan löd: *"Tror du feedbacken påverkade utförandet av resten av testet? I så fall hur?"*. Syftet med denna fråga var att försöka skapa en uppfattning om eventuell påverkan var explicit eller implicit.

Internal motivation to respond without prejudice scale (IMS) och *external motivation to respond without prejudice scale (EMS)* (Plant & Devine, 1998) användes. Båda formulären består av 5 påståenden om attityder mot invandrare på en skattningsskala från 1-9 (*1= Instämmer inte alls, 9= Instämmer helt*). Ett exempel på ett påstående från IMS är *"Jag försöker handla fördomsfritt eftersom det är viktigt för mig"*. Ett exempel på ett påstående från EMS är *"Jag försöker undvika att vara fördomsfull mot invandrare för att slippa andras ogillande"*. Plant och Devine visade på en acceptabel nivå av samstämmighet mellan frågorna (internal consistency). Cronbach alpha koefficienten för EMS och IMS varierade mellan 0,76 -0,85 i 3 stickprov. I föreliggande studie var Cronbach alpha koefficienten 0,83 för EMS och 0,71 för IMS.

För att mäta ingrupsidentifiering användes Schmaders (2002) modifierade *Collective Self-Esteem Scale*, vars syfte är att mäta ingrupsidentifiering i relation till sitt genus. Originalkällan skapades av Luhtanen och Crocker (1992). Syftet med modifieringen var att isolera påståendena på skalan som syftar till att mäta ingrupsidentifiering i relation till sitt genus (Schmaders, 2002). Detta formulär var sedan översatt till svenska av Eriksson och Lindholm (2007), och det är den översättningen vi har använt oss av. Cronbach alpha koefficienten på den översatta skalan var 0,66. I denna studie byttes ingruppen *kvinnor* från Erikssons och Lindholms översatta formulär ut till *svensk*. Formuläret bestod av fyra påståenden på en skala 1-9 (*1 = Instämmer inte alls, 9 = Instämmer helt*). Exempelvis var ett påstående ”Att vara svensk är en viktig del av den jag är.” Denna skala visade sig ha en god samstämmighet mellan frågorna (internal consistency), Cronbach Alpha koefficienten var 0,81.

Försöksdeltagarna blev ombudda att skatta hur varma och kalla deras känslor var mot svenskar respektive araber. Syftet med detta var att det skulle fungera som ett explicit attitydmått. Skattningen bestod av två påståenden, ett rörande svenskar och ett rörande araber. Man skattade på en skala 1-9 (*1 = Svalast möjliga känslor, 9 = Varmast möjliga känslor*).

Följande fråga konstruerades med syftet att skapa medvetenhet om försöksdeltagarna gissade hypoteserna: *Vad tror du vi ville komma fram till med studien?* Försöksdeltagarna fick även uppge demografisk data, samt svara på om de gjort ett liknande attitydtest förut. Detta för att veta om de eventuellt hade lärt sig att manipulera testet.

Procedur

Försöksdeltagarna blev informerade om att de skulle medverka i ett attitydtest på 15-20 minuter, att det var frivilligt, anonymt och att de kunde avbryta under testets gång. De blev visade in i testlokalen. Personerna testades framför en avskärmd dator och 1-5 försöksdeltagare testades samtidigt. Minst en försöksledare var alltid närvarande i rummet för att svara på eventuella frågor.

SPF började med att försöksdeltagarna fick instruktioner om hur de skulle sortera ansiktena och orden i de 4 olika kategorierna. Stimulit de skulle kategorisera presenterades i mitten av skärmen. Personerna placerade ordet och bilden i rätt fack med hjälp av tangenterna D, E, K och O. Varje tangent representerade en viss kategori. Vid felkategorisering visades ett rött X på skärmen, och försöksdeltagaren måste korrigera misstaget för att gå vidare. Tiden det tar för personerna att korrigera för misstaget räknas som reaktionstid.

Efter en provomgång på 32 repetitioner började första delen av testet. Detta bestod av 64 repetitioner. När försöksdeltagaren var klar visades ett meddelande på skärmen.

Meddelandet varierade beroende på vilken grupp personen randomiserats till. Alla grupperna fick följande information:

Testet heter Sorting Paired Features Task (skapad av Bar-Anan, Nosek & Vianello) och mäter implicita eller automatiska attityder. Resultatet beräknas genom att kolla på responstiden i de olika kategorierna. Desto närmare två begrepp är associerade med varandra. Desto snabbare svarar man när man skall kategorisera dessa.

Grupp 1 och 2 fick även falsk feedback. Grupp 1 fick meddelandet att de hade en stark automatisk positivitet till svenskar. Grupp 2 fick meddelandet att de hade en stark automatisk negativitet mot araber. Kontrollgruppen fick ingen falsk feedback.

Efter att försöksdeltagarna fått meddelandet startades andra delen av testet genom att de tryckte på mellanslag. Del 2 var en upprepning av del 1.

Det första formuläret bestod av frågorna angående feedbacken. Anledningen till att detta formulär gavs ut först var att frågorna angående korrigering rörde hypoteserna, och därför var det av stort intresse att svaren skulle bli så lite påverkade som möjligt. När försöksdeltagarna var klara med detta räckte de upp handen och fick de övriga frågeformulären att fylla i. När de var klara förklarade en av försöksledarna syftet med studien och samtliga deltagare fick förklarat att den feedback de fått under testet inte återspeglade deras verkliga attityder utan att det var en del av manipulationen.

Resultat

Först screenades all data. I enlighet med Bar-Anan et al. (2009) togs alla responser som var snabbare än 400 ms och långsammare än 5000 ms bort. Efter detta kontrollerades att ingen person fick mer än 10 % av sina responser i någon kategori borttagna. Ett t-test för oberoende medelvärden utfördes för att undersöka om det fanns någon skillnad mellan deltagare som var födda i Sverige och deltagare som var utlandsfödda i deras skattning av ingrupsidentifiering med svenskar. Ingen signifikant skillnad mellan grupperna uppstod och grupperna antogs därför vara jämlika. Därmed kunde alla deltagare behållas i analysen. Samtliga analyser gjordes på logtransformerade data men för att underlätta tolkningen rapporteras resultatet i millisekunder.

Hypoteserna var att resultaten skulle uppvisa skillnader i skattningen av ingruppen mellan omgång 1 och omgång 2 för gruppen som fått feedback att de uppvisar en stark automatisk positivitet till svenskar, och skillnad i skattningen av utgruppen mellan omgång 1 och omgång 2 för gruppen som fått feedback att de uppvisar en stark automatisk negativitet mot araber. Gruppen som fått svensk positiv feedback skulle korrigera för denna påstådda bias genom att uppvisa en långsammare respons till bilder av svenskar med positiva ord

medan övriga responsmönster inte förväntades förändras. Vidare hypotiserades att gruppen arab/negativ skulle uppvisa ett långsammare responsmönster till arab och negativ i omgång 2 medan övriga responsmönster inte skulle förändras. En interaktionseffekt predicerades mellan gruppen som fått svensk/positiv feedback och gruppen som fått arab/negativ feedback. Gruppen som inte fått någon feedback förväntades inte uppvisa någon signifikant skillnad i responstiderna mellan omgång 1 och omgång 2. För att testa dessa hypoteser utfördes en repeated measures ANOVA med en mellangrupsfaktor för feedback (svensk/positiv, arab/negativ, ingen).

Hypoteserna att gruppen som fått svensk/positiv och gruppen som fått arab/negativ feedback skulle ändra responsmönster under omgång 2 fick inte stöd, $F(2,87) = 0,161$, $p > 0,05$. Detta innebär att feedbacken inte påverkade grupperna på ett sätt så att responstiden förändrades mellan gångerna.

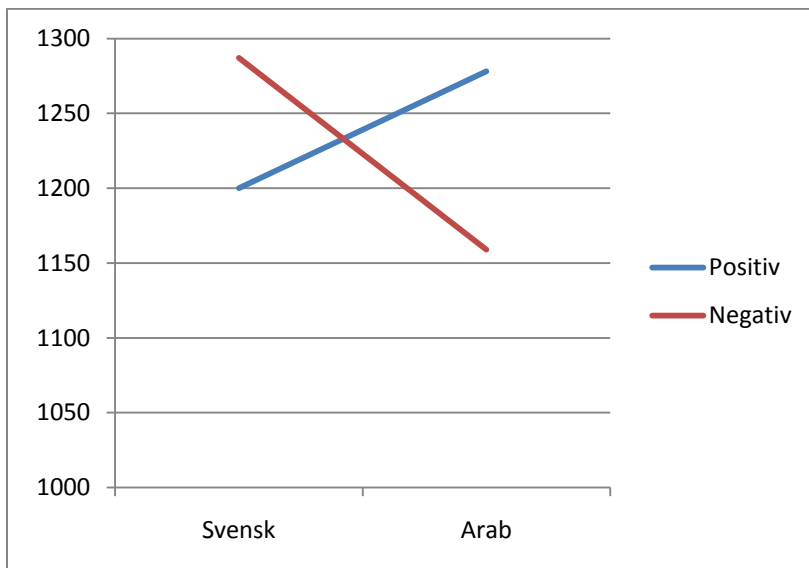
Vid närmare analys av resultatet framkom en nästan signifikant tendens till att samtliga grupper (svensk/positiv, arab/negativ, ingen) hade olika skattningsmönster mellan omgång 1 och omgång 2, $F(1,87) = 3,728$, $p = 0,057$, $\eta^2 = 0,041$. I omgång 1 var samtliga försöksdeltagare långsammare på att para ihop svensk med negativa ord ($M = 1285$ ms) än svensk med positiva ord ($M = 1215$ ms) och långsammare på att para ihop arab med positiva ord ($M = 1223$ ms) än arab med negativa ord ($M = 1182$ ms). I omgång 2 så var deltagare långsammare på att para ihop svensk med negativa ord ($M = 1167$ ms) än svensk med positiva ord ($M = 1122$ ms) och snabbare på att para ihop arab med positiva ord ($M = 1057$ ms) än arab med negativa ord ($M = 1059$ ms). Denna icke signifikanta tendens verkar påvisa att även kontrollgruppen blev påverkad av informationen som gavs mellan omgång 1 och omgång 2 alternativt att en övningseffekt har uppstått och därmed uppstod inga meningsfulla manipulationsskillnader för feedbacksfaktorn.

En interaktionseffekt fanns mellan etnicitet (svensk, arab) och valens (positiv, negativ) $F(1,87) = 15,822$, $p < 0,000$, $\eta^2 = 0,154$. I samtliga betingelser (svensk/positiv, arab/negativ, ingen) var försöksdeltagare långsammare på att para ihop bilder på svenskar med negativa ord ($M = 1226$ ms) i förhållande till att para ihop bilder på araber med negativa ord ($M = 1120$ ms) och förhållandevis lika snabba på att para ihop både bilder på svenskar med positiva ord ($M = 1168$ ms) och araber med positiva ord ($M = 1140$ ms). Dessa resultat är i enlighet med resultat i studien av Bar-Anan et al. (2009).

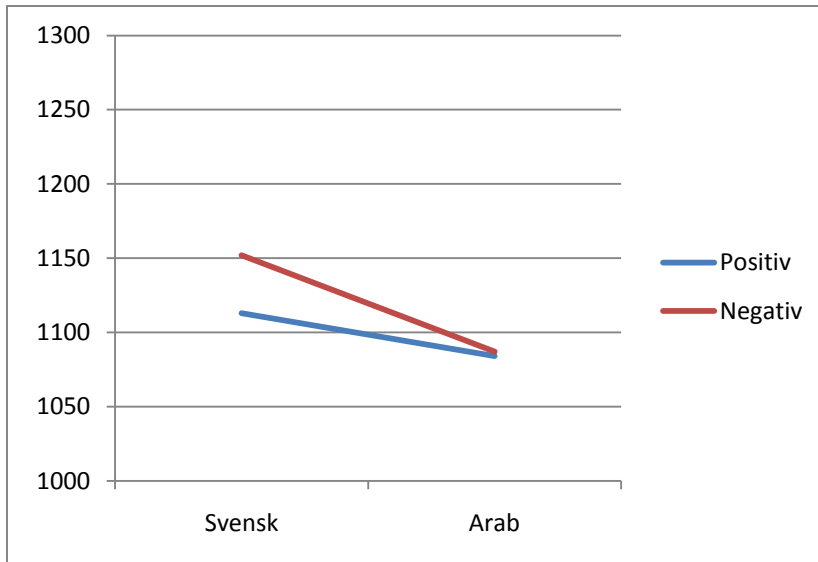
En intressant fyrvägs interaktionseffekt uppstod vid analys av responstid i SPF och vår explicita attitydmätning mellan omgång (1, 2), etnicitet (svensk, arab), valens (positiv, negativ) och skattningen av explicit motivationsskala (EMS) som gjorts om till dikotom

variabel (hög, låg) med en mediansplit, $F(1,87) = 4,062, p = 0,047, \eta^2 = 0,045$.

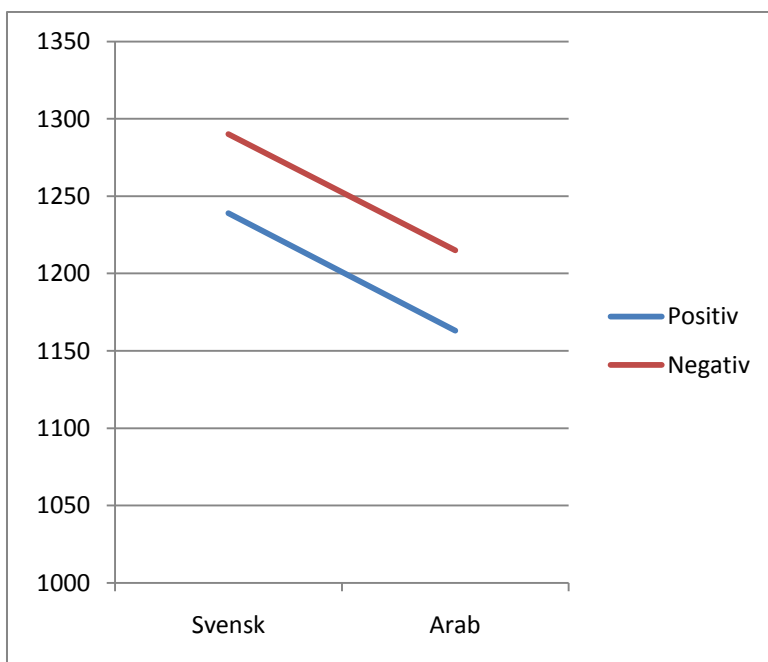
Försöksdeltagare med hög EMS skiljde sig åt mellan omgång 1 och omgång 2, $F(1,47) = 10,800, p = 0,002, \eta^2 = 0,187$. I omgång 1 hade deltagare med hög EMS snabbare reaktionstid på bilder på svenskar med positiva ord ($M = 1200$ ms) än bilder på svenskar med negativa ord ($M = 1287$ ms) och snabbare reaktionstid på bilder på araber med negativa ord ($M = 1159$ ms) än araber med positiva ord ($M = 1278$ ms) (se *Figur 2*). I omgång 2 så uppvisade försöksdeltagare med hög EMS en annan responstid från omgång 1 och det fanns i denna omgång ingen preferens för svenskar över araber (se *Figur 3*). Vidare analys av resultaten visar att försöksdeltagare med låg EMS inte uppvisade en signifikant skillnad mellan omgång 1 och omgång 2, $F(1,40) = 0,019, p > 0,05, \eta^2 = 0,000$. Deltagare med låg EMS uppvisade ingen bias i varken omgång 1 eller omgång 2 (se *Figur 4* och *Figur 5*).



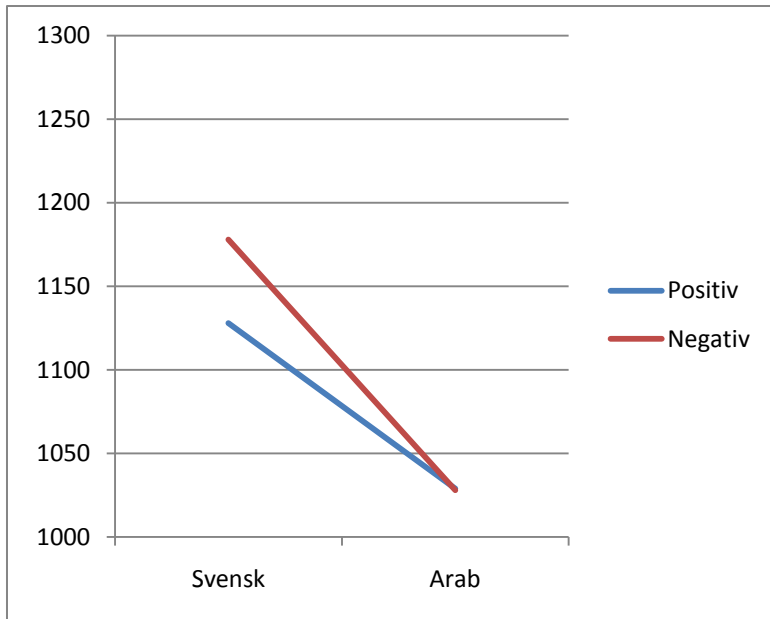
Figur 2. Responstider till bilder på svenskar och araber med positiva och negativa ord för deltagare med höga poäng på *external motivation to respond without prejudice scale* under omgång 1.



Figur 3. Responstider till bilder på svenskar och araber med positiva och negativa ord för deltagare med höga poäng på *external motivation to respond without prejudice* scale under omgång 2.



Figur 4. Responstider till bilder på svenskar och araber med positiva och negativa ord för deltagare med låga poäng på *external motivation to respond without prejudice* scale under omgång 1.



Figur 5. Responstider till bilder på svenskar och araber med positiva och negativa ord för deltagare med låga poäng på *external motivation to respond without prejudice scale* under omgång 2.

Hypoteserna att gruppen som fått svensk positiv feedback och gruppen som fått arab negativ feedback skulle korrigera för denna påstådda bias under omgång 2 kunde inte stödjas av denna studie.

Diskussion

Syftet med denna studie var att undersöka om feedback angående ingrupsfavorisering eller utgruppshostilitet i kontexten av ett implicit test leder till en korrektion av denna påstådda bias. I vissa sammanhang är det inte önskvärt att vara positiv till ingruppen eller för den delen negativ mot utgruppen. Hypotesen var att deltagare som fick feedback om att de uppvisade stark automatisk positivitet för svenskar skulle ha ett långsammare responsmönster till svensk och positiva ord i omgång 2 av SPF, medan övriga responsmönster inte signifikant skulle förändras. Vidare hypotiserades det att deltagare som fick feedback om att de uppvisade stark automatisk negativitet mot araber skulle ha ett långsammare responsmönster till arab med negativa ord i omgång 2 av SPF, medan övriga responsmönster inte signifikant skulle förändras. Denna studie har inte kunnat bekräfta dessa hypoteser. I påföljande diskussion framläggs olika förklaringar till varför hypoteserna inte fick stöd, samt vilka förbättringar av studien som kan göras vid framtida forskning. Ett oväntat resultat i form av en fyrvägs interaktionseffekt för omgång x etnicitet x valens x EMS

uppstod. Innebörden och tolkning av denna interaktionseffekt diskuteras i förhållande till tidigare forskning.

En möjlig förklaring till varför hypoteserna inte fick stöd är att stereotyperna var så pass automatiserade att en korrektion blev omöjlig. Bargh (1999) menar att automatiserade attityder aktiveras utanför vår medvetenhet och att det är praktiskt taget omöjligt att kontrollera detta, vilket i sig gör möjligheten att minska stereotyper nästan obefintlig. För att kunna kontrollera en automatiserad stereotyp måste individen först vara medveten om att en omedveten automatiserad stereotyp finns och sedan vara motiverad samt ha möjlighet att korrigera denna. Vidare menar Bargh att även lågt fördomsfulla individer kan uppvisa stereotypiska antaganden när de tolkar andra människors beteenden. Oavsett hur individer går till väga kommer det enligt Bargh bli fel när de försöker kontrollera sina automatiserade stereotyper, och under tid leder detta till att individer stereotypiserar ännu mer. Problemet med Barghs resonemang är att det faktiskt visat sig möjligt att korrigera automatiska stereotyper (Blair, 2002; Lowery et al., 2001; Wallaert et al., 2010). Det är fortfarande en öppen fråga om stereotyper är så pass automatiserade som Bargh hävdar eller om dessa stereotyper i viss mån är påverkbara. Rimligtvis finns det en dissociation mellan en automatiserad och en kontrollerad process, men det är möjligt att dessa processer ibland är kombinerade som Olson och Fazio (2009) föreslår. En korrektionsprocess kan i sig vara svår när stereotyper är automatiserade, men det borde inte vara omöjligt, speciellt inte för motiverade individer förutsatt att de har tillräckligt kognitiva resurser fria, att göra denna korrigering.

Vidare har Björklund et al. (2010) visat att kontexten påverkar i vilken grad människor uppvisar ingrupsfavorisering. Blair (2002) menar att ett attitydobjekt i olika kontexter kan ge upphov till olika automatiska associationer. Det var detta som testades i denna studie då det predicerades att associationen av svensk och arab borde bli annorlunda i ljuset av feedbacken som deltagarna fick. I studien som Björklund et al. framlade var skattningen explicit, men flertalet deltagare uppgav efteråt att de inte påverkats. Därför testades i föreliggande studie om detta även fungerade under ett implicit test. Självklart innebar det mindre kognitiv börda att korrigera för en ingrupsfavorisering i det sammanhang som Björklund et al. visade. I vår studie var försöksdeltagare under en kognitiv börda i form av att de skulle hitta rätt kategorisering till bilder och ord i fyra grupper, och detta kan i sig ha minskat möjligheten att korrigera för ingrupsfavorisering eller utgruppshostilitet.

En möjlig förklaringsmodell till varför hypoteserna inte fick stöd är att deltagare var motiverade att korrigera men inte hade möjlighet. Det kan tänkas att den falska feedbacken

leder till en känsla av skuld. I enlighet med Monteith et al. (2002) kan skuldkänslor leda till att individen upplever en diskrepans (förutsatt att denna individ är lågfördomsfull) mellan den automatiskt aktiverade attityden och hur individen upplever att de borde respondera i förhållande till sina egna personliga åsikter. Denna diskrepans kan då tänkas motivera individen att korrigera sina responser. Enligt Olson och Fazio (2009) är motivation och möjlighet en nödvändighet för att en individ skall kunna korrigera en automatiskt aktiverad respons. Försöksdeltagarna borde av den feedback de fick under testet blivit motiverade att korrigera för bias men det är inte säkert att de skulle haft möjlighet. I det explicita frågeformuläret svarade en av deltagarna på frågan *”Tror du att feedbacken påverkade utförandet på resten av testet? I så fall hur?”* på följande sätt: *”Ja, det kan det säkert ha gjort. Jag vill inte ha automatiska negativa attityder mot araber! Så jag försökte, vet dock ej om det lyckades då jag inte vet hur testet är utformat.”* Detta kan tolkas som att deltagaren inte hade möjlighet att korrigera för bias. Forskning har visat att människor kan korrigera automatiska responser vid implicita test (Blair, 2002; Lowery et al., 2001; Wallaert et al., 2010). Lowery et al. och Wallaert et al. använder sig av *Implicit Association Test* och i föreliggande studie används SPF. Till vår kännedom så har SPF inte använts i detta sammanhang tidigare och det är möjligt att SPF är ett test som gör det svårare för människor att korrigera för bias eftersom det är en större kognitiv börda med SPF än *Implicit Association Test*. Om så är fallet är detta enbart positivt då det innebär att om det är svårt att fejka ett resultat på SPF så är testet reliabelt.

Det predicerades i enlighet med Björklund et al. (2010) att deltagare i gruppen svensk/positiv skulle uppvisa en mindre positivitet till ingruppen svenskar i omgång 2, då det i situationen inte borde vara en önskvärd reaktion att vara starkt positiv till svenskar. Vidare predicerades att deltagare i gruppen arab/negativ skulle uppvisa mindre negativitet mot utgruppen araber i omgång 2. En korrigering för denna påstådda bias kan tänkas uppkomma eftersom det inte borde vara önskvärt att i denna situation vara starkt negativ till araber. Det skulle även vara möjligt att få en omvänd reaktion så att gruppen som fick svensk/positiv feedback skulle bli mindre negativ mot araber och gruppen arab/negativ skulle bli mindre positiv mot svenskar, men det finns inte något stöd för detta i resultatet. Det är dock troligt att en korrigeringsprocess sker inom den kategori som deltagare får feedback om då feedbacken i sig inte uttalar sig om hur individen presterar i de andra kategorierna.

Då det verkar som att samtliga grupper inklusive kontrollgruppen har blivit påverkade mellan omgång 1 och omgång 2 så finns det en möjlighet att meddelandet om att testet mäter responstid och associationsstyrka som samtliga grupper fick påverkade responserna på testet.

Denna icke signifikanta tendens visade att samtliga grupper fick en förändring i skattningen av arab positiv i omgång 2 så att en snabbare reaktionstid på arab positiv uppstod. Detta kan bero på att alla grupper blev påverkade av informationen mellan omgångarna eller så uppstod det en övningseffekt. Vid en replikering av studien föreslår vi en förändring i form av att införa två kontrollgrupper där en grupp endast skall få meddelande om att det är paus i testet, och den andra kontrollgruppen får meddelandet om hur testet är utfört. Om en grupp inte får meddelande om hur testet är uppbyggt så kan det kontrolleras om eventuell förändring mellan omgångarna beror på en övningseffekt eller en manipulation. Anledningen att kontrollgruppen i denna studie fick detta meddelande var för att behålla testbetingelserna konsekventa så att det endast var variation om feedback på ingrupsfavorisering och utgruppshostilitet.

Det är även möjligt att feedbacken om ingrupsfavorisering tolkades annorlunda än från hur det var tänkt. Enligt Brewers (1999) idé behöver ingrupsfavorisering och utgruppshostilitet inte vara sammanlänkade, men det är fullt möjligt att människor i allmänhet tänker att en positivitet mot ingruppen innebär en negativitet mot utgruppen när de blir ställda inför det faktum att de favoriserar sin egen grupp. Det är därmed möjligt att människor tolkar denna ingrupsfavorisering som en negativitet mot utgruppen. Om så är fallet är det möjligt att anledningen till att det fanns en tendens till att lättare para ihop arab med positiva ord i omgång 2 var för att samtliga grupper upplevde att de uppvisade bias mot utgruppen. När kontrollgruppen läste feedbacken om att testet mäter responstid och associationsstyrka är det fullt möjligt att individerna själva upplevde att de var snabbare på att para ihop svensk med positiv. Detta är en förväntad respons förutsatt att svenskar är ens ingrupp. Då är det också möjligt att individerna koncentrerade sig extra mycket för att justera så att arab/positiv skulle bli bättre i omgång 2. Som svar på frågan *"Tror du att feedbacken påverkade utförandet av resten av testet. I så fall hur?"* svarade en deltagare som tillhörde gruppen svensk/positiv på följande sätt: *"Man tänkte lite extra på att snabbare lyckas med arab delen"* och en annan deltagare som också tillhörde svensk/positiv-gruppen uttryckte sig på följande sätt: *"Man vill ju gärna vara fördomsfri så kanske ansträngde jag mig ännu mer för att vara snabbare när det gällde positiv och arab"*. Detta kan tolkas som att dessa två deltagare upplevt att en positivitet mot ingruppen innebär en negativitet mot utgruppen, och därmed har de försökt justera för denna negativitet. Det är därför en möjlighet att den icke signifikanta tendens till att samtliga grupper blivit påverkade av informationen efter omgång 1 beror på att feedbacken tolkades som negativitet mot utgruppen i samtliga betingelser, och att deltagare försökte korrigera för detta. Av de 60 deltagare som tillhörde experimentgrupperna och svarade på ovanstående fråga gällande ifall de hade blivit påverkade av feedbacken så var det endast 12

personer som uppgav att de inte blivit påverkade av feedbacken. Eftersom detta var en följdfråga som endast besvarades av grupperna som fått feedback så är det svårt att uttala sig om huruvida kontrollgruppen upplevde att de blev påverkade. Vid en replikation av studien som föreslagits ovan med fyra grupper så bör detta kontrolleras i samtliga grupper.

Det uppstod en interaktionseffekt som visade att individer med hög och låg extern motivation att kontrollera fördomar reagerade annorlunda i omgång 2 oavsett vilken feedback som hade getts. Deltagare med låga poäng på EMS uppvisade ingen signifikant skillnad på vare sig araber eller svenskar med positiva eller negativa ord mellan omgång 1 och omgång 2. De uppvisade inte heller någon ingrupsfavorisering eller utgruppshostilitet i någon av omgångarna. Deltagare med höga poäng på EMS visade en stark preferens för svenskar över araber i omgång 1 men i omgång 2 fanns ingen signifikant preferens för svenskar över araber. Det faktum att individer med höga poäng på EMS uppvisade en starkare bias än individer med låga poäng på EMS är inte ett förvånade resultat då forskning visat att individer med höga poäng på EMS uppvisar en starkare bias vid implicita test än personer som inte uppvisar höga poäng på EMS (Devine et al., 2002; Hausmann & Ryan, 2004).

Det som skiljer sig i resultatet i föreliggande studie från tidigare forskning är att individer med höga poäng på EMS har lyckats korrigera för denna bias under ett implicit test medan de med låga poäng EMS inte uppvisade bias i första omgången, och därmed inte hade något att korrigera. Bargh (1999) föreslår att individer som försöker korrigera för bias kommer att få ett slags backslasheffekt så att det resulterar i mer bias. Hausmann och Ryan (2004) som undersökte just individer med höga poäng på EMS i förhållande till personer som har höga poäng på EMS fick evidens för att individer med höga poäng på EMS inte kunde korrigera för denna bias under ett *Implicit Association Test*. Plant och Butz (2009) diskuterar även detta i relation till *Self-Determination Theory*, där de påstår att idealet bör vara internaliserat för att man skall ha möjlighet att korrigera på ett tillfredställande sätt. I relation till MODE-teorin borde det inte vara möjligt att korrigera för ett implicit anonymt test om motivationen är extern. Olson och Fazio (2009) menar att korrigering av attityder är omöjlig om motivationen saknas. Det är osäkert om motivationen att framstå som fördomsfri inför andra är tillräcklig för att ge individer verktygen som krävs för att korrigera ett implicit anonymt test. Det skulle kunna vara så att personer som har höga poäng på EMS inte brydde sig om att korrigera för denna bias under ett vanligt *Implicit Association Test* som Hausmann och Ryan (2004) utförde. I föreliggande studie är det en möjlighet att feedback om att individen uppvisade bias ledde till en extern motivation att vilja korrigera för att inte visa sig fördomsfull inför andra människor. Självklart var vår undersökning anonym och detta

informerades samtliga deltagare om men frågan är om det känns anonymt att göra ett test bland andra människor. Eventuellt skulle testledarnas närvaro i denna studie kunnat leda till en önskan om att uppvisa sig på ett socialt önskvärt sett. Det är en möjlighet att detta skulle kunna leda till en korrigering av responser hos individer som uppvisade hög extern motivation. Det är då förutsatt att deltagarna med höga poäng på EMS upplevde att andra kommer få ta del av deras resultat, men eftersom testet var anonymt så borde de inte uppleva det så. Det kan vara frågan om en intern motivation om att inte uppvisa fördomar men det fanns inga signifikanta resultat för IMS, och personer med höga poäng EMS korrigerade sina resultat oavsett deras poäng på IMS. Det bör noteras att vi inte hade valt deltagare till denna studie beroende på deras nivå på EMS och IMS utan det är frågan om en naturlig variabilitet. För att undersöka detta närmare så kan deltagare först testas på dessa skalor och individer som endast uppvisar höga poäng på EMS och låga poäng på IMS används i undersökningen för att utreda vad som leder individer med höga poäng på EMS till att korrigera sina responser på detta test.

Då denna studie utfördes i en laboratoriemiljö och feedbacken inte var sann, bör man vara försiktig med att generalisera dessa resultat. Trots detta är det rimligt att anse att denna typ av forskning bidrar till värdefull information om huruvida människor försöker korrigera för eventuell bias när de blir medvetna om att de uppvisar ingrupsfavorisering eller utgruppshostilitet. För att öka den externa validiteten föreslår vi att man utför denna studie (och då även tillämpar de ändringar gällande feedbacken som diskuteras ovan) i olika miljöer, detta för att undersöka hur människor reagerar på bias gällande in- eller utgruppen i olika kontexter. På samma sätt som Björklund et al. (2010) visade att kontexten bidrar till olika värderingar av ingruppen, är det fullt möjligt att detta är sant även på implicit nivå, för utvärdering av både in- och utgruppen. Det är även möjligt att olika kontexter gör det mer eller mindre enkelt för människor att skilja på ingrupsfavorisering och utgruppshostilitet. Exempelvis är det möjligt att vid anställningssituationer, till exempel om vita anställer vita och män anställer män blir favoriseringen av ingruppen mer framträdande.

Så vitt vi vet är detta den första studien som undersöker om korrigering av ingruppskattningar uppstår i vissa situationer under ett implicit test. Björklund et al. (2010) undersökte detta i ett explicit sammanhang och med tanke på att det är möjligt att manipulera responser på implicita test (Lowery et al., 2001; Wallaert et al., 2010) så valde vi att testa detta med SPF. Vi känner inte till någon studie där korrektion har undersökts vid användandet av SPF och därför är det med försiktighet som vi vill uttala oss om våra resultat. Det är en möjlighet att det är svårare för individer att manipulera sina responser på SPF då detta test

innebär större kognitiv börda än ett *Implicit Association Test*. Resultaten i denna studie ger indikationer på att det skulle kunna vara så. Därför föreslår vi att SPF utsätts för hårdare test. Det skulle till exempelvis vara möjligt att göra en replikation av Wallaerts et al. studie, och använda SPF istället för *Implicit Association Test*. Om fallet är att det är svårt att manipulera resultat på SPF så innebär detta att ett starkt reliabelt mätninginstrument för implicita attityder finns tillgängligt för den framtida implicita attitydforskningen, och att detta kan leda till en djupare förståelse för människors underliggande attityder.

Referenser

- Agerström, J., & Rooth, D.-O. (2009) Implicit prejudice and ethnic minorities: Arab-Muslims in Sweden. *International Journal of Manpower*, 30, 43-55. doi: 10.1108/01437720910948384
- Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task a measure of association strengths. *Experimental Psychology*, 56, 329–343. doi: 10.1027/1618-3169.56.5.329
- Bargh, J.A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. I S. Chaiken & Y. Trope (Red.), *Dual-process theories in social psychology*. (ss. 361-382). New York: Guilford Press.
- Björklund, F., Bäckström, M., & Jörgensen, Ö. (In press). In-group ratings are affected by who asks and how: Interactive effects of experimenter group-membership and response format. *Journal of Social Psychology*.
- Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 3,242-261. doi: 10.1207/S15327957PSPR0603_8
- Brewer, M.B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*,55, 429-444. doi: 10.1111/0022-4537.00126
- Butz, D. A., & Plant, E. A. (2009). Prejudice control and interracial relations: The role of motivation to respond without prejudice. *Journal of Personality*, 77, 1311-1341. doi: 10.1111/j.1467-6494.2009.00583.x
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87, 546-563. doi: 10.1037/0033-2909.87.3.546
- Eriksson, K., & Lindholm, T. (2007). Making gender matter: The role of gender-based expectancies and gender identification on women's and men's math performance in Sweden. *Scandinavian Journal of Psychology*, 48, 329-338. doi: 10.1111/j.1467-9450.2007.00588.x
- Deci, E.L., & Ryan, R.M. (2000). The what and the why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Enquiry*, 11, 227-268. doi: 10.1207/S15327965PLI1104_01
- Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18. doi: 10.1037/0022-3514.56.1.5

- Devine, P.G., Plant, E.A., & Amodio, D.M. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, 82, 835-849. doi: 10.1037/0022-3514.82.5.835
- Dunton, B.C., & Fazio, R.H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23, 316-326. doi: 10.1177/0146167297233009
- Fazio, R.H., Jackson, J.R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013-1027. doi: 10.1037/0022-3514.69.6.1013
- Greenwald, A.G., McGhee, D.E., & Schwatz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480. doi: 10.1037/0022-3514.74.6.1464
- Hausmann, L.R.M., & Ryan, C.S. (2004). Effects of external and internal motivation to control prejudice on implicit prejudice: The mediating role of efforts to control prejudiced responses. *Basic and Applied Social Psychology*, 26, 215-225. doi: 10.1207/s15324834basp2602&3_8
- Lowery, B.S., Hardin, C.D., & Sinclair, S. (2001). Social influences effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81, 842-855. doi: 10.1037/0022-3514.81.5.842
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem Scale: Self-evaluation of one's social identity. *Society for Personal and Social Psychology*, 18, 302-318. doi: 10.1177/0146167292183006
- McConahay, J.B., Hardee, B.B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, 25, 563-579.
- Monteith, M.J., Ashburn-Nardo, L., Voils, C.I., & Czopp, A.M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, 83, 1029-1050. doi: 10.1037/0022-3514.83.5.1029
- Mummendey, A., & Schreiber, H.-J. (1983). Better or just different? Positive social identity by discrimination against, or differentiation from outgroups. *European Journal of Social Psychology*, 13, 389-397. doi: 10.1002/ejsp.2420130406
- Olson, M.A., & Fazio, R.H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. I R. E. Petty, R.H. Fazio, & P. Brinol (Red.),

Attitudes: Insights from the new implicit measures (ss. 19-63). New York: Psychology Press.

Plant, E. A., & Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832. doi: 10.1037/0022-3514.75.3.811

Reynolds, K.J., Turner, C.J., & Haslam, A. (2000). When are we better than them and they are worse than us? A closer look at social discrimination in positive and negative domains. *Journal of Personality and Social Psychology*, 78, 64-80. doi: 10.1037/0022-3514.78.1.64

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194-201. doi: 10.1006/jesp.2001.1500

Wallaert, M., Ward, A., & Mann, T. (2010). Explicit control of implicit responses: Simple directives can alter IAT performance. *Social Psychology*, 41, 152-157. doi: 10.1027/1864-9335/a000022