Litteraturgranskningar

THOMAS DENK: *Värnpliktsutbildningen – en politisk socialisationsagent?* Karlstad University Studies 1999:10, Karlstad 1999.

I sin avhandling ställer Thomas Denk frågan om värnpliktsutbildningen är en politisk socialisationsagent. Med politisk socialisationsagent avses en aktör som påverkar individens politiska kultur. Politisk kultur innefattar två dimensioner: orientationer och politiska strukturer. Vardera dimensionen har tre kategorier. I anslutning till socialpsykologisk teori säges orientationer bestå av politisk kognition (uppfattningar om hur det politiska systemet är beskaffat), politisk affektion (känslor inför det politiska systemet) och politisk evaluation (omdömen om det politiska systemet). De politiska strukturer som individen orienterar sig till är politisk samfällighet (den grupp personer som ingår i det politiska systemet), politisk regim (relationen mellan medlemmarna i den politiska samfälligheten, regimnormer samt rollstrukturer) och politisk aktörsstruktur (relationen mellan aktörer som syftar att utföra politiska handlingar).

De kommunikativa handlingar som agenten utför bildar en politisk socialisationsprocess. Avhandlingen grundas på en empirisk undersökning av värnplikten som politisk socialisationsstruktur. Till grund för slutsatserna ligger två enkäter. Den ena riktades till ett urval om 733 värnpliktiga vid fem förband; materialet insamlades mot slutet av de värnpliktigas utbildningstid. Den andra besvarades av 630 mönstrande ungdomar; denna grupp betraktas som en "kontrollgrupp" (s. 51). Enkäterna innehöll framför allt frågor om individernas uppfattningar om det politiska systemet, exempelvis vad som anses känneteckna en god medborgare, hur man kan påverka politiska beslut samt omdömen om politiker. Dataanalysen grundas dels på en jämförelse mellan värnpliktiga och mönstrande, dels på skillnader inom gruppen värnpliktiga. Avhandlingen försöker också bedöma betydelsen av andra socialisationsagenter såsom television, tidningar, vänner, familj, arbete, organisationsmedlemskap och gymnasieutbildning.

Författaren drar slutsatsen att värnplikten påverkar det stora flertalet av de undersökta delaspekterna av politisk kultur. Avhandlingen utmynnar i konklusionen att "värnpliktsutbildningen intar rollen som politisk socialisationsagent" och att värnpliktsutbildningen sålunda bör "inkluderas i teoribildningen om politiska socialisationsagenter" (s. 168).

Slutsatserna är direkt avhängiga av den empiriska undersökningen. Dessvärre har undersökningen flera brister i metodhänseende.

Undersökningsuppläggningen betecknas i avhandlingen som en "falsk panel" (s. 63). Det av författaren myntade uttrycket är vilseledande, eftersom det över huvud taget inte rör sig om upprepade observationer av samma individer. Någon panelanalys är därmed inte möjlig att genomföra. Även om det hade varit praktiskt möjligt att genomföra en sådan panelstudie av en grupp värnpliktiga före, under och efter deras utbildning är det inte säkert att en sådan uppläggning hade varit den lämpligaste. En bedömning av värnpliktsutbildningens betydelse fordrar en jämförelsepunkt, naturligen med en kontrollgrupp som inte genomgått värnpliktsutbildning.

Avhandlingen innehåller en ansats till en sådan experimentell tankegång. Jämförelsen försväras emellertid av att kontrollgruppen består av ett urval av samtliga mönstrande, dvs. både sådana som uttagits till värnpliktsutbildningen och sådana som ej uttagits. Visserligen gör författaren en jämförelse mellan dessa båda undergrupper i urvalet av mönstrande, men de observerade differenserna mellan värnpliktiga och mönstrande är genomgående vanskliga att tolka. Det finns nämligen en betydande åldersskillnad. Detta förhållande redovisas inte i avhandlingen, men författaren har haft vänligheten att tillmötesgå min förfrågan om en särskild databearbetning på denna punkt. Genomsnittsåldern är bland de mönstrande 17 år och 10 månader och bland de värnpliktiga 20 år och 4 månader. Åldersskillnaden mellan de båda urvalen är sålunda två och ett halvt år. Många andra "socialisationsagenter" kan påverka individerna under dessa år. Den relativa betydelsen av värnpliktsutbildningen är därmed svårbedömd.

Avhandlingens redovisning av urval, fältarbete och bortfall är ofullständig. Urvalet av värnpliktiga baserades ursprungligen på tolv förband, men endast fem finns med i undersökningen. Effekterna av att sju förband utelämnats diskuteras inte. Urvalet av mönstrande är otillräckligt dokumenterat. Det framgår inte om urvalet är slumpmässigt och om de dragits i flera steg. Bortfallets eventuella effekter diskuteras knappast alls. Avhandlingen innehåller ingen uppgift om tidpunkten för datainsamlingen.

Av jämförelser mellan urvalen och respektive population framgår att det finns vissa systematiska fel. Flygvapnet är exempelvis överrepresenterat och flottan saknas. Plutonbefälsvärnpliktiga är överrepresenterade, medan värnpliktiga uttagna till depå- och bevakningstjänst är underrepresenterade. I urvalet av mönstrande finns det fler som uttagits till värnpliktsutbildning än i populationen. Avhandlingen innehåller ingen diskussion om hur dessa skevheter kan påverka slutsatserna.

Det slumpmässiga fel som man alltid måste beakta vid statistiska urvalsundersökningar beräknas i avhandlingen efter antagandet om ett obundet, slumpmässigt urval. Författaren erkänner själv att felmarginalen är relativt stor för undergrupper i de redan relativt små urvalen. Beräkningen av felmarginalerna "visar en begränsning hos undersökningen att generalisera resultat till specifika grupper utifrån den empiriska undersökningen" (s. 51). Ändå grundar sig åtskilliga av avhandlingens slutsatser på jämförelser mellan sådana specifika grupper.

Avhandlingen skiljer inte mellan statistisk signifikans och teoretisk signifikans. Om en skillnad visar sig vara statistisk säkerställd antas, utan någon diskussion, att den också är tillräckligt stor för att tillmätas betydelse utifrån avhandlingens teori och hypoteser. I praktiken är det därmed urvalsstorleken som kommer att avgöra hur många hypoteser som bekräftas.

Dataanalysen bygger genomgående på indexkonstruktioner. De olika aspekterna av politisk kultur mäts genom additiva index baserade på serier av intervjufrågor. Huruvida dessa frågor bildar skalbara mönster eller homogena dimensioner prövas aldrig. Avhandlingen avstår från att utnyttja den samhällsvetenskapliga metodlärans inventarium av faktoranalys, kumulativ skalning och andra sätt att pröva huruvida det finns empiriskt stöd för antaganden om skalbarhet. Indexkonstruktionerna vilar därför inte på empiriskt stöd utan på författarens egna antaganden.

Avhandlingen är inte begränsad till att undersöka värnpliktsutbildningen, utan syftet är också att studera betydelsen av andra socialisationsagenter. Med hänvisning till socialisationsteoretisk litteratur understryks betydelsen av att analysera den relativa betydelsen av olika socialisationsagenter. Syftet är vällovligt. Det kan förhålla sig så att den observerade differensen mellan värnpliktiga och övriga har uppkommit genom skillnader i andra avseenden, exempelvis i fråga om utbildning och medievanor. Läsaren väntar att hypotesen om andra socialisationsagenter skulle prövas genom någon gängse teknik för multivariat kausalanalys, såsom multipel regression eller motsvarande. Någon sådan analys genomförs emellertid inte. Det enda som redovisas är de parvisa korrelationerna mellan kontakterna med andra socialisationsagenter (appendix C2). Dessa beräkningar är emellertid inte tillräckliga för att dra några slutsatser om relativa effekter.

Lika problematisk är analysen av skillnaderna inom gruppen värnpliktiga. Tanken är att mäta betydelsen av värnpliktsutbildningens innehåll. De aspekter som uppmärksammas är försvarsgren, utbildningstid, befattningskategori och förbandstillhörighet. Varför just dessa aspekter valts ut motiveras aldrig med utgångspunkt från några teoretiskt grundade hypoteser. Urvalet framstår därför som ad hoc-artat, främst betingat av uppgifternas lättillgänglighet. Möjligheterna att dra några säkra slutsatser på grundval av denna databearbetning begränsas också, som redan nämnts, av de små urvalen och de därmed sammanhängande felkällorna.

Avhandlingen glider oförväget från korrelation till orsaksverkan. En observerad skillnad avseende förbandstillhörighet tas till intäkt för att värnpliktsutbildningens innehåll påverkat individernas politiska uppfattningar. Mothypotesen är givetvis att skillnaderna inte uppkommit genom socialisation utan genom selektion. Skillnaderna kanske helt enkelt beror på att värnpliktiga som uttagits till ett visst förband har högre utbildning än genomsnittet. Materialet hade gjort det möjligt att genom sedvanlig kausalanalys pröva ett antal sådana mothypoteser. Möjligheten har dessvärre inte utnyttjats.

Den del av individens politiska kultur som benämns politisk kognition operationaliseras genom att utnyttja informationen om vilka individer som svarat "vet ej" eller "osäker" på intervjufrågorna. Åsiktsförekomst används som mått på politisk kognition. Författaren finner att de värnpliktiga har signifikant högre värden på politisk kognition; hypotesen om att värnpliktsutbildningen fungerar som en politisk socialisationsagent anses därmed på denna punkt blivit bekräftad. Det finns emellertid uppenbara mothypoteser till att de mönstrande oftare svarar "vet inte" eller "osäker". Man vet från många andra undersökningar att åsiktsförekomst, liksom andra aspekter av politiskt engagemang, har ett tydligt samband med ålder. Inväxandet i politiken tar sig uttryck i stigande engagemang och växande säkerhet i de egna uppfattningarna. Åldersskillnaden på två och ett halvt år mellan mönstrande och värnpliktiga kan vara en viktig förklaring till den observerade skillnaden i andelen vet inte-svar. Det kan därför vara helt andra påverkansfaktorer än värnplikten som satt sina spår. Några sådana mothypoteser prövas dock inte.

De två återstående aspekterna av politisk kultur (politisk affektion och politisk evaluation) prövas på samma material som redan använts i kapitlet om politisk kognition. Visserligen är det denna gång inte endast andelen vet inte-svar som beaktas, utan också de övriga svarsalternativen. Procentberäkning och indexkonstruktion har emellertid utförts på sådant sätt att vet inte-svar även i dessa delar av avhandlingen påverkar resultaten. Ett exempel kan illustrera. Enligt tabell 5.1 är det 75,8 procent av de värnpliktiga och 69,3 procent av de mönstrande som svarar "ja, alltid" på frågan om man anser att en medborgare bör rösta i de allmänna valen. Differensen på 6,5 procentenheter tolkas som ett mått på värnpliktsutbildningens betydelse. Men denna skillnad beror till stor del på att de mönstrande oftare har kryssat för alternativet "vet ej/osäker", ett resultat som redan använts för att dra slutsatser i en tidigare del av avhandlingen. Läsaren kan själv räkna om procenttalen under antagande att vet inte-svaren exkluderas. Man finner då att procentsskillnaden sjunker från 6,5 till 2,5 procentenheter. Prövningen av de olika aspekterna av politisk kultur är därför inte oberoende av varandra. Samma empiriska resultat används dubbelt. Framställningen blir därmed vilseledande.

Ett sammanfattande omdöme måste beakta att avhandlingen har sina positiva sidor. Frågeställningen är både inomvetenskapligt och utomvetenskapligt intressant; värnpliktens eventuella betydelse för den politiska opinionsbildningen har varit ett outforskat område. Den inledande diskussionen om socialisationsprocessen visar en förtrogenhet med den socialisationsteoretiska litteraturen. Framställningen är systematisk och relativt överskådlig.

Men som denna recension illustrerat har avhandlingen svagheter när det gäller undersökningens metod. Enligt författarens egen sammanfattning (tabell 7.1) påstås avhandlingen ha visat att värnpliktsutbildningen påverkar 16 av 17 undersökta delaspekter av politisk kultur. Flera av dessa avser osäkra jämförelser mellan olika delkategorier inom gruppen värnpliktiga. I några fall finns en stark misstanke att resultaten endast speglar en åldersskillnad som kan förklaras på annat sätt än med värnpliktsutbildning. Dubbelanvändningen av samma data överdriver också antalet bekräftade hypoteser.

Man tvingas därmed dra motsatt slutsats än avhandlingens. Undersökningen har inte kunnat belägga att värnpliktsutbildningen har betydelse för individens politiska kultur.

Olof Petersson

Replik

Inledning

Å ena sidan är recensionens kritik saklig. Å andra sidan återges flera delar av avhandlingen ofullständigt. Framställningen är delvis så ofullständig att den i flera aspekter återger avhandlingen felaktigt. Det gäller avhandlingens design, analys och slutsats. Som en konsekvens av detta kan stora delar av recensionens kritik inte uppfattas som relevant. I ett försök att återge avhandlingen och bemöta recensionens kritik kommer detta genmäle att ägnas åt a) avhandlingens slutsats, b) avhandlingens design och analys, c) frågan om ålder och selektion samt d) operationalisering.

Avhandlingens slutsats

Den övergripande slutsatsen för avhandlingen är, som framgår av recensionen, att värnpliktsutbildningen bör inkluderas i teoribildningen om politiska socialisationsagenter (s 168). Det är dock med betydande reservation som slutsatsen formuleras, vilket inte framkommer i recensionen. Reservationen utgår ifrån de metodiska och teoretiska avgränsningar som återfinns i den empiriska undersökningen: a) sammansättning av urvalet, b) uppläggningen av undersökningen, c) insamlingsmetoden och undersökningseffekter samt d) avgränsning av teoretiska aspekter (s 168ff). Vad avhandlingen därmed genererar är en hypotes om att värnpliktsutbildningen är en politisk socialisationsagent (s 171).

Vad beträffar urval konstateras i metodkapitlet att urvalets systematiska fel medför att undersökningens resultat inte kan generaliseras utanför urvalet: "...de resultat som presenteras i den fortsatta texten som gäller specifika grupper [kommer] att enbart avse *de intervjuade* i respektive grupp." (s 51). Som konsekvens av detta fastslås det i avhandlingens slutkapitel att "...det återstår att empiriskt undersöka och fastlägga om så är fallet [att värnpliktsutbildningen påverkar de värnpliktigas politiska kultur] i andra situationer och med andra förutsättningar än de som är aktuella för den genomförda undersökningen." (s 171). Kritiken som gäller urval och möjligheten att generalisera undersökningens resultat förlorar därmed i relevans, eftersom undersökningen avgränsas till att gälla urvalet av mönstrande och värnpliktiga. En avgränsning som varken framkommer i recensionen eller som recensionen tar hänsyn till.

Avhandlingens design och analys

Enligt recensionen bygger avhandlingens design på en jämförelse mellan mönstrande och värnpliktiga. Därtill har designen, enligt recensionen, en experimentell ansats. Detta är minst sagt en sanning med modifikation. För det första är det bara en av fyra delar i analysen som bygger på jämförelsen mellan mönstrande och värnpliktiga. För det andra är det inte jämförelsen mellan mönstrande och värnpliktiga som är avgörande för att värnpliktsutbildningen betecknas som politiska socialisationsagent.

Den empiriska analysen i avhandlingen består av fyra delar. Den första delen jämför de mönstrande och de värnpliktiga som ingår i urvalet. Syftet med denna jämförelse är att få en grov bild av om och hur den politiska kultur förändras under värnpliktsutbildningen, vilket innebär att analysen inte har den experimentella ansats som antyds i recensionen.[1] Det är inte utifrån jämförelsen mellan mönstrande och värnpliktiga som värnpliktsutbildningen anses vara en politisk socialisationsagent. Jämförelsen är inte tillräcklig för slutsatser om värnpliktsutbildningens eventuella roll som politisk socialisationsagent. Det är endast som en grov indikator på utvecklingen av den politiska kulturen under värnpliktsutbildningen som jämförelsen ingår i analysen.

Det är främst två svagheter som begränsar jämförelsen mellan mönstrande och värnpliktiga. Dessa svagheter är i avhandlingen explicita, vilket inte framgår av recensionen. De diskuteras både i metodkapitlet och avslutningskapitlet (s 63 och s 169f). Den första svagheten är att jämförelsen bygger på ett antagande: att de värnpliktigas politiska kultur när de påbörjar värnpliktsutbildningen motsvaras av den politiska kultur som de mönstrande har. Svagheten i antagandet är att den politiska kulturen kan förändras hos de mönstrande under de drygt 14 månader (inte 2,5 år som hävdas i recensionen) som i genomsnitt ligger mellan mönstring och värnpliktsutbildningens start. Eftersom det saknas empiriskt underlag för att klarlägga *om* den politiska kulturen förändras och *hur* den politiska kulturen förändras under perioden inkluderar avhandlingens slutsats en reservation för antagandet.

Den andra svagheten i jämförelsen mellan mönstrande och värnpliktiga är att den endast möjliggör jämförelse på agregerad nivå. Det är två grupper av olika personer som jämförs (därav uttrycket "falsk panelstudie"). Detta innebär att individuella förändringar inte indikeras av jämförelsen. Som en konsekvens av detta reserveras avhandlingens slutsats även för denna svaghet (s 169f och s 171). En "äkta" panelstudie, som undersöker samma värnpliktiga före och efter utbildning, skulle kunna uttrycka förändringar på både individ- och agregeradnivå. Tyvärr fanns det inte praktiska möjligheter att genomföra denna typ av studie.

Den andra delen i avhandlingens analys undersöker om det finns en selektion i urvalet till värnpliktsutbildningen. Det är möjligt att det förekommer en urvalsselektion vid mönstringen, som innebär att det är mönstrande med viss politisk kultur som väljs ut till värnpliktsutbildningen. Därför jämförs de mönstrande som är uttagna till värnpliktsutbildningen med de mönstrande som inte är uttagna till värnpliktsutbildningen vad beträffar politisk kultur. Resultatet av analysen är emellertid att det, med undantag för en delaspekt, inte finns en selektion i urvalet till värnpliktsutbildningen.

Varken den första eller andra delen av analysen kan indikera aktörers roll som politisk socialisationsagent. En jämförelse mellan mönstrade och värnpliktiga är inte tillräcklig för att avgöra om värnpliktsutbildningen är en politiska socialisationsagent. Det enda som kan konstateras utifrån jämförelsen är utvecklingen på den aggregerade nivå. Vilka faktorer som påverkar de värnpliktigas politiska kultur kan inte identifieras utifrån denna typ av jämförelse. Syftet med analysens tredje del är därför att undersöka om värnpliktsutbildningen påverkar de värnpliktigas politiska kultur och därmed intar rollen som politisk socialisationsagent. Det som är avgörande i analysen för att värnpliktsutbildningen ska betraktas som politisk socialisationsagent är att det finns samband mellan aspekter av värnpliktsutbildningen och de värnpliktigas politiska kultur. Detta framgår av den analytiska strategi som formuleras i metodkapitlet (s 57f). Det finns med andra ord en medvetenhet om skillnaden mellan statistisk signifikans och teoretisk signifikans. Denna medvetenheten återspeglas inte minst i respektive analysavsnitt, som innehåller en relativ tydlig åtskillnad mellan teoretisk nivå och operationell nivå.

I avhandlingens analysmodell ingår fyra aspekter av värnpliktsutbildningen: försvarsgren, värnpliktskategori, förbandstillhörighet och utbildningstid. De tre första aspekterna betraktas som innehållsliga aspekter, medan den sista aspekten avser exponeringstid. Att valet av dessa aspekter kan förefalla vara ad-hoc-artad, vilket recensionen ger uttryck för, är överraskande. Elementär kunskap om värnpliktsutbildningen är att utbildningen i sin planering är uppbyggd kring dessa aspekter.

I analysens fjärde del undersöks om andra aktörer utanför värnpliktsutbildningen påverkar de värnpliktigas politiska kultur. Avhandlingens analysmodell inkluderar sju aktörsgrupper, som den befintliga teoribildningen betecknar som politiska socialisationsagenter: familjen, vänner och bekanta, organisationer, television, tidningar, arbete och gymnasieutbildning. I den fjärde delen ingår analyser av samband mellan kontakten med dessa aktörer och de värnpliktigas politiska kultur. Den analytiska strategin för att avgöra om en aktörsgrupp ska betraktas som politisk socialisationsagent motsvarar strategin för värnpliktsutbildningen: om det finns samband mellan kontakten med aktörsgruppen och de värnpliktigas politiska kultur, så kommer aktörsgruppen att betraktas som politisk socialisationsagent. Resultatet av den empiriska undersökningen är att det aktörsgruppernas betydelse varierar med vilken delaspekt som undersöks.

En strävan i avhandlingens analys är att fastställa vilken relativ betydelse som respektive agent har. Med detta tar analysen hänsyn till dels att individen har kontakt med flera agenter samtidigt, dels att kontakterna med agenterna kan vara beroende av varandra. Det kan till exempel vara så att värnpliktiga uttagna till gruppbefäl i större utsträckning är medlemmar i organisationer och läser tidningen oftare än värnpliktiga som är meniga. En skillnad mellan värnpliktiga gruppbefäl och meniga skulle då inte förklaras av värnpliktskategori, utan av medlemskap i organisationer och konsumtion av tidningar. För att klargöra eventuella interkorrelationer genomförs tre analyser. Den första studerar sambandet mellan aspekter i värnpliktsutbildningen (Appendix C1). Sambandet mellan kontakter med aktörer utanför värnpliktsutbildningen studeras i den andra analysen (Appendix C2). Den tredje analysen studerar sambandet mellan aspekter i värnpliktsutbildningen och kontakter med aktörer utanför värnpliktsutbildningen (Appendix C3).

Tillsammans indikerar de tre interkorrelationsanalyserna att det endast finns ytterst svaga samband mellan de ingående indikatorerna, som i de flesta fallen dessutom inte är signifikanta (s 58ff). Som en konsekvens av detta kan indikatorerna betraktas som oberoende av varandra.[2] Att genomföra multivariat korrelationsanalys med oberoende faktorer på nominal nivå som inbördes är oberoende av varandra och dessutom har variation i antalet kategorier (nivåer) är inte fruktbart.[3] Det ger ingen ny information om faktorerna relativa betydelse. Faktorernas relativa betydelse motsvaras därför av de samband som framkommer av de bivariata analyserna. Detta är också resultatet av de multivariata analyser som har genomförts av det insamlade underlaget, men som inte redovisas i avhandlingen.

Ålder, selektion eller socialisation?

Ett argument som framförs i recensionen är att skillnaderna mellan mönstrande och värnpliktiga som framkommer i analysen beror på åldersskillnader och inte på socialisation. Även om jämförelsen mellan mönstrande och värnpliktiga inte är avgörande i analysen, vilket har konstaterats tidigare, finns det ändå anledning att bemöta argumentet. En utgångspunkt för studier i politisk socialisation är att det finns skillnader mellan åldersgrupper. Att då förklara skillnader mellan åldersgrupper med ålder blir tatulogiskt. Ålder eller tid kan inte vara en förklaring till förändring av politisk kultur. Istället måste förklaringen finnas i det som händer i tiden. Att klargöra vad det är som händer i tiden och som påverkar politisk kultur är därför ett syfte ned forskningen om politisk socialisation. Detta resonemang återfinns i avhandlingens teorikapitel (s 42), vilket inte beaktas i recensionen.

Det finns en möjlighet att delvis pröva tesen om ålder på det urval som ingår i undersökningen. Flertalet av dem som ingår i urvalet av värnpliktiga (drygt 80 procent) är i åldern 20 och 21 år vid utbildningens slut. Men utbildningstid varierar inom åldersgrupperna, vilket innebär att perioden mellan mönstring och värnpliktsutbildning också varierar. Om tesen, som framförs i recensionen, är korrekt, skulle det förväntade resultatet vara att de värnpliktiga som började värnpliktsutbildningen senare (äldre: 21 år) skulle ha en annan politisk kultur än de som påbörjade värnpliktsutbildningen tidigare och som har samma utbildningstid (yngre: 20 år), eftersom de som påbörjade senare har haft ett år extra mellan inskrivning och utbildningen än de som började tidigare samt att de under detta år kan ha blivit påverkad av (till exempel) andra agenter. Resultatet visar emellertid att så inte är fallet. Det finns ingen skillnad mellan åldersgrupperna 20 och 21 år vad beträffar deras politiska kultur.[4] Resultatet gäller oavsett vilken utbildningstid som studeras.

Ett annat argument som framförs utan empiriskt underlag i recensionen är att de skillnader inom värnpliktsutbildningen som framkommer i analysen beror på selektion. Urvalet till värnpliktsutbildningen skulle, enligt recensionen, vara beroende av vilken politisk kultur som de mönstrande har. Det är möjligt, men inte troligt att urvalet skulle innebär en selektion baserad på politisk kultur. För det första finns det inget i de tester och undersökningar som genomförs vid mönstringen som är direkt relaterade till den politiska kulturen hos de värnpliktiga eller kontakter med politiska socialisationsagenter. För det andra visar analysen av sambandet mellan aspekter i värnpliktsutbildningen och kontakter med aktörer utanför värnpliktsutbildningen att det endast finns ytterst svaga samband (Appendix C). Det är därför inte troligt, som exempel, att de uttagna till ett viss förband har en speciell socialisation, eftersom deras kontakt med politiska socialisationsagenter inte avviker från värnpliktiga vid andra förband. För det tredje visar analysens andra del att det inte finns någon selektion vad beträffar urvalet till värnpliktsutbildningen. Tyvärr erbjuder inte materialet möjligheter att undersöka om urvalet till typ av värnpliktsutbildning innebär en selektion. Främsta orsaken till detta är att det inte fanns möjligheter att testa tesen att urvalet till förband innebär en selektion (även om det i recensionen hävdas att så är fallet). Med andra ord, även om det finns skäl att uppfatta tesen som orimlig, så återstår det att undersöka om tesen har en empirisk giltighet. Tesen om selektion kan därför inte empiriskt undergräva avhandlingens slutsats. Det går inte heller att avvisa tesen med empiriskt underlag.

Operationalisering

Eftersom indexen för politisk affektion och politisk evaluation baseras på samma frågor som indexen för politisk kognition så är, hävdar recensionen, framställningen missvisande. Frågan gäller emellertid inte om samma empiriska underlag används dubbelt, utan hur de tre begreppen är relaterade. Ett alternativ, som möjligen recensionen utgår ifrån, är att de tre begreppen refererar till egenskaper som är separerade ifrån varandra. Utifrån denna relatering är det missvisande att använda samma empiriska underlag för indexen. Avhandlingen har dock en annan relatering av begreppen. Enligt avhandlingens analysmodell är politisk kognition, som avser individernas medvetenhet om politiska objekt, en förutsättning för politisk affektion och politisk evaluation. Om vi saknar medvetenhet om objekten kan inte vårt mentala förhållningssätt till objekten laddas med känslor (affektion) eller omdömen (evaluation), då objektet inte existerar mentalt för oss. Av detta följer att individens politiska kognition får genomslag på individens politiska affektion och politiska evaluation. Att då konstruera index som *inte* tar hänsyn till denna relation skulle vara direkt missvisande. De konstruerade indexen är därför inte missvisande i förhållande till hur begreppen definieras och relateras, eftersom de tar hänsyn till relationen mellan begreppen.

Recensionen är också kritisk till att operationaliseringen inte baseras på empiriska stöd för skalbarhet. Detta är en svaghet. Avhandlingen skulle med fördel kunna kompletteras med analyser som prövar skalbarheten hos indexen. Inte minst skulle detta kunna klarlägga dels hur olika aspekter i den politiska kultur är relaterade till varandra, dels skapa underlag för utveckling av typologi vad beträffar både politisk kultur och politisk socialisation. Å andra sidan är det intressant att recensionen inte tar upp frågan om huruvida indexen indikerar de egenskaper som avses med respektive begrepp. Det är denna problematik som uttryckligen har varit vägledande för operationalisering. Index har konstruerats utifrån ambitionen att de ska indikera de egenskaper som avses med begreppen. I avvägningen mellan begreppsvaliditet och empirisk skalbarhet har avhandlingen möjligen en alltför ensidig strävan att uppfylla begreppsvaliditeten. En given framtida uppgift är därför att pröva och studera den empiriska skalbarheten hos de ingående indikatorerna.

Thomas Denk

Noter

1. Att mönstrande kallas för "kontrollgrupp" är inte tillräcklig intäkt för att hävda att analysen har en experimentell ansats.

2. Det finns tre fall av interkorrelation. Två av fallen är inte problematiska eftersom indikatorerna avser att mäta kontakten med samma aktörsgrupp (vänner och bekanta respektive tidningar). Det tredje fallet gäller sambandet mellan utbildningstid och värnpliktskategori. Efter en kompletterande analys framkommer det att utbildningstiden inte har någon betydelse när värnpliktskategorin är konstant. Därför exkluderas utbildningstid som faktor i analysen (s 58).

3. Det är till och med möjligt att två nominella faktorer vars fördelning är fullständigt korrelerade med varandra kan ha relativt oberoende betydelse på en tredje faktor om antalet kategorier hos faktorerna varierar tillräckligt. Det finns, som exempel, ett fall i den genomförda undersökningen där ett fullständigt samband mellan två nominella faktorer inte har betydelse för analysen eftersom faktorerna har olika antal kategorier. Mellan faktorerna försvarsgren och

förbandstillhörighet finns ett fullständigt samband. Men eftersom faktorn försvarsgren har tre kategorier och förbandstillhörighet har fem kategorier har sambandet ingen större inverkan på kategoriernas relativa betydelse.

4. Med detta avses att det inte föreligger något signifikanta samband mellan ålder och indexen för politisk kultur i respektive utbildningsgrupp.

Slutreplik

Thomas Denks replik innehåller ett antal direkta eller indirekta medgivanden.

1. Resultaten kan inte generaliseras utanför urvalet.

2. Trots att avhandlingen talar om en "kontrollgrupp" har analysen inte någon experimentell ansats.

3. Jämförelsen mellan mönstrande och värnpliktiga begränsas av stora svagheter.

4. Avhandlingen studerar inte några individuella förändringar, varför det är inte är fråga om någon panelstudie.

5. Mothypotesen om selektion i urvalet till värnpliktsutbildningen har endast berörts genom en enkel jämförelse avseende skillnader i den beroende variabeln, men har inte prövats genom någon multivariat kausalanalys.

6. Även om avhandlingen förvisso nämner åtskillnaden mellan teoretisk nivå och operationell nivå görs ingen skillnad mellan teoretisk och statistisk signifikans vid den empiriska hypotesprövningen.

7. De fyra aspekterna av värnpliktsutbildningen har valts ut eftersom utbildningen i sin planering är uppbyggd kring dessa aspekter. Då variablerna sålunda inte har härletts ur någon teori framstår urvalet som ad hoc-artat.

8. Interkorrelationerna mellan de oberoende variablerna är inte så låga att en multivariat analys kan negligeras.

9. Avhandlingen har inte utnyttjat möjligheten att analysera åldersskillnader. Den beräkning som nu plötsligt dyker upp i repliken förekommer inte i avhandlingen.

10. Dubbelanvändningen av data återspeglar en fundamental oklarhet om hur avhandlingens centrala begrepp logiskt och empiriskt är relaterade till varandra.

11. Det är en svaghet att indexens skalbarhet inte baseras på empiriska stöd.

Olof Petersson

BJÖRN HOLMBERG: *Passing the Open Windows. A Quantitative and Qualitative Approach to Immediate Military Balance and Escalation of Protracted Conflicts.* Report 27. Department of Peace and Conflict Research, Uppsala University, 1998.

Do Open Windows Encourage Conflict?[1]

1. Realist Theory and Military Opportunity

Changes in the balance of strength between two or more parties to a conflict are central to several realist theories of international relations, such as balance of power theory and theories of the power transition. Björn Holmberg's doctoral dissertation discusses whether changes in military strength and capabilities are likely to lead one of the parties to perceive a 'window of opportunity', which in turn is likely to increase the probability of escalation. The central question in the dissertation is formulated in the following way (p. 8): 'Do rapid shifts in the immediate military balance (that is, military opportunities) cause escalation of protracted conflicts, and if so, under what circumstances?'[2]

Holmberg's starting point, realist theory, is not a unified structure, but rather a paradigm which can be used to formulate several alternative realist theories. According to Holmberg, these theories are united by three ideas: (1) the state as the key actor in the international system and the lack of a supranational authority, (2) the state as a rational actor which strives to maximize the expected utility of its actions, and (3) power as the key tool in international relations. Beyond this, realist theories differ on such important points as whether to see states as power maximizers or security satisficers. Holmberg argues that rapid changes in military strength are likely not just to give new opportunities to change-oriented states, but also to threaten the security of status-quo oriented states. In this way he avoids having to make a choice between these two varieties of realism.

On the other hand, Holmberg breaks with realist theory when he extends its perspective beyond the relationship between the great powers. Most realist writers have described realism as a paradigm for the relationship between the great powers. However, if realists want to claim general validity for their theory, Holmberg argues, they should be able to explain relations between small and medium-sized states as well. Waltz and other key realist theoreticians have assumed that the same theoretical mechanisms must regulate the relationship between smaller states when these relationships are insulated from the major powers.

In order to answer his major question, Holmberg carries out a two-step empirical analysis. First he makes a quantitative analysis of military opportunities during the Cold War and case-studies of the relationship between India and Pakistan and between Iran and Iraq. He asserts that quantitative and qualitative studies each have their strong and their weak sides and that the best results will be obtained by combining the two methodologies, citing Blalock (1979) and King, Keohane & Verba (1994) in favor of this view.

In the following three sections I will review Holmberg's argument with a minimum of clarification and comment. I then discuss three major reservations with regard to theory, data, and methodology.

2. The Theoretical Model

Holmberg develops a fairly elaborate theoretical model with 'window of opportunity' as a key variable. This concept, said to have been invented by a Pentagon analyst in 1978, became a key term in security policy in the New Cold War in the early 1980s. US critics of détente, who came to power under President Reagan, interpreted the Soviet arms drive as an attempt to obtain a first-strike capability. Key decision makers in the new administration argued that the USA had to be able to respond in kind in order to prevent the Soviet Union from obtaining a window of opportunity which the country might want to (or even feel pressured to) exploit.

The concept 'window of opportunity' is not unambiguous, as Holmberg illustrates by means of Figure 1. Initially there is an approximate balance of power between the two. From T3 the power relationship becomes skewed and A obtains an advantage over B. From T5, A declines and at T8 the relationship is once again fairly balanced. In the terminology of Lebow (1984) A had a window of opportunity in the period T5 to T8. Snyder (1989), on the other hand, emphasizes B's window of opportunity in the period T1 to T3, assuming that B realizes that its opponent is likely to increase its military strength after T3. State B has to strike *before* state A can arm significantly.

Going beyond Holmberg's own discussion for a moment, we might speculate on the consequences of states reacting to *any* actual or prospective change in the military power relationship. Either your own state has an advantage, in which case you should exploit it while there is time and before your opponent can even the score. Or you fear that your opponent will change the power relationship, in which case you have to strike before he does. This implies that the actual power relationship has very little significance for the danger of preventive war. It is the parties' *perceptions* of the danger of a change that is the decisive factor, as long as you assume an anarchic international system. The parties do not trust one another, and the danger of preventive war is omnipresent. In that case, the danger of escalation would become a constant rather than a variable, and the question of factors which may influence the probability of escalation would be irrelevant.

Even if Holmberg does not pursue the implications this far, he does conclude that 'window of opportunity' is too broad a concept to be useful in an empirical study. Instead, he introduces the concept of *military opportunity*, defined as 'a

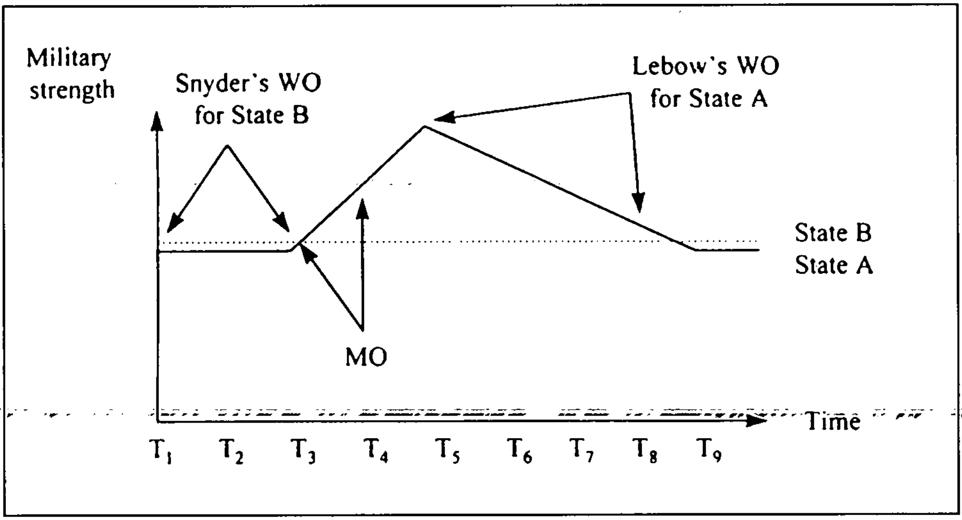

Figure 1. Window of Opportunity and Military Opportunity. The unbroken line represents the power level of A, and the dotted line is B's power level.
Source: Holmberg, 1998: 25, Figure 2.

rapidly evolving military advantage' (p. 25), corresponding to the time period from T3 to T4 in figure 1.[3] State A has a military opportunity during this period and state B has a military vulnerability.

Holmberg wishes to *combine* elements from a rational actor approach with organization theory, in stated contrast to Allison (1971) in his classical analysis of the Cuban Missile Crisis. Holmberg assumes that rational state actors have the survival of the state as their highest-ranking goal and that the more militarized a conflict is, the greater is the threat to the state. From this, he concludes that the more militarized the conflict, the greater the danger of escalation. Figure 2 indicates that conflicts in the upper half of the ladder are characterized as militarized. One consequence of this, not explicitly stated, is that the higher a conflict has been escalated, the greater the probability of a further escalation. This hypothesis is strengthened by a point of view which Holmberg identifies with organization theory. He assumes that the military organization strengthens its influence in the decision making process once the conflict has become militarized. This contributes to increasing the probability of escalation with military means, assuming that the military opportunity remains.

Holmberg also draws on learning theory. He assumes that experience from earlier conflicts provides the actors with experience in the use of different tools of conflict behavior. Positive as well as negative learning builds on experience in previous escalations. Winning a war is a positive experience, while losing it involves a negative experience. Because learning is a characteristic of the nation (i.e. a variable measured at the monadic level) Holmberg feels that this variable is inappropriate in a quantitative test on dyadic data and saves it for his later qualitative analysis.[4] However, the main emphasis in the empirical analysis is on the two other frameworks.

3. Testing the Model

In order to operationalize the concept of military opportunity Holmberg might have followed in the steps of strategic analysts and looked at *the size of standing forces* and the presence or absence of particular weapons systems. In the early 1980s, the development of new weaponry, such as the neutron bomb, intermediate-range missiles, and the SDI, stood at the center of the US debate on windows of opportunity Alternatively, in the tradition of arms race studies, he

Table 1. Events which Create Military Opportunities between A and B.

A. Events which weaken state B
1. State B experiences an internal revolt or escalated civil war
2. State C, a third party, launches an offensive against State B
3. The alliance State B has with State C is terminated
4. State C dramatically reduces its military presence in State B

B. Events which strengthen state A
5. State A receives extensive deliveries of military aid during a short period of time
6. State A forms an alliance with State C
7. State C dramatically increases its military presence in State A

C. Events which assume armed conflict between A and B
8. State A defeats large parts of B's forces

Source: Holmberg, 1998: 57-58.

might have focused on *the development of the arms spending* of the two countries. Instead, he lists eight relevant events which can create a military opportunity (Table 1):

Although the list is quite comprehensive, it does not seem to be theoretically derived, and one wonders if it is exhaustive. When he gets to the problem of operationalization (p. 57), Holmberg leaves out nos. 5 and 8 because they would involve extensive additional data collection and coding. As for no. 8, the Correlates of War project (on which Holmberg generally relies) does not provide data on the size of the military forces and does not even give data on military losses disaggregated to the annual level. The figures for military losses in this data set include the entire war and thus cannot be used in an analysis of escalation. However, the remaining six events can be operationalized by means of the Correlates of War project data on civil war and interstate war, together with some additional data on the escalation of civil war and the military presence of a third state.

A military opportunity for A automatically involves a military vulnerability for B. A rival hypothesis suggested in the dissertation is that it is not State A, but State B which attacks during the military opportunity. The logic is that B is afraid that its military situation will be weakened even further and feels bound to intervene while the situation may yet be saved.

The dependent variable is the *escalation* of the conflict. An escalation is conventionally understood as the increase in the severity of the conflict behavior by one or both parties to a conflict. Escalation means crossing some threshold. It is a stepwise process from the first initiation of conflict behavior to the most severe forms of conflict behavior in war. Holmberg briefly flirts with the idea that geographical proliferation of a conflict can also represent a form of escalation, but he sets this aside and concentrates on increases in intensity.

On the basis of on the Militarized Interstate Disputes (MID) data set of the Correlates of War project[5], Holmberg pictures an escalation ladder with five steps above 'no dispute' (Figure 2). These steps have been taken directly from the MID data, except for the top step: 'strategic offensive', by which Holmberg means 'a decisive move to achieve a military victory' (p. 60). The operational criterion here is whether a party will deploy a major part of his forces in order to settle the war once and for all.

In order to study the escalation process empirically, Holmberg looks at what he calls *protracted conflicts* – conflicts drawn out in time – and selects two versions of such conflicts. The

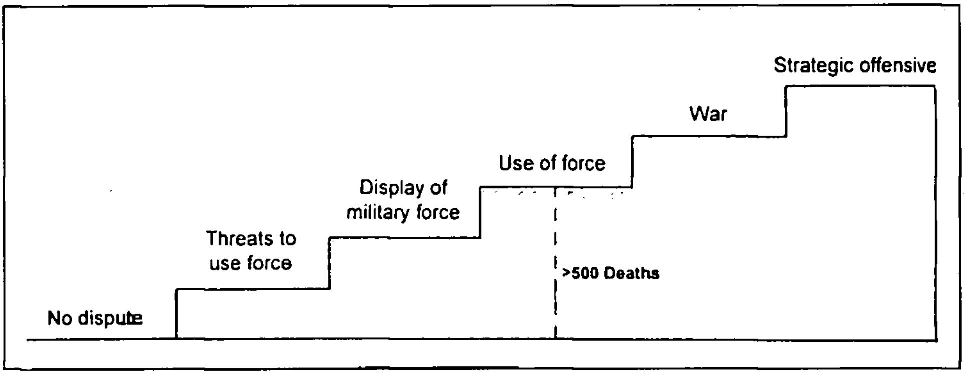

Figure 2. The Escalation Ladder. Source: Holmberg, 1998: 59, Figure 8.

first is all wars which extend over a period of at least two years. This criterion brings three wars into his empirical domain. Secondly, he includes 'enduring rivalries' (Goertz & Diehl, 1993) – extended conflictual relationships which do not necessarily exhibit active conflict behavior at all times. In the literature we find somewhat different definitions of enduring rivalry which build on the MID data set. Holmberg settles for a definition from Wayman & Jones, which he cites from Geller (1993)[6], and which yields 10 enduring rivalries. Of his total of 13 protracted conflicts, he excludes 4 because they are likely to influence each another. For instance, the conflict between Afghanistan and Pakistan is left out because it is partly overlapping in time and space with the one between India and Pakistan, and the latter is the longer conflict. This rather elaborate selection process leaves him with nine protracted conflicts which are listed in Table 2.

The unit of analysis in the quantitative study is the dyad year, giving an n of 260, adequate for statistical analysis. Of these dyad years, 30 satisfy at least one of the criteria for military opportunity.

Holmberg introduces several control variables. The theory of the democratic peace leads us to expect that, for dyads consisting of two democracies, escalation to a high level of violence is unlikely. Another control variable is included for statistical reasons: Beck, Katz & Tucker (1998) and others have pointed out that, in studies using the dyad year as the unit of analysis, lack of independence between adjacent observations can induce overconfidence or deflated standard errors for the coefficient estimates.

Table 2. Protracted Dyadic Conflicts.

Dyad	Duration	Category
Vietnam-Cambodia	1975-79	Protracted war
Ecuador-Peru	(1858) 1945-86	Enduring rivalry
Greece-Turkey	1958-86	Enduring rivalry
India-Pakistan	1947-86	Enduring rivalry
Iran-Iraq	(1934) 1945-86	Both
North Vietnam-South Vietnam	1973-75	Protracted war
Somalia-Ethiopia	1960-86	Enduring rivalry
Syria-Israel	1948-86	Enduring rivalry

Source: Holmberg, 1988: 56, Table 4.

Everything else being equal, the probability of conflict between two countries in year *t* is greater if there was conflict between the same two countries in the year *(t-1)*. Holmberg corrects for such temporal dependence by including the dependent variable with time-lags of one, two, and three years. Since he does not use conflict but rather the *escalation* of conflict as the dependent variable, he cannot be completely sure that the problem is solved through this correction. However, assuming that a conflict is unlikely to be escalated in successive years, Holmberg is probably well insured against temporal dependence.

Table 3. Escalation and Military Opportunity, Nine Protracted Conflicts (% of Dyad-years).

		Military Opportunity?		
		Yes	No	All
	Yes	10	6	7
Escalation?	No	90	94	93
	All	100	100	100

Calculated on the basis of Holmberg, 1998: 66; n = 260, p = .41.

4. Results

Given how much work Holmberg has put into his theoretical model and in the operationalization of his variables, his quantitative analysis is surprisingly brief. He first performs a bivariate table analysis (p. 66), which shows that most conflicts are *not* escalated. However, this table does not show what fractions of conflicts are escalated in conflicts where there is no military opportunity. By calculating this on the basis of other information provided in the dissertation, I obtained Table 3. The probability of conflict when a military opportunity is present is nearly doubled, but the difference is not significant – hardly a surprise in view of the small numbers of military opportunities observed.

There is no positive relationship between military vulnerability and escalation. In the subsequent bivariate logistic regression analysis, Holmberg merges the presence of a military opportunity and a military vulnerability, so that the effects of the two can no longer be separated. The combined variable is not significant, and Holmberg concludes that the bivariate analysis provides no support for his model.[7]

Proceeding to the multivariate logistic regression analysis, he adds an interaction term between militarization two years earlier and the combined variable for military opportunity or vulnerability. This interaction emerges as the only one that comes close to being significant. Most of the other variables have *p*-values above 0.5. This finding I interpret to mean than once a conflict has gone beyond a certain level and military means have been employed, military opportunity does contribute to further escalation.

In an attempt to decompose this interaction effect, Holmberg takes a special look at the interaction effect between military opportunity and escalation for a low degree of militarization, i.e. below the level of force characterized as war. Here he finds a clearly significant relationship between military opportunity and escalation; this relationship becomes even stronger when he filters out the democratic dyads. However, the relationship is not significant at the highest level of escalation. He explains this by saying that events in a war may proceed so quickly that a model built on annual changes is not very satisfactory.[8]

Holmberg concludes that a model built solely on realist assumptions does not do very well. Realist hypotheses must be supplemented by other assumptions, in particular assumptions derived from organization theory.

Following the quantitative analysis, Holmberg proceeds to case-studies of India vs. Pakistan in 1970–71 and Iran vs. Iraq in 1979–88. These two cases are chosen deliberately to include military opportunities as well as escalation. For the first conflict, where a single year is studied, both are present. The military opportunity for India consisted of an outbreak of civil war in Pakistan, eventually leading to the breakaway of Bangladesh. India had been in conflict with Pakistan since the partition at independence in 1948. By intervening in the civil war, India could acquire a decisive strategic advantage relative to its arch-rival. In the other conflict Holmberg finds two cases of coinciding military opportunity and escalation. One occurred in 1981, when civil war broke out between the Mujahedin and the government in Iran. The other was in 1985, when the conflict between the Iraqi government and the Kurds escalated. In both cases, one of the two warring parties was significantly weakened and the other party saw an opportunity to get on the offensive.

Holmberg concludes that India did see a military opportunity when civil war broke out in Pakistan. That the war ended relatively quickly, Holmberg attributes to pressure from the great powers. Perhaps India also lacked the will to give Pakistan a decisive military defeat and rested content with having ensured that the partition of Pakistan remained permanent. In line with an alternative hypothesis which I have not discussed, Holmberg holds that India in this case replaced further escalation with a de-escalation of the conflict once the initial goal had been attained.

Holmberg finds that Iraq clearly had a military opportunity vis-a-vis Iran in the first phase of the war. The military opportunity for Iran in 1985 which Holmberg discerns in the quantitative analysis is, however, rejected in the qualitative analysis, because Iraq did not need to set aside much of its military power – one out of 32 military divisions – to combat the Kurdish revolt.

Right at the end Holmberg tries to combine lessons from the quantitative and the qualitative analysis. He asserts (p. 115) that 'realism was ... partly defeated on its own turf', although his most important aim was not to combat realism, but to determine under which circumstances there might be a relationship between military opportunity and escalation.

Holmberg finds that a degree of militarization strengthens the relationship between military opportunities and escalation, in line with his theoretical expectations. In the qualitative analysis he finds three military opportunities in the course of the two wars that he subjects to special scrutiny. Only one of them is accompanied by escalation, while two are accompanied by de-escalation and an end to the armed conflict. Here he suggests that the same cause may have multiple effects. However, he is unable to identify general conditions for when one or the other outcome is more likely. Holmberg suggests that his analysis supports the stress that neorealism places on the security of the state, in contrast to the traditional realist emphasis on the expansion of power under all circumstances.

The hypotheses derived from organization theory do not find much support, according to Holmberg. The role of military in the decision making process does appear to increase with the militarization of the conflict, but this does not necessarily imply that military opportunities are pursued with greater vigor. Holmberg thinks that military authorities may have failed to interpret these situations as military opportunities even if they fulfill the quantitative criteria. Therefore, he does not reject the assumptions from organization theory, but he feels that improved data on the decisionmaking processes may form a better basis for successful hypothesis testing.

Holmberg concludes that his two-pronged methodological approach has withstood its test. 'Methodological pluralism' emerges as 'a winner' (p. 121). The central tenet of realism – the significance of military opportunities – has failed, although a threatened state may imply a greater threat to its environment. Holmberg assumes that his model can also be applied to domestic conflict, the numerically dominant form of conflict today (Wallensteen & Sollenberg, 1998). As far as interstate conflict is concerned, the statistical data should be extended to the period prior to World War II; there is also a need for more detailed case-studies. Likewise, military preferences should be studied more closely. The bottom line, Holmberg suggests, is that military windows of opportunity do exist, but that most states do not leap through them.

5. Some Praise and Some Reservations

Holmberg has taken hold of a problem in international relations which is of obvious academic as well as political interest. Despite the increased standing of neoliberal ideas in recent years (Kegley, 1993), we should welcome an attempt to dissect the precise assumptions of the realist school and test them, as Holmberg has done here. Holmberg has put a great deal of effort into the statistical testing as well as the case-studies. He has not shied away from the task of mastering an extensive quantitative literature as well as more specialized literature on his two cases. Most doctoral candidates would probably have felt satisfied to review one of these literatures. This being said, I have three major reservations[9] – on theory, data, and methodology.

5.1 Theory

In his test of realist theory, Holmberg wishes to make an original contribution by applying the theory to the relationship between medium-sized and small states. He quotes twice (p. 20, p. 126) Kenneth Waltz (1959: 73) who has argued: 'The theory once written also applies to lesser states that interact insofar as their interactions are insulated from the intervention of the great powers of a system, whether by the relative indifference of the latter or by the difficulties of communication and transport.' It is difficult to reconcile this with Holmberg's list of protracted conflicts (Table 1). For instance, Holmberg puts the starting-point for the war between North Vietnam and South Vietnam at 1973. By then, however, there had been more or less continuous war in Vietnam for 27 years – first a colonial war ending in a partitioning of the country along a border drawn by the great powers, then war between the two parts of the country with increasing foreign participation, eventually to the point where the COW project defines the war as interstate. North Vietnam fights this war against one of the two superpowers with extensive material support from the other. It is hard to interpret this as 'interaction ... insulated from the intervention of the great powers'. The war between Vietnam and Cambodia 1975–79 was also heavily influenced by superpower rivalry in the region. For most of the period studied here, Somalia and Ethiopia were solidly positioned on separate sides of the Cold War, although after the fall of Haile Selassie they switched sides. Several of Holmberg's protracted conflicts take place in the Middle East, where the great powers have long vied for influence. And so on down the list. In fact, only for two of the conflicts (Chile vs. Argentina and Peru vs. Ecuador) does it make sense to see them as purely local conflicts, and then only if you ignore the Monroe Doctrine and the pervasive influence of the USA on the entire region. If this data set is to be used to test realist theory on smaller powers, the justification can hardly be found in the quotation from Waltz.

Holmberg is not immune to such ideas. For instance, the conclusion or suspension of a military alliance with a third country is one of the events which can give rise to a military opportunity. In his final chapter he also discusses briefly the US involvement in the Gulf War of 1987–88 and how the positions of the major powers may have influenced the war between India and Pakistan in 1971. But he does not squarely face the problem that relations between the partners in these protracted conflicts influence the course of events throughout the conflict.

Björn Holmberg would probably have been better off by selecting a much larger sample, even the universe of dyads over the same period, and explicitly taking account of major power status or relations to major powers as a separate variable. Bremer (1992) finds 'major power' to be a useful predictor of dyadic armed conflict, as do Raknerud & Hegre (1997). Membership in the same, or opposing, alliances might have been another useful variable, as would a variable indicating whether or not an allied country is at war in the relevant year. This would have required a more complex model, but it would also have allowed Holmberg a much larger *n*, so that more of his relationships might have proven to be significant.

5.2 Escalation

The escalation ladder of Figure 2 has several steps. However, Holmberg uses only two of them. He defines escalation as one of two actions: Either the use of force resulting in more than 500 deaths, or the introduction of a strategic offensive. Thus, most of idea of a stepwise process is lost. If you are already in a war, you can escalate only once. If you are not, you can escalate twice – first by engaging in actions which lead to more than 500 battle-deaths, and subsequently by engaging in a strategic offensive. In fact, there are only 17 escalations in the data, and they occur in only 6 of the conflicts. In three of the nine conflicts, there is no escalation at all. Chile and Argentina were in a protracted conflict for 33 years, Ecuador–Peru for 42 years, and Greece and Turkey for 29 years. Not once during these 104 dyad years was there a single case of escalation! Escalation is measured differently for enduring rivalries (outbreak of war) and for protracted wars (strategic offensive), and only the Iran–Iraq conflict is both. Thus, for most conflicts, only one escalatory step occurs, and de-escalation is not measured. Surely this is a rather limited concept of escalation.

The main reason why Holmberg has difficulties in measuring escalation lies in the source of his conflict data. The Militarized Interstate Disputes data do, of course, measure the intensity of a conflict on a ladder, as shown in Figure 2. However, any MID is characterized by the highest level of force engaged in by the actors. Thus, the only way to measure escalation in a MID is by comparing the force level of an initiator with the force level of the other state. In real life, we observe conflicts which start with a threat and are met with, say, a show of force, whereupon the first state resorts to force, which may in turn be reciprocated. Here we have a series of escalatory moves which may or may not end in full-scale warfare. However, the way the MID data have been coded, such a sequence of actions is not possible. In a MID that lasts for two years, the force level of the entire MID is defined by the highest level of force engaged in. This has some curious consequences. For instance, the interstate war data set of the Correlates of War project records the Spanish–American war as having started on 21 April 1898. In the MID data set, the war starts over three months earlier, on 3 January. The militarized dispute that started on that date eventually escalated into a war, and so it is coded as a war right from the start. Worse, the Mexican–American war starts more than three years prematurely in the MID data set, on 12 May 1843, whereas the starting date is identified as 23 August 1846 in the interstate war data set.[10]

None of the conflicts studied in this dissertation seem to suffer from such major errors in timing. However, the problem remains that the MID data are inherently unsuitable for studying escalation or de-escalation within a MID which does not escalate to war.

Actually, there does exist a more detailed data set which might perhaps have been used in this study, the Incidents data set. The first published description of the MID data set (Gochman & Maoz, 1984: 589) reveals that the basic building blocks of the MIDs are individual military acts, or 'incidents', which are grouped together 'into temporally bounded disputes', according to three rules: First, the same or an overlapping set of actors have to be involved. Second, the incidents have to involve the same issue or set of issues; and third, the elapsed time between the incidents cannot exceed six months. A dispute has been coded as ended 'when there were no codable incidents for a period of six months or when a war terminated'.

No data on incidents are found in the MID computer file from the Inter-university Consortium on Social and Political Research (Singer & Small, 1994). Neither are they available on the Peace Science Society homepage on the web,[11] from which the MID data can also be downloaded. Björn Holmberg may be forgiven for not realizing that this data set exists. Indeed, I shared his ignorance until I stumbled across Carlson (1995), who tests a theory of escalation on the MID data. Correspondence with that author revealed that she had in fact used the Incidents data set, but only for conflicts involving at least one major power (p. 526).[12] I have checked the original Incidents data set for Holmberg's conflicts.[13] In fact, it contains very limited information about the incidents in the relevant MIDs. This is reassuring, in the sense that Holmberg has not missed a major source of information. On the other hand, had he realized that almost no incidents are recorded for the protracted conflicts discussed in this publication, he might have second thoughts about the reliability of the dependent variable.

The Correlates of War data set on interstate wars contains information on events that most informed people have already heard about and which they can check in encyclopedias and historical reference works. Any error in that data set is likely to have been discovered and corrected a long time ago. The MIDs, however, frequently refer to little-known events.[14] In the MID data set very little information is given about each MID, so it is extremely difficult to identify it on the basis of standard sources. The overall reliability of the data set is unknown, probably unknowable. For analyses involving hundreds of MIDs this is probably not too serious, since the coding errors are unlikely to crowd out the correct information. But for studies such as Holmberg's dissertation, where the *n* balances on the edge of what can be analyzed statistically, such uncertainty about the quality of the data is very problematic.

Fortunately, apart from the three wars, the conflicts studied by Holmberg do not consist of a single MID. The Ecuador–Peru case, for instance, contains no less than 20 MIDs over the 42-year period. In principle, then, escalation should be measurable. Whenever you find a MID in one year and no MID in the year before, that means that someone has escalated the conflict. Even if you have MIDs back to back (which does not occur in the Ecuador–Peru conflict) one MID could be at a higher level than the previous one. But Holmberg limits his ability to measure escalation in this way by requiring that the conflict must have escalated to the use of force involving more than 500 dead. The Ecuador–Peru conflict is a century-old disagreement about the border. From time to time, the conflict has led to border clashes, frequently involving the deaths of soldiers. By any standard of ordinary language, there have been a number of escalations and de-escalations in the period 1945–86. Yet none are recorded in this study.

I conclude that the MID data set is not really suitable for a study of escalation, at least outside the realm of the great powers. Holmberg might have considered using the International Crisis Behavior data set.[15] Alternatively, he could have taken the Interstate War data set of the Correlates of War project and done additional data collection to introduce disaggregated codings of levels of violence. In any case, Holmberg 's solution is not suited to task at hand, and he should at least have discussed this problem in the dissertation.

5.3 Methodological Pluralism?

Holmberg argues that quantitative and qualitative studies are complimentary, citing several authorities in favor of his view. However, the thrust of the argument of those who have argued for a pendulum movement between quantitative and qualitative studies is that quantitative (or rigorously comparative) studies should be used to test a theoretical argument. Qualitative studies involving a few cases can be used to explore a richer set of variables, in order to form new hypotheses which are then tried out in a new rigorous test. I do not get the impression that Holmberg disagrees with this. Nevertheless, when he arrives at to his case-studies, he succumbs to the common temptation of referring to them as 'tests'. Even the abstract of the dissertation refers to the case-studies as strengthening the conclusions of the large-*n* study. However, only two cases are examined. If evidence from these two cases were to be adduced in favor of the nomothetic argument advanced in the theoretical part of the dissertation, Holmberg would have to subscribe to a deterministic view of social structure. It seems particularly ill-advised to invoke King, Keohane & Verba in favor of Holmberg's 'methodological pluralism'. Their book is a strong plea for strengthening the scientific method in comparative studies – not for using one or two cases as *evidence* of a general relationship.[16]

Holmberg might have used his case-studies to check whether the causal argument made in the theoretical section seems to fit the detailed examination of the cases. If it does not, new hypotheses might be formulated, which could then be re-tested in a new quantitative study. However, even to be useful in this role, Holmberg would probably have needed much more detailed information. His sources of information are, at best, second-hand. In the case-study of India and Pakistan, for instance, he relies to a large extent on an extensive quotation from Wilcox (1973: 35), published by a conservative think-tank and almost contemporary with the crisis itself (1970–71). It is difficult to see the Wilcox quotation as demonstrating much beyond the fact that Wilcox himself is a realist, which makes it hardly surprising that he should describe the crisis in realist terms.[17]

The study of major Cold War crises like the Korean War and the Cuban Missile Crisis has recently been brought forward by the release of important documents from the most sensitive parts of the decision making process.[18] In the future, such studies are likely to provide a major source of inspiration for theorists of international relations. Even so, the actual testing of generalizing theories will have to be done in a quantitative fashion, or in systematic comparative studies as discussed by King, Keohane & Verba.

To sum up my criticism in a pointed way: Holmberg has not tested the theory he intends to test, he has used data which are basically unsuited for the task, and he combines two methodologies in a problematic way. These are problems he shares with many of his fellow scholars in peace research and international relations. The good news is that he has nevertheless produced a challenging and interesting study, which should inspire further work on the empirical testing of realist ideas.

Nils Petter Gleditsch

Notes

1. This article is built on my opening statement at Björn Holmberg's public defense of his doctoral dissertation (Holmberg, 1998) at Uppsala University, 26 May 1998. I am grateful to the other members of the committee, Peter Wallensteen and Dag Sörbom, for their inputs to my discussion. Stuart Bremer, Lisa Carlson, Mats Hammarström, Susan Høivik, and Björn Holmberg made useful comments on an earlier version. I also acknowledge the financial support of the Norwegian Research Council for my research.

2. Holmberg's emphasis on *rapid* shifts distinguishes his inquiry from the literatures on balance of power, power transition, and long cycles, which focus on long-term shifts and which Holmberg notes only in passing (p. 6).

3. It is not fully clear to me why the period T4 to T5 is not also a military opportunity.

4. However, most quantitative analyses of dyadic behavior (such as Bremer, 1992) include variables measured at the national level, such as democracy. Indeed, so does Holmberg's study.

5. The original source is Gochman & Maoz (1984) and the most updated version, not cited by Holmberg, is described in Jones, Bremer & Singer (1996).

6. This definition requires at least 5 reciprocated MIDs involving the same two states, each dispute lasting at least 30 days. There must also be a minimum of 25 years between the outbreak of the first dispute and the settlement of the most recent one. Finally, if there is a gap of more than 10 years between two disputes, the territorial or other issues at the core of the dispute must remain unsolved during that period, and the gap must not exceed 25 years.

7. Given his low n and a relationship fairly close to significance ($p = .12$), I find his summary dismissal of the bivariate results to be overly pessimistic.

8. Another possible explanation is that when war has broken out, there is only one more step left on the escalation ladder. With a more fine-grained variable for war, he might have found a significant relationship here as well.

9. Some minor reservations: I miss a clarification of how some of the variables are measured and a precise specification of which variables are included in what statistical tests. On the technical side, there are several errors in the references: Bremer, 1992 is cited with an incomplete title; the name of the British publisher Macmillan is misspelled, a particularly serious error for a political scientist with an interest in contemporary history; the first name of the Mr. Hopkins who founded the famous university was Johns and not John, etc. Finally, the fact that the standard errors for all the variables in Tables 12, 13, and 14 are identical except for rounding errors, indicates that there is something wrong with the multivariate statistical analysis.

10. I am grateful to Kristian Gleditsch for calling my attention to these anomalies.

11. http://pss.la.psu.edu/datares.htm. The war data, the disputes data, and other data sets from the Correlates of War project can be downloaded from this site.

12. The coordinator of the Peace Science Society webpage, Stuart Bremer, who was also involved in

generating the new data set, does not consider the revised incidents data to be accurate enough to identify the precise sequence and timing of events. However, they should be accurate enough to assess the highest stage of intensity reached in a dispute. The incidents data have not been released, since it is feared that they might be misused to answer questions for which they are not suitable (pers. comm. 27 April, 15 December 1998).

13. I am grateful to Zeev Maoz for sending me this data set and to Håvard Hegre for analyzing it for me.

14. For instance, Iran and Iraq were involved in 20 MIDs during the period 1953–92. The Iran–Iraq war (1980–88) is probably the only event familiar to most of us.

15. Cf. Brecher & Wilkenfeld (1997) and http://www.colorado.edu/IBS/GAD/spacetime/data/ICB.html.

16. Cf. King, Keohane, & Verba (1994: ch. 6), where they argue against 'crucial case studies', but also suggest ways of increasing the *n* by adding more observations from geographical or other subunits, different time periods, or alternative dependent variables.

17. Holmberg notes this problem (note 120, p. 106) in relation to his more varied sources for the other case-study.

18. A good place to start looking is the *Bulletin* published by the Cold War International History Project at the Woodrow Wilson Center for Scholars, Washington, DC.

References

Allison, G T, 1971. *Essence of Decision: Explaining the Cuban Missile Crisis.* Boston, MA: Little, Brown.

Beck, N; J N Katz & R Tucker, 1998. 'Beyond Ordinary Logit: Taking Time Seriously in Binary Times-Series-Cross-Section Models', *American Journal of Political Science* 42, 1260–1288.

Blalock, H M, Jr, 1979. *Social Statistics.* Second edition. New York: McGraw-Hill.

Brecher, M & J Wilkenfeld, 1997. *A Study of Crisis.* Ann Arbor, MI: University of Michigan Press.

Bremer, S, 1992. 'Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816–1965', *Journal of Conflict Resolution* 36, 309–341.

Carlson, L J, 1995. 'A Theory of Escalation and International Conflict', *Journal of Conflict Resolution* 39, 511–534.

Geller, D, 1993. 'Power Differentials and War in Rival Dyads', *International Studies Quarterly* 37, 173–193.

Gochman, C S. & Z Maoz, 1984. 'Militarized Interstate Disputes, 1816–1976. Procedures, Patterns, and Insights', *Journal of Conflict Resolution* 28, 585–615.

Goertz, G & P F. Diehl, 1993. 'Enduring Rivalries: Theoretical Constructs and Empirical Patterns', *International Studies Quarterly* 37, 147–171.

Holmberg, B, 1998. *Passing the Open Windows. A Quantitative and Qualitative Approach to Immediate Military Balance and Escalation of Protracted Conflicts. Report* 27. Department of Peace and Conflict Research, Uppsala University.

Jones, D M; S A Bremer & J D Singer, 1996. 'Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns', *Conflict Management and Peace Science* 15, 163–213.

Kegley, C, 1993. 'The Neo-Idealist Moment in International Studies? Realist Myths and the New International Realities', *International Studies Quarterly* 37, 131–146.

King, G; R O Keohane & S Verba, 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton, NJ: Princeton University Press.

Lebow, R N, 1984. *Between Peace and War. The Nature of International Crisis.* Baltimore, MD: Johns Hopkins University Press.

Raknerud, A & H Hegre, 1997. 'The Hazard of War: Reassessing the Evidence for the Democratic Peace', *Journal of Peace Research* 34, 385–404.

Singer, J D & M Small, 1994. *The Correlates of War Project: International and Civil War Data, 1816–1992 (ICPSR 9905).* Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Snyder, J L, 1984. 'Perceptions of the Security Dilemma in 1914', pp. 153–179 in R Jervis, R N Lebow & J G Stein, eds, *Psychology and Deterrence.* Baltimore, MD & London: Johns Hopkins University Press.

Wallensteen, P & M Sollenberg, 1998. 'Armed Conflict and Regional Conflict Complexes, 1989–97', *Journal of Peace Research* 35, 621–634.

Waltz, K N, 1979. *Theory of International Politics.* Reading, MA: Addison-Wesley.

Wilcox, W, 1973. *The Emergence of Bangladesh.* Washington, DC: American Enterprise Institute for Public Policy Research.

Quest for Theory Oriented Research and Methodological Pluralism

In his review of my dissertation, *Passing the Open Windows* (1998), Nils Petter Gleditsch undertook an extensive and interesting analysis, and proposed a number of possible enhancements. These ranged from how to interpret theory to which data sets should be used to test the theory. However, and perhaps not surprisingly, I disagree with some of Gleditsch's interpretations of my work, although I do agree with most of his ideas and positions concerning research in general. There are a number of comments that should be made regarding theory, data, and the combination of quantitative and qualitative empirical analysis. These are the main focuses of Gleditsch's criticism and at the same time, hopefully, they should be of general interest to the research community. I leave the minor issues aside.[1]

The theoretical model of the dissertation is illustrated in Figure 1. It is applied to a very specific population of non-great power states, all of which have experienced protracted armed conflict at different levels of intensity. These are relationships where the logic of realism, being a theory of competition and conflict, should be valid. In his criticism, Gleditsch argues that realism does not cover the relationships of non-great powers if they are not, as Waltz (1979) noted, insulated from great-power interference. Given this assumption, I would agree with the critic that my more general purpose, namely, to question some aspects of realism, did not succeed. However, I did not base the deductive logic of the theoretical model on Waltz's neorealism. Nor was my specific purpose to combat realism in general but rather, to test an important realist proposition.

As discussed in Chapter 2, there are several versions of realism, and Doyle (1997) identified four. They all differ substantially, although they share certain assumptions. In the concluding section of Chapter 2, I emphasize that it is important to find the deductive base for the relationship between military opportunity and escalation, the most important relationship in the study. Focusing on earlier modern realists, called *realists*, such as Carr (1946), Kaplan (1957), and Morgenthau (1979) on the one hand and *neorealists*, such as Waltz on the other hand, leads to the following conclusions:

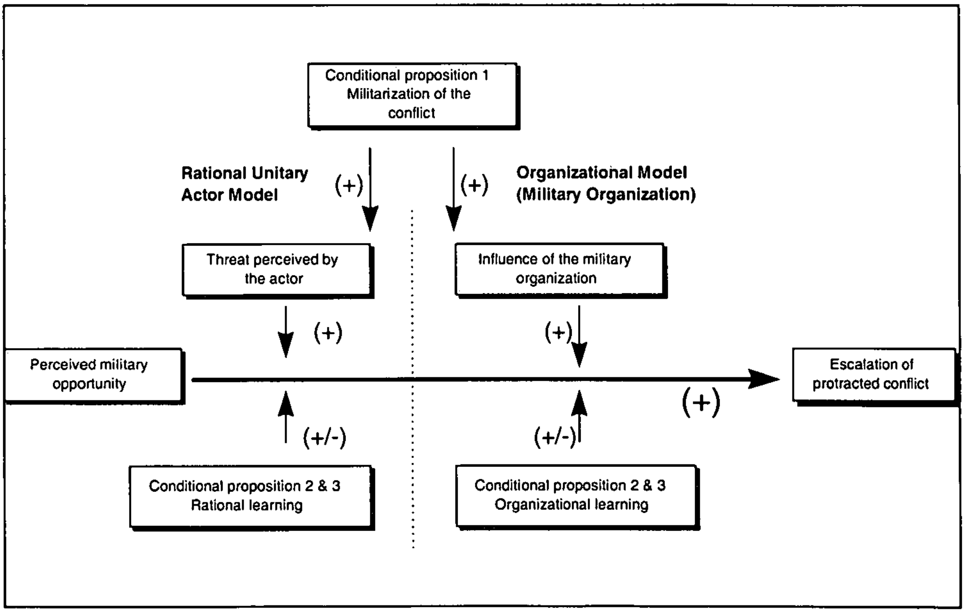

Figure 1. Theoretical Model of Military Opportunity and Escalation.

First, realists would say that a rapid shift in the military balance is an incentive for escalation (that is, it increases the probability of an escalation). Remember Kaplan's (1957:23) phrase 'Fight rather than pass up an opportunity to increase capabilities.' Second, neorealists would say that dramatic shifts in military balance would lead to action if security were threatened. Otherwise, this depends on whether power is a goal in itself as it is for states that seek 'universal domination' (Waltz, 1979:118). (Holmberg, 1998: 19)

It is thus clear that realism, in contrast to Waltz's neorealism, gives a rather solid base for the view that military opportunity is strongly related to escalation. Consequently, this version of realism was the one put to test in the dissertation. Throughout the dissertation realism is separated from neorealism, and the latter paradigm even lends its view to the inductive conclusions from the case studies, still to be empirically tested. That is, revolutionary processes in a state (State A) might increase the level of threat perceived by its opponent (State B) and thus increase the probability of escalation by State B in case of a military opportunity. As an example from the dissertation, the revolution of 1979 crippled Iran militarily, thus creating a military opportunity for Iraq, but also producing a threat towards the regime of Iraq in that the revolution could spread to Iraq. Consequently, Iraq attacked Iran in 1980 partly as a result of the 1979 revolution. This is in line with the findings of Walt (1992) and Maoz (1989) that revolutions are associated with international conflict. The inference in this dissertation is that there might be an important interaction effect between threat perceptions and military opportunity, both caused by the revolutions, as shown by the example above.

Gleditsch's initial question concerning the validity of applying realism to non-great powers is interesting. Several scholars state that realism, in general, is a great-power theory or paradigm (for example Levy, 1989; Reiter, 1996). Gleditsch's criticism seems well founded. However, as shown above, Waltz's statement on insulated non-great powers is not the deductive foundation of the dissertation. Furthermore, by saying, 'the desire to attain a maximum of power is universal', Morgenthau (1973:208) lets us assume realist behavior from non-great powers. This line was cited in the dissertation and is the strongest theoretical reason for testing a *realist* proposition on non-great powers that are a part of the *universe*. Nevertheless, while I agree that this application is non-conventional it is still important to extend theory testing beyond its classical domains, as in the case of realism. If not, there would be little progress in the field of conflict research.

The second major area of criticism, data, is a very important one. I agree with most of Gleditsch's objections to using the *Militarized Interstate Disputes* data (MID) to measure the general concept of escalation. Still, this is basically a question of how the central variables are defined, and in what way they are interrelated in theory. First, I was not trying to explain escalation of protracted conflicts in general but to investigate the specific relationship between military opportunity and escalation—in other words, if military opportunity was a *sufficient* condition for escalation. Second, not all types of escalation were of interest. The bivariate relationship between military opportunity and escalation had its deductive ground in realism. That is, if there is a dramatic shift in power the stronger will attack the weaker to convert the imbalance into larger capabilities. As cited before 'Fight rather than pass up an opportunity to increase capabilities.' (Kaplan 1957:23). The brutish manner of a realist would make us expect a larger attack by the stronger state on the weaker state. In addition, in the enhanced theoretical model based on rational choice and organization theory I expected the rational actor under threat to be more willing to escalate (shown as an interaction effect in figure 1). This was debated in the dissertation leading to a theory-based operational definition of escalation,

Another theoretical reason not to include disputes below this threshold [*use of force*] is that we expect a clear and decisive move from the actor as he is interested in removing the threats that his adversary is projecting. A minor border clash as a result of a military opportunity would make little sense in our model. An empirical benefit of this approach is that underreported minor observations, which are also likely to suffer from systematic measurement errors, are excluded. (Holmberg, 1998:59)

The conclusion must thus be that crisis escalation and minor escalations are of little theoretical interest in the dissertation under discussion. A consequence of this, as stated in the quote, is that the most serious validity problem with MID is avoided. Most of the escalations were verified through the reliable Correlates of War data set (COW) or through manual coding[2]. However, when considered as an interaction effect (see degree of militarization in 1) lower levels of violence were used. This is problematic, as discussed by Gleditsch.

This leads me to a general observation. Even if Gleditsch's criticism of the data is very important to the field of quantitative analysis it is not really relevant here. It is more a question of theory than of data. There seems to be a widespread tendency in quantitative analysis to emphasize technical operational variables. This without clear reference to their underlying theoretical variables and the theoretical context in which they are placed. In a study that was basically deductive, as in the case of the dissertation under debate, the link between theory, method, and data is essential, as the theory would otherwise not be tested.

Finally, Gleditsch asserts that I combine quantitative and qualitative methods in a problematic way. The question of which method to use is a long-standing battle between heated contenders. Most Ph.D. candidates are taught by their mentors that the choice of method should be directed by the research problem. However, this seldom seems to be the case in the real world dissertation. I have tried to combine methodologies with the best intentions, although it has inevitably meant that each part, the quantitative and the qualitative, has become less extensive. But, with the belief that the combination amounts to more than the sum of its parts, it was an easily justifiable choice.

Gleditsch's criticism is mainly directed towards my use of the term 'tests' when referring to the case studies and the fact that I use two cases as evidence of a general relationship. The first problem is rather a question of semantics—perhaps I should have used another term. But the second issue is more serious. Let me first cite the dissertation before debating the issue

As shown in the discussion, the qualitative phase could validate the large-*n* results; that is, some of the supportive observations were put to the test and sustained. At the same time, some observations could be found and others could be dismissed. This might lead the reader to the conclusion that the case study method is superior; however, I will argue that the combination of the two methodologies, quantitative and qualitative, has shown its advantages. Looking at only the case studies might have given the impression that military opportunity generally causes escalation or de-escalation. The unstructured selection of cases, which is not too uncommon, may lead to selection on the independent and dependent variables and thereby generate a biased result. By making a general application to a large sample, selected through theoretically motivated criteria, we could conclude that military opportunity is not associated with escalation. Further, the weight of threat perception and the initiation of the Iran-Iraq war would not have been discovered without the case studies. Methodological pluralism appears to be the winner. (Holmberg, 1998:120f)

In these lines in the final chapter, I clearly state that a general conclusion based on a few cases would have led to the wrong conclusions. The magic of the dual method is that the results are *combined*. That is, trends found in the quantitative phase could be further investigated in the case studies through a very important theoretical link. Thus, two of the most important cases that supported the theoretical model were subject to a second 'test'. In so doing, I was being particularly severe on my own theory. We could call it an advanced recording of cases essential for the quantitative conclusions.[3] In addition to this, the two 'cases' of India–Pakistan 1970–1971 and Iran–Iraq 1979–1988 generated six observations (cases) where at least one of the theoretical variables was present (see page 110). In other words, there were not two cases but six as I used the method, cited by Gleditsch, to increase the number of observations by using different time periods. At least this shows that if I had committed, in my own words, the crime of drawing general conclusions based on the case studies, there would at least have been six, not two as Gleditsch implies. I also investigated whether the causal links suggested by the theory were present, and drew some inductive conclusions that refined the theoretical model. (The former being an approach suggested by Gleditsch). However, debating the inductive conclusions I stated,

Ideally, the results of the reconstruction should be tested in a new large-*n* study. This is necessary if the scholar wishes to bring the theory beyond the stage of loose *ad hoc* hypotheses. (Holmberg, 1998:11)

Consequently, I agree with Gleditsch that case studies, standing alone, can give us theoretical generalizations but in order to be empirically generalizable theory must be tested on a properly selected sample.

In finalizing this response to Nils Petter Gleditsch's interesting criticisms of my dissertation, I must say that in general we seem to agree with each other in many areas. However, in contrast to Gleditsch's ending words, I must emphasize that I did test the theory I intended to test (as shown in figure 1). I also tested a *realist* proposition on its on turf while extending its classical domain of empirical application. The data that were used were *theoretically* most suitable for the task, and the research effort combined two methodological fields that are seldom merged. This resulted in more refined and interesting conclusions, and also led to an improved theoretical model.

Björn Holmberg

Notes

1. I want to express my gratitude to Assistant Professor Erik Melander in the Department of Peace and Conflict, Uppsala University, for comments on this text. I need to comment on the standard errors in the logistic regression. A small error in my use of SPSS 6.1 resulted in identical standard errors for all variables. Redoing the calculation gave me correct standard errors and the conclusions in the dissertation were not affected. An article based on the dissertation is forthcoming. See www.peace.uu.se/holmberg.htm for future information.

2. The manual coding might have been more clearly described in the dissertation.

3. Ideally, this should be done with all cases, as observations that did not support the model might be found to do so if they were recoded. However, by only looking at the supportive observations at least I did not inflate the results in my own favor.

References

Carr, Edward Hallet, 1946. *The Twenty Year' Crisis 1919-1939: An Introduction to the Study of International Relations*. London: Macmillan.

Doyle, Michael W., 1997. *Ways of War and Peace: Realism, Liberalism, and Socialism*. New York: W.W. Norton & Company.

Holmberg, Björn, 1998. *Passing the Open Windows: A Quantitative and Qualitative Approach to Immediate Military Balance and Escalation of Protracted Conflicts*. Uppsala: Department of Peace and Conflict Research.

Kaplan, Morton A., 1957. *System and Process in International Politics*. New York: John Wiley & Sons.

Levy, Jack S., 1989. "The Causes of War: A Review of Theories and Evidence", pp. 209–333 in Philip E. Tetlock; Jo L. Husbands; Robert Jervis; Paul C. Stern & Charles Tilly, ed. *Behaviour, Society, and Nuclear War*. New York: Oxford University Press.

Maoz, Zeev, 1989. "Joining the Club of Nations: Political Development and International Conflict", *International Studies Quarterly*, vol. 33, no. 2, pp. 199–231.

Morgenthau, Hans J., 1973. *Politics Among Nations: The Struggle for Power and Peace*. New York: Alfred A Knopf.

Reiter, Dan, 1996. *Crucible of Beliefs: Learning, Alliances, and World Wars*. Ithaca and London: Cornell University Press.

Walt, Stephen M., 1992. "Revolution and War", *World Politics*, vol. 44, no. 3, pp. 321-368.

Waltz, Kenneth N., 1979. *Theory of International Politics*. Reading: Addison-Wesley.

Rejoinder

1. I recognize that Holmberg acknowledges that there are different school of realist thought. But do any of them advocate studying realist behavior on a sample where great powers are deliber-

ately excluded? Since he quoted Waltz twice, I assumed that Holmberg based his unusual choice on this quotation. Apparently, we now agree that this statement by Waltz cannot justify his selection of units.

2. I do not understand on what theoretical basis Holmberg derives the idea that he must limit himself to escalation to war and ignore all other forms of escalation. I would have thought that a rational decisionmaker behaving according to realist principles would try to avoid war if he could coerce his opponent by less violent means. Holmberg defends his choice by referring to 'the brutish manner of a realist'; but this seems like an ad hoc explanation to me.

3. I concede that our disagreement on the third point, concerning the means of testing the theory, may be more semantic than I thought when I read the dissertation. I still think that Holmberg displayed lack of caution in his use of the word 'test' in connection with the case studies. And if we agree that statistical hypothesis-testing is the way to go, an *n* of 6 is not terribly satisfactory.

*

Since writing my original article, I have discovered two additional articles which study escalation using the MID dataset. I have been unable to incorporate them in my main argument, but I take this opportunity to call the readers' attention to them:

Partell, P, 1997. 'Escalation at the Outset: An Analysis of the Targets' Responses in Militarized Interstate Disputes', *International Interactions* 23: 1-36.

Peter, P & G Palmer, 1999. 'Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes', *International Studies Quarterly* 43: 389-405.

Of these two, the first makes use of the Incidents file and is, as far as I know, the only other published work to do so, besides Carlson, 1995. I am grateful to Glenn Palmer for these additional references.

Nils Petter Gleditsch

Till redaktionen insända skrifter

Karl-Göran Algotsson, 2000. *Sveriges författning efter EU-anslutningen*. Stockholm: SNS.

Ulf Bernitz, Sverker Gustavsson, Lars Oxelheim, red, 2000. *Europaperspektiv. Årsbok 2000*. Stockholm: Santérus Förlag.

Erik Beukel, Kurt Klaudi Klausen, Poul Erik Mouritzen, ed, 2000. *Elites, Parties and Democracy. Festschrift for Professor Mogens N. Pedersen*. Odense: Odense University Press.

Torben Bech Dyrberg, Allan Dreyer Hansen, Jacob Torfing, red, 2000. *Diskursteorien på arbejde*. Frederiksberg: Roskilde Universitetsforlag.

Tor Egil Førland, Dag Harald Claes, 1999. *Europeisk integration*. Lund: Studentlitteratur.

Carl Johan Gardell, 2000. Islamisk fundamentalism. *Världspolitikens Dagsfrågor* 1 2000.

Ann-Sofi Jakobsson Hatay, 1999. Fredsprocessen på Nordirland. *Världspolitikens Dasgfrågor* 12 1999.

Mats Karlsson, 1999. *Vår man i Moskva. En studie över den svenske ambassadören Rolf Sohlmans syn på Sovjetunionen och dess utrikespolitiska intentioner 1947-1950*. Stockholm: Historiska institutionen.

Magnus Killander. 1999. Vägval för Iran. *Världspolitikens Dagsfrågor* 11 1999.

Elisabeth Kronqvist, 1999. *Institutionalisering i nordiska partier*. Åbo: Åbo Akademi.

Riikka Kuusisto, 1999. *Western Definitions of War in the Gulf and in Bosnia*. Helsingfors: Finska Vetenskaps-societeten.

Assar Lindbeck m fl, 2000. *Politisk makt med oklart ansvar. Ekonomirådets rapport 2000*. Stockholm: SNS.

Alf Ross, 1999. *Ret som teknik kunst og videnskab*. Udvagt af Isi Foighel, Hans Gammeltoft-Hansen, Henrik Zahle. Köpenhamn: Jurist- og Økonomforbundets Forlag.

SNS 1999. *Sveriges konstitutionella urkunder*. Stockholm: SNS.

SOU 1999:112. Civilsamhället som demokratins arena. Demokratiutredningen. Skrift nr 29.

SOU 1999: 121. Avkorporativisering och lobbyism. Demokratiutredningen. Forskarvolym XIII.

SOU 1999: 126. Politikens medialisering. Demokratiutredningen. Forskarvolym III.

SOU 1999:129. Demokratins estetik. Demokratiutredningen. Forskarvolym IV.

SOU 1999: 130. Demokratins trotjänare. Demokratiutredningen. Forskarvolym X.

SOU 1999: 131. Marknaden som politisk aktör. Demokratiutredningen. Forskarvolym XI.

SOU 1999:144. Demokrati på remiss. Demokratiutredningen. Skrift nr 30.

SOU 1999:150. Vad hände med Sveriges ekonomi efter 1970? Demokratiutredningen. Skrift nr 31.

Medarbetare i Statsvetenskaplig Tidskrift

Dag Anckar är professor i statskunskap vid Åbo Akademi.

Fil dr Ingegerd Municio-Larsson är verksam vid Södertörns Högskola.

Lena Dahlberg är doktorand i statsvetenskap i Stockholm och verksam vid Dalarnas forskningsråd.

Fil dr Mikael Axberg är verksam vid Statsvetenskapliga institutionen i Uppsala.

Rolf Hugoson är verksam vid Statsvetenskapliga institutionen i Umeå.

Fil dr Bengt-Ove Boström och Marie Uhrwing är verksamma vid Statsvetenskapliga institutionen i Göteborg.

Bo Rothstein är professor i statsvetenskap i Göteborg.

Professor Olof Petersson är forskningsledare vid SNS i Stockholm.

Fil dr Thomas Denk är verksam vid universitetet i Karlstad.

Professor Nils Petter Gleditsch är verksam vid International Peace Research Institute i Oslo.

Fil dr Björn Holmberg är verksam vid avdelningen för freds- och konfliktforskning i Uppsala.