

Reflektivt ekvilibrium: tre utvidgningar

1. INTRODUKTION

De flesta moralfilosofer säger inte särskilt mycket om vilken metod de använder. Ofta befinner de sig inom en viss tradition eller kontext där merparten gör saker på ett likartat sätt och grunderna för arbets sättet ifråga tenderar då att lämnas outtalade. När moralfilosofer inom den analytisk-filosofiska traditionen numera ändå säger något av metodologisk art handlar det ofta om att de strävar efter *reflektivt ekvilibrium*. Denna term introducerades av John Rawls i *En teori om rättvisa* och tankegångarna ifråga har sedan utvecklats dels av Rawls, dels av andra.¹ Efter att först ha karakteriserat grunddragen hos tanken om reflektivt ekvilibrium kommer jag här framför allt att gå in på tre olika sätt att utvidga den relevanta processen och diskutera i vilken utsträckning dessa tre utvidgningar är rimliga, dels var för sig, dels tagna ihop.

Teorin om reflektivt ekvilibrium ses ofta som en koherentistisk teori om epistemiskt rättfärdigande. Här kommer jag dock i första hand att diskutera reflektivt ekvilibrium som ett arbets sätt inom moralfilosofi och politisk filosofi, framför allt då i olika försök att konstruera normativa teorier inom dessa discipliner (och i det som följer kommer jag att fokusera på normativ etik);² den exakta statusen hos den produkt som eventuellt kommer ut ur detta arbete kommer jag däremot inte att diskutera i någon större utsträckning.³ Jag kommer att peka på en del pragmatiska skäl för att betrakta tillstånd av reflektivt ekvilibrium som tankemässiga framsteg, men detta skall inte tas som uttryck för en underliggande pragmatisk teori om epistemiskt rättfärdigande.

¹Framför allt av Norman Daniels i en serie uppsatser som finns samlade i hans (1996).

²En sådan förståelse av reflektivt ekvilibrium har bland annat förts fram av Michael DePaul (1986) som menar att det handlar om en metod för teorikonstruktion och att den därmed är förenlig med fundamentism ifråga om epistemiskt rättfärdigande.

³För en ingående diskussion av dessa frågor, se Folke Tersman (1993).

2. PRINCIPER OCH GENOMTÄNKTA OMDÖMEN

Även om det finns de som menar att principer inte har någon intressant roll att spela inom moraliskt tänkande⁴ är den vanligaste ansatsen inom normativ etik ändå att försöka identifiera några få grundläggande principer, ibland rentav en enda högsta princip, alltså ett försök att systematisera vårt moraliska tänkande och att få fram något slags moralens kärna.

I grundmodellen för reflektivt ekvilibrium finns två kategorier som vi utgår från: (A) genomtänkta moraliska omdömen (för vilka vi har starkt förtroende, besitter relevant information om och vilka dessutom inte bäst förklaras av fördomar eller egenintresse), samt (B) moraliska principer (vilka i utgångsläget vanligen är generaliseringar av typen ”det är fel att ljuga”). Även om vi är övertygade om att våra principer har undantag (ibland är det till exempel rätt att ljuga) kan vi fortfarande ställa upp på dem just som generaliseringar: de gäller för det mesta. Kategori (A) kan dock också innehålla exempel för vilka vi i utgångsläget inte har någon relevant princip.

Det exakta innehållet i de åsiktsmängder som utgörs av (A) och (B) kan naturligtvis variera från person till person, men tenderar att vara relativt spretigt hos de allra flesta av oss. Tanken med reflektivt ekvilibrium är dock att vi kan arbeta oss fram mot en mer välordnad åsiktsmängd genom att gradvis jämka samman principer och enskilda omdömen till en mer genomtänkt och sammanhängande helhet. Därmed öppnas det också upp en möjlighet för att vissa moraliska åsiktsmängder har ett försteg framför andra. Rent pragmatiskt är det i alla fall svårt att tänka sig någon inleda en debatt med att säga ”Mina åsikter är visserligen inte genomtänkta och inte hänger de samman heller, men lyssna nu på vad jag har att säga i denna fråga!”.

Vad är det då som skall hända för att vi skall kunna sägas ha rört oss mot ett tillstånd av reflektivt ekvilibrium? Motsägelsefrihet verkar vara ett rimligt utgångskrav,⁵ men ett reflektivt ekvilibrium bör vara något mer än bara det: en samling åsikter som överhuvudtaget inte har med varandra att göra kan vara motsägelsefri, men skulle knappast kallas sammanhängande för det. Ett krav man skulle kunna ställa är härledbarhet *från principer till omdömen*, att varje omfattat omdöme om någon specifik situation logiskt följer från någon princip plus en beskrivning av den relevanta situationen (ett slags syllogism med en moralisk princip som översats, en icke-normativ situationsbeskrivning som undersats och det enskilda normativa omdömet som slutsats). Detta skulle dock förutsätta att den moraliska

⁴Den ledande företrädaren för denna hållning är Jonathan Dancy (2004).

⁵Om man släpper detta krav är det nog rimligt att säga att resultatet blir en annan metod, se till exempel Ben Eggleston (2010) för en diskussion av vad han kallar *praktiskt ekvilibrium*, vilket är ett tillstånd som kan tillåta vissa motsägelser.

omdömesförmågans roll i att komma fram till enskilda omdömen skulle kunna elimineras och den enda realistiska möjligheten för det är förmodligen en moralisk monism, där moralen reduceras ner till en enda princip. Så snart vi har en mångfald av principer kommer det också att uppstå sammanvägnings- och prioriteringsfrågor där vi förmodligen inte kommer att kunna identifiera några enkla regler för hur de skall lösas; det blir upp till vår omdömesförmåga. Redan att identifiera en samlad uppsättning principer tycks dock kunna utgöra ett steg framåt och dessutom ge oss viss vägledning i hur vi skall använda vårt omdöme, så även om härledbarhet kan vara eftersträvsvärt är det nog inte rimligt som ett krav. Den främsta monistiska kandidaten på marknaden, utilitarismen, har ju för övrigt också haft stora svårigheter med att hantera olika enskilda omdömen för vilka vi har mycket stark tillit.⁶

Ett annat möjligt krav är vad som kan kallas *täckning av omdömen*, att varje omdöme skall gå att begripliggöras i termer av någon eller flera principer. Det skall inte finnas några ”fritt svävande” omdömen i en sammanhängande moralisk åsiktsmängd utan varje omdöme faller under en eller flera principer, även om de inte logiskt följer ur dessa givet någon viss situationsbeskrivning.⁷ Detta krav är förenligt med moralisk pluralism och det tycks onekligen som att det skulle vara något av ett misslyckande om det efter genomgången reflektionsprocess i ens åsiktsmängd fanns omdömen för vilka man har tillit, men som inte har någon täckning i de principer man omfattar.

Om man strävar mot att formulera en uppsättning principer torde också ett par desiderata, om än inte krav, vara dels att de principer man omfattar som grundläggande är relativt få, dels att de befinner sig på en likartad generalitetsnivå. Det skulle vara något otillfredsställande med ett system där man både har principer som ”Det är fel att döda” och principer som ”Det är fel att gå mot röd gubbe när småbarn är i närheten och kan se en göra det” som grundläggande principer; båda dessa principer kan ju visserligen vara tillämpliga på olika enskilda fall, men den senare är så specifik att den ligger väldigt nära att vara ett enskilt omdöme och man kan tycka att efter en genomgången reflektionsprocess borde man ha en mer generell princip

⁶Exakt vilka problem som uppstår beror på vilken variant av utilitarismen som är aktuell; den klassiska hedonistiska utilitarismen kan till exempel inte göra reda för att de flesta av oss anser att en person agerar fel om han torterar ett litet barn för sitt eget nöjes skull även om det skulle vara så att han fick ut mer lust ur det än vad barnet led.

⁷Eventuellt skulle man kunna karaktärisera den relevanta relationen som abduktiv snarare än deduktiv, men den förra termen är inte lika väldefinierad som den senare så att bara tala om täckning lämnar öppet exakt hur vi skall förstå relationen.

som täcker in de fall som täcks in av den, i ett hierarkiskt system borde den vara en princip på lägre nivå, inte på den högsta.

Så långt vad som är involverat i en strävan mot reflektivt ekvilibrium i dess enkla eller snäva form; som redan nämnts har dock många velat utvidga processen på olika sätt och även om skälen till det kan vara flera kommer vi här i första hand att bry oss om ett par besläktade problem som båda har att göra med att processen alltid utgår från specifika personers åsiktsmängder och att var vi hamnar kommer att bero på var vi startar. Till att börja med kan man tänka sig att en person med vad vi uppfattar som befängda moraliska åsikter, säg en fanatisk nazist, genomgår processen och bara kommer ut på andra sidan som en mer sofistikerad moralisk idiot. Vad är då värdet på den kvalitetsstämpel som reflektivt ekvilibrium tänks utgöra? Det andra problemet är att även inom klassen av människor med vad vi uppfattar som någorlunda vettiga moraliska åsikter kan man tänka sig att de genomgår processen och snarare än konvergerar i stället divergerar. Inom både moralen och politiken är det ju viktigt att kunna nå fram till gemensamma hållningar och om det inte finns något som på förhand talar för att processen skall leda någonvart på denna punkt, vad är då poängen? Om reflektivt ekvilibrium bara är ett tillstånd i en privat bubbla är det kanske i slutändan ändå inte så intressant. De tre idéer som i litteraturen förts fram om att utvidga processen och som vi nu skall gå vidare till att undersöka kan alla sägas karaktäriseras av att de hakar i något som är externt till individens moraliska åsiktsmängd och därmed eventuellt kan fungera som ett slags tyglar som styr upp hur vi kan röra oss inom ramen för vår strävan mot reflektivt ekvilibrium.

3. UTVIDGNING AV INNEHÅLLET: BAKGRUNDSTEORIER

Den första utvidgningen som vi skall titta på har framför allt förts fram av Norman Daniels⁸ och den grundläggande tanken är att man inte bara skall se till normativa principer och omdömen utan att man också skall utsträcka den aktuella åsiktsmängden till att inkludera (C) relevanta bakgrundsteorier, till exempel teorier om vad det innebär att vara en person och hur samhällen fungerar. Exakt vad som skall ingå i (C) kan naturligtvis diskuteras, och varken Daniels eller Rawls har något tolkningsmonopol här, men på ett allmänt plan tycks det i alla fall rimligt att acceptera att ens moraliska övertygelser är starkt knutna till olika föreställningar om aktörskap både i individuell och social bemärkelse och att anta att våra moraliska åsikter redan i utgångsläget inte existerar

⁸Se framför allt (1996, kapitel 2 [ursprungligen publicerad 1979]).

som något separat, utan är integrerade med en förståelse av sig själv som aktör och samhällsmedlem. Om man till exempel har en atomistisk syn på människan kommer man förmodligen att dras till liberala principer, medan om man ser människan som en i djup mening social varelse kommer man förmodligen att finna andra principer rimliga i ljuset av detta. Givet att denna koppling redan existerar, är det då inte också rimligt att ta in dessa bakgrundsteorier i processen?

Ett skäl att svara nekande på den frågan kan vara en hotande komplexitet. Måste man för att bedriva normativ teori nu också kunna hantera statsvetenskap, sociologi, psykologi, ekonomi och kanske ett par discipliner till? Inte nödvändigtvis. Visserligen skapar denna utvidgning eventuellt kontaktytor mot dessa discipliner men för det filosofiska tänkandet handlar det förmodligen mer om att abstrakta modeller av personers och samhällens natur kan fungera som delar i en reflektivt ekvilibrium-process; det handlar alltså i första hand om till exempel filosofiska teorier om personskap. Dessa bör naturligtvis inte vara ur fas med vad empiriska vetenskaper har att säga, men huruvida så är fallet är knappast något som måste avgöras i samband med varje försök till normativt teoretiserande och den enskilde tänkaren behöver inte delta i djupgående diskussioner om även dessa bakgrundsteorier för att kunna knyta an till dem i sitt normativa teoretiserande. Komplexiteten stiger, men det finns skäl att anta att den redan är högre än vad en snävare ansats som bara fokuserar på principer och omdömen kan ge sken av.

Ett annat problem är att i den utsträckning som dessa vetenskapliga eller filosofiska teorier är deskriptiva, på vilket sätt kan de då sägas vara relevanta för (A) och (B) som båda är normativa domäner? Det är en sak att vi till vardags ofta baserar våra moraliska åsikter i dem, men det kanske helt enkelt bara är för att vi på ett felaktigt sätt tenderar att gå från *är* till *bör*.⁹ Det enda vi kan få i relation till (C) är kanske då en trivial motsägelselfrihet eftersom moraliska och icke-moraliska påståenden är av olika art och inte kan motsäga varandra (mellan "Du ljög när du sade det" och "Du borde inte ha ljugit när du sade det" finns till exempel ingen logisk motsättning). Här är det dock viktigt att komma ihåg att även mellan (A) och (B) var inte kravet att vi skulle kunna deducera omdömen ur principer; idealet om täckning innebär bara en svagare relation som att kunna begripliggöras av eller att vara rimlig i ljuset av någonting¹⁰ och sådana relationer kan mycket väl tänkas föreligga mellan deskriptivt och normativt, framför allt om man antar olika sidoprinciper som "böra im-

⁹Margaret Holmgren (1989, s. 51–52) kritiserar brett reflektivt ekvilibrium bland annat utifrån detta problem.

¹⁰Se Catherine Elgin (1996, s. 128).

plicerar kunna”¹¹ eller metodologiska principer som ett värdesättande av enkelhet och tankeekonomi. Det kan till exempel föreligga olika former av isomorfism mellan normativa och deskriptiva föreställningar, och givet detta kan vissa föreställningar passa bättre ihop med andra, även om det inte föreligger några strikta logiska relationer mellan dem. Bara för att man har en starkt individualistisk uppfattning om hur människor faktiskt fungerar måste man inte vara politiskt liberal, men den isomorfism som föreligger mellan dessa uppfattningar gör att det tankemässiga steget från det första till det andra ändå är mindre än till att vara politisk kollektivist. Bara för att man anser att relationen till ens framtida jag är en relativt svag relation på det metafysiska planet, behöver man inte landa i uppfattningen att våra skäl för att bry oss om oss själva inte står i någon särklass jämfört med våra skäl att bry oss om andra, men något slags passform finns ändå mellan dessa båda uppfattningar.¹²

Att ta in bakgrundsteorier öppnar upp för att metoden inte blir fullt så koherentistisk som Rawls och Daniels tänker sig den, tvärtom skulle olika metafysiska teorier, kanske framför allt om vad det innebär att vara en person, kunna bli ett slags fast mark som vi kan bygga våra moraliska och politiska teorier på. Anledningen till att de båda inte tror att det finns någon sådan fast mark är inte heller så mycket epistemologisk eller metodologisk som metafysisk. De tänker sig att personer är i grunden plastiska¹³ och att vad det innebär att vara en person kommer därmed att bero, åtminstone till en betydande del, på vilket slags samhälle man lever i. Hur vi fungerar som personer kommer alltså att till stor del bygga på vilka moraliska och politiska föreställningar som dominerar i det samhälle där vi formats som personer. Den eventuella arkimediska punkten som bakgrundsteorierna erbjuder är, om detta stämmer, kanske fast nog för att ta spjårn mot ifråga om vissa moraliska föreställningar (om man till exempel resonerar inom ramen för ett samhälle redan karaktäriserat av en liberal individualism, då har en stark kollektivism eller tankar om naturliga hierarkier ingen god passform med den typ av människor som

¹¹Principen som säger att för att det skall kunna vara så att vi bör göra något måste vi också kunna göra det omfattas av många moralfilosofer. I sin enkla form kan denna princip rentav tillåta oss att deducera vissa moraliska slutsatser från vissa beskrivningar; men det är också fullt möjligt att svagare versioner av principen, som att om något är väldigt svårgenomförbart blir det klart mindre rimligt att ställa som moraliskt krav på oss, kan vara acceptabla och då blir det mer en fråga om avvägningar. Sådana svagare versioner skulle kunna vara ett exempel på när bakgrundsteorier kan bli relevanta för att bedöma vissa normativa uppfattningar.

¹²En av de centrala argumentationslinjerna hos Derek Parfit (1984) bygger på just detta.

¹³Se framför allt Daniels (1996, kapitel 7).

befolkar sådana individualistiska samhällen), men det är knappast fråga om någon fast berggrund på vilka stora teorier kan byggas.

Hur kan då denna typ av utvidgning lösa eller lindra de problem som togs upp med den snäva ansatsen? Till att börja med bygger en del hållningar som många av oss finner moraliskt stötande på deskriptiva föreställningar som kan underkännas, till exempel förutsätter en nazists moraliska hållning vissa idéer om raser som inte kan sägas ha någon rimlig metafysisk eller vetenskaplig bas och då är inte heller dessa moraliska åsikter rimliga hur sammanhängande och systematiserade de än må vara. Också givet personers plasticitet kommer en del föreställningar om hur människor fungerar fortfarande att vara falska. När det gäller problemet med divergens verkar det troligt att så länge vi utgår från samma typ av samhälle kommer bakgrundsteorierna att kunna fungera som ett slags ankare som begränsar hur långt ifrån varandra vi kan röra oss i vår strävan efter reflektivt ekvilibrium. Även om det kan finnas oenighet också ifråga om bakgrundsteorierna är utrymmet för kreativitet inte lika stort där som ifråga om vilka moraliska uppfattningar man som enskild individ kan tänkas ha, särskilt som empiriska vetenskaper nog ändå måste antas ha visst inflytande över bakgrundsteorierna även när dessa är av ett mer abstrakt slag.

4. UTVIDGNING AV PROCESSEN: ALTERNATIVA UPPFATTNINGAR

Även om Daniels kanske är den vars diskussion om breddat reflektivt ekvilibrium varit mest inflytelserik var det Rawls själv som först introducerade en distinktion mellan snävt och brett (1974–75, s. 8). För Rawls handlade det dock då inte i första hand om en skillnad i vad som ingår i ett tillstånd av reflektivt ekvilibrium utan om hur vägen fram mot detta tillstånd sett ut. Å ena sidan kan man tänka sig en process där man bara utgår från sin egen åsiktsmängd och sedan försöker hitta kortaste vägen till ett tillstånd av koherens, å andra sidan en process där man öppnar upp för olika alternativa uppfattningar och olika argument för dessa, för att sedan därutifrån arbeta med sin egen åsiktsmängd. Även om man inte direkt plockar in delar av dessa alternativa uppfattningar i sin egen åsiktsmängd, kommer den tillit som man har till olika principer och omdömen i denna med stor sannolikhet att förändras. Det finns naturligtvis inga garantier för att denna typ av process kommer att innebära mer än att man tar en längre väg till samma reflektiva ekvilibrium som man ändå hade kommit fram till, men det är inte heller det viktiga. Poängen är att den längre vägen ger en ytterligare tyngd åt den åsiktsmängd som man till slut når fram till. För att återigen göra ett pragmatiskt test, om någon skulle inleda en debatt med att säga ”Visserligen har jag överhuvudtaget inte tagit in och reflekterat över andras ståndpunkter när jag

format mina egna, men lyssna nu noga på vad jag har att säga!” skulle man förmodligen inte lyssna särskilt noga.

Medan den första utvidgningen låg mer tydligt inom ramen för något slags koherentism handlar denna alltså snarare om att utvidga vad som är involverat i vår strävan mot reflektivt ekvilibrium så att den blir till något mer än just bara strävan mot koherens; som Rawls själv uttrycker saken handlar det, draget till sin spets, om att sträva efter den moraluppfattning som ”skulle överleva ett rationellt övervägande av alla tänkbara moraluppfattningar och alla rimliga argument för dessa” (1974–75, s. 8, min övers.). Det blir alltså en fråga om vad man kan kalla för ett *reflektivt överlevnadstest*.

Rawls noterar dock också att en sådan process knappast är till fullo genomförbar och föreslår att vi utgår från den filosofiska traditionen och de uppfattningar som tenderat att omfattas inom den. Sidgwick's jämförande studie av moraluppfattningar i *The Methods of Ethics* tas som ett exempel på hur detta kan gå till.¹⁴ Men varför då fokusera på uppfattningar ur den filosofiska traditionen? Rawls ger inget utförligt svar på den frågan, men rimligtvis är tanken att även om man naturligtvis hittar moraliska och politiska uppfattningar också utanför filosofin så är det framför allt där man kan finna dels klart formulerade versioner av sådana uppfattningar, dels en mer utarbetad argumentation kring dessa. Att ta in dessa alternativ i den egna reflektionsprocessen skall handla om att utmana sina egna uppfattningar och då måste man ju ställa dem mot de skarpaste versionerna av alternativen.

Denna form av utvidgning är inte alls lika diskuterad som den variant av brett reflektivt ekvilibrium som Daniels fört fram, ofta noteras inte ens skillnaden mellan dem,¹⁵ kanske delvis för att reflektivt ekvilibrium ofta ses som en form av koherentism och medan den första utvidgningen kan förstås inom ramen för detta, är denna andra utvidgning som sagt av ett annat slag. Är man intresserad av metoden just som en form av koherentism kan det alltså finnas skäl att inte anamma denna andra utvidgning alternativt att se den som ett tänkbart tillägg, men att det

¹⁴I slutändan landar ju Sidgwick i en klassisk utilitarism, men vägen dit går genom en systematisk jämförelse och diskussion av andra uppfattningar (även om man nog likt Rawls (1974–75, s. 9) kan invända att framför allt den form av egoism som Sidgwick diskuterar gärna kunnat ersättas med något slags perfektionism/eudaimonism).

¹⁵Tersman (1993, s. 252) är ett undantag; han påpekar att det nog finns en skillnad mellan Rawls och Daniels på denna punkt, men sedan är det Daniels han diskuterar. Rawls själv gör också i realiteten den första utvidgningen, och den är kanske rentav viktigare för hans teori, men även i den sista versionen som Rawls publicerade av teorin involverar hans föreställning om reflektivt ekvilibrium fortfarande också denna andra utvidgning (2001, s. 31).

är ett tillägg av ytterligare en metod snarare än en utvidgning av den ursprungliga. Samtidigt är det kanske rimligt att ett tillstånd av reflektivt ekvilibrium inte bara skall vara en fråga om en balans som råder i systemet som sådant, till följd av dess interna koherens, utan också en balans i relation till systemets omgivning som gör att det dessutom är en stabil jämvikt. I sig kan ju ett tillstånd av jämvikt vara ytterst känsligt för extern påverkan, men om man redan har utsatt sin åsiktsmängd för denna externa påverkan och därefter befinner sig i jämvikt, då tycks det finnas god grund för att uppfatta tillståndet ifråga som ett slags reflektiv slutpunkt och inte bara en tillfällig vilopunkt.

Är man i första hand intresserad av reflektivt ekvilibrium som ett arbetssätt och kanske oroas över den tidigare nämnda risken för att det kan bli en alltför privat angelägenhet då finns ytterligare skäl att följa Rawls och göra denna andra utvidgning i och med att den innebär en process där man inte bara arbetar med sin egen åsiktsmängd utan även tar in andras uppfattningar. Eftersom denna utvidgning handlar om ett slags överlevnadstest blir det naturligtvis värt något först då (i) personen som genomgår det verkligen går in i det på ett öppet och seriöst sätt, och (ii) det sker på ett ingående och systematiskt sätt som verkligen prövar personens uppfattningar. Och eftersom människors benägenhet kanske snarare är att söka bekräftelse än att verkligen pröva sina uppfattningar är det förmodligen en typ av process som bäst äger rum i ett socialt sammanhang, till exempel genom ett deltagande i den diskussion som sker inom den akademiska filosofin, där det alltid tycks finnas någon som är villig att försöka få en att inse orimligheten i ens uppfattningar. Denna andra utvidgning pekar alltså i riktning mot moraliskt och politiskt tänkande som en kollektiv process, där vi som företrädare för olika alternativa uppfattningar genom vår diskussion hjälper varandra till uppfattningar som kanske i slutändan inte är identiska men som fortfarande bygger på att vi tagit in varandras resonemang. En rimlig hypotes här är nog dessutom att en viss tendens till konvergens kommer att finnas om vi deltar i ett sådant utbyte och det tycks också sannolikt att olika former av moralisk och politisk extremism kommer att ha svårt att överleva i en sådan process (jämfört med en process som bara skulle involvera en själv eller ens gelikar).

5. UTVIDGNING AV MÅLSÄTTNING: GEMENSAMMA UPPFATTNINGAR

Även om reflektivt ekvilibrium som metod naturligtvis inte är något som Rawls har sista ordet om går det inte att komma runt att det till stor del är i hans arbeten som metoden har utarbetats. Det innebär också att den tenderar att evolvera tillsammans med hans teori om rättvisa. En av de

stora förändringarna som skett i hans filosofiska projekt är den starka betoningen i hans senare arbeten på den mångfald av fullt rimliga uppfattningar som föreligger i moderna samhällen. Han tänker sig då att i utarbetandet av moraliska och politiska teorier måste vi utgå från detta pluralismens faktum och försöka arbeta oss fram mot idéer som kan delas av alla människor som utgår från fullt rimliga uppfattningar. Denna utvidgning är inte något som Rawls talar om som en form av brett reflektivt ekvilibrium utan snarare som ett *allmänt* reflektivt ekvilibrium, men det utgör i vilket fall som helst en tredje form av utvidgning av vad ett reflektivt ekvilibrium bör handla om: idealet är alltså ett brett och allmänt reflektivt ekvilibrium (2001, s. 30–31).

I och med denna utvidgning tar inte bara rörelsen mot reflektivt ekvilibrium in andras åsikter som ett sätt att förbättra de egna utan hela processen blir till stor del till ett kollektivt projekt: att resonera sig fram till vad som rimligen kan vara vår gemensamma moraliska eller politiska hållning. Hos den senare Rawls är denna ambition knuten till hans idé om att på ett samhällligt plan uppnå vad han kallar för överlappande konsensus (1993, s. 144). Det handlar om ett tillstånd där vi visserligen har en rad filosofiska, existentiella eller religiösa föreställningar i ljuset av vilka de principer vi står för tycks oss rimliga, men där vi i relation till den offentliga diskussionen av principer skjuter dessa föreställningar åt sidan och försöker nå fram till en samstämmighet kring principerna för vårt samhälle i stället. Det kommer fortfarande att vara viktigt för oss att förankra denna mellannivå av delade principer i våra egna omdömen och mer generella föreställningar och teorier, men konsensus kan uppnås på den nivån just genom att vissa frågor skjuts åt sidan inom ramen för den offentliga diskussionen. Den grundläggande tankegången här går dock att tillämpa inte bara politiskt utan även moraliskt och ett par exempel på andra som resonerar på detta sätt är Martha Nussbaum (2000) i relation till utvecklingsfrågor och Charles Taylor (1996) i relation till mänskliga rättigheter.

Med denna tredje utvidgning blir processen till ett mer invecklat projekt. Den individuella versionen har en uppenbar fördel: individer har redan existerande åsiktsmängder på ett sätt som kollektivt inte har. Men det finns också uppenbara problem med att hålla projektet individuellt; dels är moralen och politiken saker som reglerar förhållanden mellan människor, dels handlar många av de svåra diskussioner som filosofer ofta hoppas kunna bidra till om hur vår gemensamma lagstiftning bör se ut. Det kan alltså tyckas finnas något poänglöst med att bara sträva efter sitt eget högst personliga reflektiva ekvilibrium, även om det är produkten av ett åsiktsutbyte med andra. Det kanske skulle vara ett enklare projekt, men man vill ju knappast vara som mannen som letar efter sina tappade nycklar under gatlyktan för att det är ljusst där, trots att han egentligen

är rätt säker på att han tappade dem borta vid dörren. Det individuella projektet kvarstår dessutom i viss utsträckning eftersom tanken med ett överlappande konsensus inte är att vi skall nå en gemensam hållning i allt utan att vi utformar våra individuella hållningar i relation till en gemensam hållning om vissa grundläggande institutioner och principer.

Givet en orientering mot allmänt reflektivt ekvilibrium är det tydligt att vi inte resonerar i vår privata bubbla, men även om man i teorin skulle kunna tänka sig en djupt dialogisk process där vi tillsammans genom ett ständigt utbyte arbetar utifrån vissa gemensamma åsikter i riktning mot ett reflektivt ekvilibrium är det i praktiken naturligtvis ogenomförbart, särskilt som det ”vi” vi talar om här inte är ”vi som sitter i seminarierummet” (redan där skulle det vara svårt) utan mer något i stil med ”vi människor” eller möjligen ”vi i vårt samhälle”. Sättet som den enskilde filosofen kommer att arbeta torde i stället vara relativt likartat som inom den individuella versionen, med skillnaden att man hela tiden måste sträva efter att ta in andra ståndpunkter och försöka göra reda för dem såtillvida att de tillämpbara principer man arbetar i riktning mot skall kunna vara acceptabla utifrån en mångfald av rimliga utgångspunkter. Man skulle kunna säga att det handlar om att arbeta både vertikalt (inom ramen för den egna åsiktsmängden) och horisontellt (inom ramen för den åsiktsmängdsnivå där vi tänks etablera ett överlappande konsensus med andra kring vissa institutioner eller principer). Det kan fortfarande vara så att man förkastar vissa ansatser, som kantian kanske man till exempel (likt Rawls) förkastar utilitarismen, men man måste samtidigt arbeta mot att hitta samhälleliga lösningar på moraliska och politiska frågor som kan vara rimliga även utifrån sådana hållningar som man själv förkastat.

Helt klart finns det en stark konvergenstanke inbyggd här och olika former av moralisk och politisk extremism torde ha svårt att hävda sig. Ofta kommer de nog inte ens att falla inom ramen för mångfalden av rimliga uppfattningar; för att göra det, åtminstone enligt Rawls,¹⁶ måste de nämligen kännetecknas av en grundläggande tolerans för andras uppfattningar och en öppenhet för att resonera tillsammans med andra inom offentlighetens arena. Här finns inte utrymme att gå djupare in på detta, men även om man oroar sig över hur snävt reflektivt ekvilibrium kan vara alltför internt kan man på denna punkt, liksom Chantal Mouffe (2005, s. 223), eventuellt oro sig över om inte Rawls i alltför hög grad bygger in en liberal hållning redan i föreställningen om rimlighet och att det trots allt på en metodologisk nivå inte är rimligt med en sådan syn på rimlighet. I princip kan man dock tänka sig att denna tredje utvidgning

¹⁶För mer om hur Rawls uppfattar vad som krävs för rimlighet, se (1993, s. 58–61).

görs samtidigt som man är mer öppen ifråga om vad som krävs för att en uppfattning skall räknas som rimlig och därmed ingå bland parterna i strävan mot överlappande konsensus. Det är i vilket fall som helst den utvidgning av de tre som tagits upp här som allra tydligast tar arbetet med våra åsiktsmängder ut ur vår egen privata bubbla genom att knyta det arbetet till ett gemensamt projekt.

6. MÖJLIGA SPÄNNINGAR

Givet att det finns skäl som talar för alla tre av dessa utvidgningar tagna var för sig är det inte då också rimligt att anamma alla tre som ett paket? På ett principiellt plan tycks det inte finnas något hinder för det och framför allt Rawls företräder ju tanken att vi bör göra det, men då skall det också sägas att hans hållning nog i viss mån bygger på dels att de relevanta bakgrundsteorierna är relativt tunna, dels att han mer eller mindre köper den typ av samhälle som redan föreligger i moderna västerländska nationalstater som utgångspunkt. En komplicerande faktor här är dessutom att särskilt i hans senare arbeten utarbetar han den politiska liberalismen som teori samtidigt som han utarbetar metateorierna om reflektivt ekvilibrium och överlappande konsensus. Flätar Rawls samman dessa eller blandar han samman dem? Åtminstone den tredje utvidgningen tycks ju onekligen bygga på ett demokratiskt ideal om offentlig åsiktsbildning där vi resonerar tillsammans utifrån en grundläggande jämlikhet och respekt för varandra, särskilt som Rawls är väldigt noga med att poängtera att ett överlappande konsensus bör vara något mer än en kompromiss eller ett *modus vivendi* (1993, s. 146–48). Man kan här, som redan påpekats, oroa sig för att idealet om ett överlappande konsensus, med den starka koppling det har till tanken om ett välordnat samhälle, i grund och botten inte är ett epistemiskt eller metodologiskt ideal utan ett politiskt och detta oavsett var man lägger ribban för att något skall räknas som en rimlig uppfattning.

Det finns dessutom en möjlig spänning mellan den första och den tredje utvidgningen. Det kan finnas fylliga bakgrundsteorier om såväl personskap som samhället som i slutändan visar sig vara fullt rimliga och som står i strid med olika trosuppfattningar som är vanligt förekommande i de flesta samhällen, särskilt eftersom de bakgrundsteorier som rent faktiskt spelar roll för människors moraliska och politiska åsikter inte sällan är metafysiska teorier av ett slag som är föga informerade av empirisk vetenskap utan snarare handlar om religiösa dogmer eller andra helt ogrundade föreställningar om människan, samhället och världsalltet. Hos Rawls och Daniels uppstår inte det problemet i och med att de rent substantiellt har en oerhört tunn teori om personskap, men dels kan man

ifrågasätta om teorin om personers plasticitet verkligen är rimlig, dels kan man ifrågasätta om det är rimligt att ansatsen som helhet, med alla tre utvidningar, faktiskt förutsätter att man anammar just den teorin.

En annan möjlig spänning föreligger mellan den kritiska diskussion som förmodligen krävs för att den andra utvidningen skall fungera på ett bra sätt och den ekumeniska hållning som den tredje utvidningen verkar, om inte förutsätta, så åtminstone starkt peka mot. Den engagerade debattörens sätt att sikta mot konsensus tenderar väl snarare att handla om en vilja att omvända andra till den egna ståndpunkten än att sikta mot en form av konsensus som tillåter oss alla att hålla kvar vid våra grundläggande utgångspunkter. Samtidigt förutsätter rimligen ett åsiktsutbyte som verkligen prövar våra hållningar och de skäl vi har för dem just engagerade debattörer. Det kan säkerligen finnas plats för den ekumeniska hållningen i vissa sammanhang, till exempel när man skall formulera principiella värdegrunder för vården¹⁷ eller andra verksamheter, men som en allmän metodologisk hållning är den mer tvivelaktig.

Den gemensamma nämnaren ifråga om dessa spänningar är alltså den tredje utvidningen och det gör att det också är den som det verkar rimligt att förhålla sig mest skeptisk till. Den typ av överlappande konsensus som Rawls förordar är en typ av konsensus som tycks vara tänkt att mer eller mindre kunna etableras här och nu, eller rent av redan föreligger i viss utsträckning i många av de västerländska samhällen vars politiska system ligger relativt väl i linje med den liberalism som Rawls företräder. Man kan dock också tänka sig att man i likhet med C. S. Peirce ser konsensus eller konvergens snarare som en bortre gräns för tänkandet och undersökandet.¹⁸ Givet den andra utvidningen kan det tyckas rimligt att anta att processen, framför allt givet hur människor fungerar som sociala var-elser, åtminstone har en inneboende tendens mot konvergens, även om det knappast finns skäl att anta att denna bortre gräns representerar mer än en möjlighet av att uppnå en stabil form av jämvikt. Möjligheten och eventuellt förhoppningen om denna form av konvergens handlar dock inte om att här och nu försöka jämka samman våra åsikter på vissa punkter, även om vi i praktiken naturligtvis alltid kommer att behöva nå olika kompromisser, utan om att vi genom argumentation och utbyte i slutändan mycket väl kan nå fram till en gemensam hållning. Detta är dock kanske något som man bäst förstår i termer av en inte helt ogrundad förhoppning snarare än något som skall byggas in i själva metoden som sådan.

Mellan den första och andra utvidningen tycks det inte föreligga några

¹⁷Ett exempel på detta är Beauchamp och Childress (2008).

¹⁸Hos Peirce (1878) finns dock en tanke om oundviklighet hos denna konvergens som eventuellt förutsätter någon form av realism, något det kan finnas skäl att inte förutsätta i ett moraliskt och politiskt sammanhang.

framträdande spänningar; problemet med att göra båda utvidgningarna är nog bara att processen därigenom blir mer komplex. Samtidigt kan, som redan påpekats, åtminstone den andra utvidgningen sägas utgöra ett relativt litet steg, om ens ett steg alls, givet att det handlar om akademisk filosofi. Man skall inte idealisera denna verksamhet alltför mycket och utbytet av åsikter inom den kan säkert ibland degenerera till klubbar av inbördes beundran snarare än att utgöra någon prövning i djupare bemärkelse. Som ideal är det dock knappast någon främmande fågel utan snarare redan inbyggt i själva verksamheten. Inom ramen för akademisk filosofi handlar frågan som uppstår alltså snarare om hur man bäst hittar former där olika filosofiska uppfattningar kan ställas och brytas mot varandra på ett sätt som innebär en genuin prövning av dem än om huruvida man skall ha en process där en sådan prövning ingår till att börja med.

LITTERATUR

- Beauchamp, T. och J. Childress. 2008. *Principles of Biomedical Ethics*, 6 uppl. New York: Oxford University Press.
- Daniels, N. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- Dancy, J. 2004. *Ethics without Principles*. Oxford: Oxford University Press.
- DePaul, M. 1986. "Reflective Equilibrium and Foundationalism". *American Philosophical Quarterly* 23.
- Eggleston, B. 2010. "Practical Equilibrium: A Way of Deciding What to Think about Morality". *Mind* 119.
- Elgin, C. 1996. *Considered Judgment*. Princeton: Princeton University Press.
- Holmgren, M. 1989. "The Wide and Warrow of Reflective Equilibrium", *Canadian Journal of Philosophy* 19.
- Mouffe, C. 2005. "The Limits of John Rawls's Pluralism". *Politics, Philosophy & Economics* 4.
- Nussbaum, M. 2000. *Women and Human Development*. Cambridge: Cambridge University Press.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Peirce, C. S. 1878. "How to Make Our Ideas Clear". *Popular Science Monthly* 12.
- Rawls, J. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Rawls, J. 1974-5. "The Independence of Moral Theory". *Proceedings and Addresses of the American Philosophical Association* 48.
- Rawls, J. 1993. *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. 2001. *Justice as Fairness: A Restatement*. Cambridge, Mass.: Harvard University Press.
- Sidgwick, H. 1907. *The Methods of Ethics*, 7 uppl. London: Macmillan.
- Taylor, C. 1996. "A World Consensus on Human Rights?". *Dissent* 43.
- Tersman, F. 1993. *Reflective Equilibrium: An Essay in Moral Epistemology*. Stockholm: Almqvist & Wiksell.