

Moraliskt hänsynstagande och dess gränser

Enligt populärt tänkande finns en kontrast mellan egoisten, som bara tänker på sig själv, som strävar efter att tillgodose sina egna intressen, helt oanfäktad av skrupler rörande de olägenheter detta kan medföra för andra individer, och den av moralen besjälade som också tar hänsyn till andra individer och deras intressen. En del filosofer – från Platon och framåt – har menat att denna kontrast bara existerar så länge man har en snöd eller inskränkt egoism i åtanke: för den rationelle och upplyste egoisten framstår motsättningen mellan egenintresse och moraliska hänsynstaganden som en chimär. Denna ståndpunkt har den fördelen att den inte behöver postulera attityder som sympati, välvilja eller medlidande hos en individ för att kunna förklara hur denne är i stånd till att handla moraliskt. Ingen betvivlar att människor som regel är benägna att försöka realisera vad de tar för att ligga i deras eget intresse, och om det nu förhåller sig så att det ligger i ens intresse att ta hänsyn till andra, så behöver en människa bara upplysas därom för att uppföra sig hänsynsfullt. Det som talar emot denna syn på relationen egoism – moraliskt hänsynstagande är intuitionen att det senare innebär att bedöma sitt handlande och dess konsekvenser även från andra inblandade varelsers perspektiv, och även om världen skulle vara så vist och förträffligt inrättad att det som är bäst ur agentens synvinkel också är bäst ur övriga inblandades synvinkel, så kan detta på sin höjd vara en lycklig tillfällighet, och ingenting som är definitionsmässigt garanterat.

Personligen är jag av den meningen att ett moraliskt förhållningssätt innefattar att man ser sitt handlande även ur andra inblandade varelsers perspektiv i en mening som spränger egoismens ramar. Jag skall emellertid inte här försöka bevisa riktigheten av denna uppfattning. Den uppgift jag ämnar ta itu med är i stället att försöka precisera den innebörd fraser som ”att bedöma något ur andra påverkade varelsers perspektiv” kan tänkas ha när de används för att karakterisera ett moraliskt förhållningssätt. Jag skall inte försöka påvisa att många människor har den altruistiska motivation som erfordras för att verkställa moralens bud. Det tycks mig dock uppenbart att människor kan hysa genuin välvilja för andra individer (men inte i en utsträckning som gör

det rättsliga maskineriet överflödigt). Kanhända denna välvilja har sin rot i att den andre tidigare behandlat en på ett sätt som man värderar positivt, men härav följer inte att den ur välvilja framsprungna handlingen har ett egoistiskt syfte som att tillförsäkra en framtida förmåner från den välvilligt behandlade.

1. En version av utilitarismen

En egoist strävar efter att tillgodose sitt egenintresse i största möjliga mån, han söker en maximal tillfredsställelse av egenintresset. Han bekymrar sig för hur en planerad handling påverkar andra varelsers intressen endast i så måtto som dessa varelsers reaktioner har konsekvenser för hans egenintressen. Den moraliske agenten däremot åstundar att tillvarata så många och så starka intressen som möjligt, oavsett vems intressen det är; han prioriterar inte vissa intressen bara för att de råkar vara hans egna utan hans bedömningar är baserade endast på fakta som styrkegrad och förenlighet med andra intressen. Kort sagt, medan egoisten jagar maximering av sin *individuella* intressetillfredsställelse, så eftersträvar den av moralen besjälade maximering av en *generell* intressetillfredsställelse (antingen i form av största summa eller högsta individuella genomsnitt). Jag kallar denna idé om moralens natur för *generalitetstesen*. Den utgör *en* precisering (men den mest plausibla jag kan tänka mig) av den utilitaristiska kärnkonceptionen – en precisering som kan benämnas den intresse-utilitaristiska. En annan möjlig precisering talar om maximering av välbehagssensationer i förhållande till smärt- eller obehagssensationer.

Innan generalitetstesen tas upp till kritisk prövning är det lämpligt att ge ett förtydligande av en punkt i den. Vad åsyftas med frasen "att ligga i någons intresse"? Frasen är definierbar i termer av viljeattityder. Förhållandet är dock inte så enkelt att man kan jämställa att *X* ligger i en agent *A*'s intresse med att *A* vill ha *X*. En faktor som komplicerar förhållandet är att en individ kan vara felinformerad på ett område som rör hans viljeställningar: om *A* vill inmundiga en viss tablett emedan han ser den som ett medel att förbättra sitt hälsotillstånd (något som han vill i sig självt, som yttersta ändamål), men intagandet av tabletten i själva verket skulle förvärra hans hälsotillstånd, så ligger det inte i vederbörandes intresse att ta tabletten. För att *X* skall ligga i *A*'s intresse är det nödvändigt antingen (a) att *A* vill *X* som yttersta ändamål, och att denna attityd inte uppkommit eller består genom informationsbrister, eller (b) att *X* är ett medel till något som *A* vill likmätigt (a). Detta krav

är emellertid inte tillräckligt: det måste tillfogas att om X är velat som yttersta ändamål, så får X inte inkludera någon referens till en annan individs viljeattityder. Man kan inte säga att det ligger i A 's intresse att tillfredsställa en av B 's viljeattityder, om detta är något som A vill enbart som ändamål. Däremot kan det ligga i A 's intresse att "smörja" B , dvs tillfredsställandet av viljeattityder hos B kan vara ett medel för A att uppnå egna fördelar. Denna senare inskränkning förenklar problematiken utan att utelämna något, ty en persons eventuella altruistiska omsorger beaktas indirekt genom att de omhuldades egoistiska strävanden tas med i beräkningen.

Enligt mitt förmenande är generalitetstesens ohållbar, och jag skall nu övergå till att argumentera för denna uppfattning. Min avsikt är dock inte att ge en uttömmande presentation av alla argument som kan tänkas tala mot tesen, utan jag skall koncentrera mig på ett par som introducerar överväganden vilka kommer att aktualiseras även längre fram. Tesen kan kritiseras för att den är svår att tillämpa i praktiken, enär den förutsätter t ex interpersonliga jämförelser av styrkegrader hos viljeattityder. Om jag har att välja mellan att stilla A 's nikotinbegär eller B 's, så är det väsentligt att kunna avgöra vems nikotinbegär som är starkast. Denna kunskap är inte principiellt otillgänglig: exakt kunskap om de neurofysiologiska mekanismer som är korrelerade med nikotinbegär skulle ge vägledning. Men denna kunskap är inte tillgänglig för tillfället, och även om den vore det, så skulle vi ändå möta svårigheten att det är en omständlig procedur att nå kunskap om givna individers neurologiska tillstånd – en svårighet som ytterligare accentueras av att en handling normalt inte påverkar blott ett par utan en mängd varelser. Under dessa omständigheter är risken uppenbar att ett betydande mått av godtycke och fördomar smyger sig in i intressekalkylen. En annan välbekant invändning är att alla fördelningsproblem knappast kan lösas på ett sätt som är tillfyllest, om inte generalitetsidén suppleras med en rättvisepincip. Men de invändningar jag vill trycka speciellt på är följande.

1) Generalitetstesens kommer att kräva alltför mycket av den som vill hålla sig inom moralens ramar i och med att den förefaller att negligera distinktionen mellan att göra precis vad man är moraliskt förpliktad att göra och att utföra övermeriterande handlingar, att göra "mer" än moralen fordrar. (Denna problematik uppmärksammas t ex i J O Urmsons "Saints and Heroes", omtryckt i J Feinberg (utg): *Moral Concepts*, Oxford 1969.) Antag t ex att A befinner sig i en situation där han har att välja mellan att rädda livet på andra närvarande personer och offra sitt eget, å ena sidan, och att rädda sitt eget skinn men offra de andras, å andra sidan; han står kanske inför ett val mellan att kasta sig på en exploderande granat eller att kasta sig bakom ett skydd. Då handlar

A, enligt generalitetstesen, omoraliskt om han väljer det senare alternativet, ty ett val av det förra skulle säkerligen maximera den allmänna tillfredsställelsen. Intuitivt ter det sig dock rimligare att säga att *A gör mer* än moralen kräver om han offerar sitt liv – hans beteende är heroiskt – och att ett uraktlåtande av denna gärning inte är omoraliskt, ehuru mindre prisvärt. Det medges att utilitaristen kan nyansera bilden genom att skilja mellan (a) att en handling är rätt eller fel, och (b) att den bör påbjudas respektive förbjudas, vilket senare definieras genom nyttan av positiva respektive negativa reaktioner på handlingen. En övermeriterande handling blir då en rätt handling, vars utförande emellertid inte är påbjudet utan bara rekommenderat, dvs den maximala tillfredsställdheten fordrar att utföranden av den beröms men inte att underlåtanden fördöms. Att klandra folk för uraktlåtandet av dylika handlingar skulle sätta dem under psykisk press och således inverka menligt på den allmänna tillfredsheten. Det tycks mig dock i högsta grad tveklaktigt om denna psykiska press – som knappast kan vara akut med tanke på sällsyntheten av situationer som den skisserade – kan sägas uppväga den vinning i människoliv den strängare attityden skulle medföra. Betänk också att en människa sällan upplever villkor som pressande om hon vant sig vid dem sedan barnsben.

2) Enligt generalitetstesen finns det bara två frågor att beakta när man har att ta ställning till om man skall tillgodose en annan individs intresse: "Hur starkt är intresset?" och "Vilka andra intressen (hos vederbörande och andra) är det förenligt med?" Om man finner att ett intresses stillande är förenligt med en maximering av den allmänna intressetillfredsställdheten, då är det moraliskt riktigt att stilla intresset. Men detta berövar en individ möjligheten att moraliskt kritisera det som maximerar den allmänna intressetillfredsställelsen. En klarsynt individ kan inte hylla något som ett moraliskt ideal i den bemärkelsen att han anser sig berättigad att vägra att bidra till förverkligandet av *X*, oavsett hur många som har intresse av *X*. Inget engagemang hos en person kan med berättigande ställas över och utanför den form av kompromissanda och villighet att köpslå som ligger i att betrakta engagemanget blott och bart som ett element, vilket kan slås ut av andra, i en intressekalkyl. Intresse-utilitarismen är något av en smilfinksmoral, en moral som implicerar att de förhärskande intressena ovillkorligen måste strykas medhårs. Det är vanskligt att illustrera den här tankegången med slagkraftiga exempel, exempel som omöjliggör den utilitaristiska riposten att det inte rör sig om intressen utan bara om viljeattityder grundade på felinformation. Det är plausibelt att tänka sig – fastän somliga skulle mena att det är att hemfalla åt ohållbart elitänkande – att den breda massan inte har tillräcklig begåvning och sensibilitet för att uppskatta de

mest avancerade konstverken, att dess dokumenterade vurm för skäliggen andefattig konst inte bara beror på att den är vilseförd, otillräckligt upplyst etc. I så fall kan en konstnär ställas inför följande dilemma: antingen producera verk av typen *K*, vilket den konstnärliga integriteten och trohet gentemot den personliga visionen kräver, eller verk av typen *L*, vilka tilltalar det stora flertalets smak. Enligt utilitarismen är det omoraliskt att välja det förra alternativet; konstnären kan inte med fog hävda att autenticitet har större moralisk vikt än tillfredsställdheten av den förhärskande smaken. Men den sistnämnda ståndpunkten tycks mig, vare sig man instämmer i den eller ej, vara en moraliskt möjlig ståndpunkt.

2. Två former av universaliserbarhet

Dessa argument är knappast tillräckliga för att konklusivt vederlägga intresse-utilitarismen – troligen rör vi oss på ett gebit där konklusiv bevisföring är principiellt omöjlig, till följd av att känslö- och viljeattityder spelar en vital roll och att nyckelbegreppen är oförbättrligt nebulösa. Men argumenten tycks mig starka nog för att motivera att man ser sig om efter något som kan träda till i generalitetstesens ställe. Det finns en princip som sedan urminnes tider intagit en central plats i etiskt tänkande – jag åsyftar den idé som kommer till uttryck i Konfucius' sats att man inte skall göra mot andra vad man inte vill att de skall göra mot en själv (se tex Wing-Tsit Chan: *A Source Book in Chinese Philosophy*, Princeton U P 1963, s 44) och som i positiv form utgör kristendomens gyllene regel. Jag skall begagna termen "(viljans) universaliserbarhet" för att beteckna denna idé och preciseringar av den. Ty det är inte klart exakt hur denna regel skall förstås, men ett förslag är detta. Antag att *A* skall försäkra sig om att en handling, som han står i begrepp att utföra och som bara inverkar på *B*, är moraliskt riktig. *A* har då att fråga sig "Skulle jag, om jag med bibehållande av min nuvarande personlighet och uppsättning av viljeattityder befann mig i en position som i alla övriga hänseenden var identisk med *B*'s, kunna vilja eller samtycka till att bli behandlad så som handlingen ifråga behandlar *B*?". Om svaret är jakande, har handlingen moralens bifall.

Det förtjänar att understrykas hur denna idé om universaliserbarhet skiljer sig från generalitetstesen. I båda fallen kan man sägas ta hänsyn till vilka konsekvenser de till buds stående handlingsalternativen har för andra individer. Men medan generalitetstesen lägger vikten vid hur de av handlingen påverkade individerna själva, med de viljeattityder *de*

faktiskt innehar, skulle värdera dessa konsekvenser, så lägger universaliserbarhetsidén tonvikten vid hur endast *den handlande själv*, med de viljeattityder *han* har, skulle värdera dessa konsekvenser om han befann sig i de behandlades ställen.

Från denna utläggning torde det omedelbart framgå att argumentet 2) i föregående sektion inte träffar universaliserbarhetsidén. Enligt denna idé gäller det inte att betjäna *andras* intressen utan att tillmötesgå de intressen *man själv* skulle ha i en hypotetisk situation, nämligen om man befann sig i de behandlades omständigheter. Och man kan inte inta en kritisk distans till sina egna intressen så som man kan till *andras* intressen, av den enkla anledningen att de är ens egna intressen. Man kan inte samtidigt vilja att p och vilja att man inte vill att p (jag talar här, liksom i uppsatsen i övrigt, enbart om rationella viljeattityder, dvs viljeattityder som är modifierbara i ljuset av skäl). Ty de skäl som underblåser och befäster attityden av andra ordningen underminerar i samma mån attityden av första ordningen och tvärtom. Det finns därför en konstighet hos "Jag vill ha X men X är inget som är värt att (vilja) ha" som är frånvarande hos "Andra individer vill ha X men X är inget som är värt att (vilja) ha".

Universaliserbarhetsidén tycks emellertid vara öppen för invändningen att dess undvikande av argument 2) sker till priset av en total kallsinighet inför andra individers viljeattityder. Antag att A vill bli dödad om han ådrar sig vissa men, blir blind, förlamad etc, medan B vill fortsätta leva även i dessa trista omständigheter; då förefaller det som om A 's handling är att döda B , när B hemsöks av någon av de nämnda olyckorna, blir moraliskt sanktionerad enligt universaliserbarhetsidén, ty A vill ju själv bli dödad i dessa omständigheter. Men det verkar helt groteskt att detta skulle vara tillräckligt för att moraliskt legitimera A 's handling att döda B mot hans vilja, om B ådrar sig ett fysiskt handikapp. Det är kanske med exempel som detta i åtanke som G B Shaw har sagt "Handla inte mot andra som du vill att de skall handla mot dig. Deras smak är kanske inte densamma som din". Universaliserbarhetsidén tycks inte lämna något rum för hänsyn till *andras* intressen när dessa inte sammanfaller med de den handlade själv skulle ha om han befann sig i de behandlades situation. Den utilitaristiska underkastelsen under *andras* intressen (givet att de har tillräcklig vid spridning), vad de än månne vara, tycks ha avlösts av nära nog sin motsats: total okänslighet inför intressen som skiljer sig från ens egna.

Svårigheten kan också formuleras så här. A 's handling är av typen att döda någon som är handikappad, men den är också av typen att döda någon, som är handikappad, *mot hans vilja*. A kan fråga sig om han är beredd att bli dödad om han själv blir handikappad, men han kan inte,

utan att spränga ramarna för den idé som nu är under granskning, ställa sig frågan om han kan samtycka till att bli dödad om (i) han själv blir handikappad och (ii) han då vill fortsätta leva, ty det är motsägelsefullt att samtidigt överflytta de nuvarande attityderna (i oförändrat skick) till den hypotetiska situationen och anta att ens attityder där skiljer sig från de nuvarande. Ändå är det förhållandet, att B vill fortsätta leva även som handikappad, uppenbarligen en relevant omständighet, en omständighet som man måste ta i beaktande om man har att bedöma riktigheten av handlingen att bringa B om livet. Det perspektiv den konfucianska eller kristna principen, så som jag hittills har utläst den, anlägger på relevanta omständigheter är således inte tillräckligt vidsträckt.

Det finns emellertid en annan, vid första påseende snarlik, tanke som med samma rätt kan göra anspråk på att vara en uttolkning av den urgamla princip som varit föremål för behandling. Jämför följande två satser:

- a) Givet sina nuvarande viljeattityder skulle A , om p , vilja q ;
- b) A 's nuvarande viljeattityder är sådana att han vill att om p så q (eller skulle vilja om p så q , om hans uppmärksamhet fästes vid detta förhållande – men låt oss för enkelhets skull bortse från denna variant).

Det ligger i öppen dag att a) och b) inte är likabetydande. I a) är p tillsammans med A 's nuvarande viljeattityder en betingelse för att A skall ha en viss viljeställning, nämligen vilja q . Placeringen av variabeln " p " är emellertid annorlunda i b): här figurerar den som en komponent i den fras som anger *innehållet* i en av A 's viljeattityder; vi har alltså här en viljeattityd med ett hypotetiskt innehåll.

Denna skillnad i variabelns placering är betydelsefull, ty den medför att det finns en klass av insättningar för " p " som ger b) empiriskt sanningsvärde, utan att ge a) det. Klassen utgörs av de satser som implicerar att A 's nuvarande viljeattityder genomgår en förändring. Dessa satser skapar komplikationer för a), eftersom p här är ett villkor tillsammans med förhållandet att A 's attityder är de nuvarande, dvs att hans attityder bevaras intakta. Det totala villkoret för att A skall vilja q blir därför motsägelsefullt. Det är kontroversiellt om hypotetiska satser med självmotsägande försatser är logiskt sanna eller logiskt falska, men deras sanningsvärde beror i alla händelser inte av empiriska omständigheter – något som gör dem otjänliga för mina syften. Insättningar av samma satser i b) vållar däremot inga besvär. Text så kan man i b) för " p " substituera " A blir fysiskt handikappad och ändrar viljeställning till att

vilja fortsätta leva” och för ” q ” ” A blir dödad”, ty en person kan inte bara vilja att han blir dödad om han blir handikappad, han kan också vilja att detta sker även om han i dessa omständigheter skulle komma att ändra inställning. Det ligger inget märkligt i att vid en tidpunkt vilja att de icke-identiska viljeattityder man har vid en *annan* (faktiskt eller tänkt) tidpunkt skall frustreras. Man kan döma ut en möjlig utveckling av en själv som degeneration och förfall. Kontrastera detta mot vad man med samma substitutioner erhåller ur a): givet att A 's nuvarande viljeattityder förblir konstanta, skulle A , om han blir handikappad och hans nuvarande viljeställning ändras så att han vill fortsätta leva, vilja bli dödad. Här görs det motsägelsefulla att antagandet att A 's viljeattityder både förändras och förblir konstanta.

Efter denna utläggning bör det framgå att jag ovan begagnade mig av a)-formen för att uttyda den konfucianska och kristna tankegången. De svårigheter som denna tolkning råkade in i är de som just reproduceras för a)-formens vidkommande. Låt oss nu se om b)-formen ger bättre underlag för en tolkning av denna centrala etiska idé. En formulering av idén som ligger nära till hands i detta sammanhang är följande: A 's vilja att utföra handlingen H har moralens sanktion om och endast om A 's vilja är universaliserbar i den meningen att A är villig att acceptera att H utförs även om han befinner sig i vilken som helst av de situationer, som i alla relevanta hänseenden är identisk med en av de positioner som en viljande varelse, vilken påverkas av H i den faktiska situationen, befinner sig i. Vid närmare besiktning visar det sig emellertid att denna formulering måste kasseras därför att den inte utesluter något som man kan kalla ”spekulation i osannolikheter”. Antag att q representerar något A skyr eller vill undvika; då blir det lättare för A att vilja att om p så q ju mer osannolikt inträffandet av p är. Av detta följer att det kan vara rationellt att vilja ha en smärre fördel som är praktiskt taget viss i den faktiska situationen, fastän man härmed riskerar att ådra sig en större nackdel om något osannolikt inträffar. För att illustrera, en person vars välbefinnande vilar på en säker grund kan resonera så här: ”Att avstå en del av mitt välbefinnande till denna hungrande människa innebär en uppoffring som väger lätt mot den skada jag åsamkas i en hypotetisk situation, där jag själv hungrar och välbefinnande människor förvägrar mig sin hjälp. Men eftersom det är så osannolikt att jag kommer att hamna i den hungrandes situation, så är jag villig att ta risken att föreskriva att jag då inte skall ges någon hjälp. Jag kan alltså med gott samvete låta bli att dela med mig av mitt välbefinnande”. Enligt ovanstående formulering drivs den här personen av moraliska bevekelsegrunder — en konsekvens som är klart grotesk.

Källan till problemen ligger i den roll sannolikhetsöväväganden till-låts spela. Helt klart är sannolikheten för att agenten skall komma att hamna vid den mottagande änden moraliskt oväsentlig, liksom sannolikheten för att andra då kommer att behandla honom så som han nu planerar att behandla de inblandade. Så länge dessa sannolikheter influerar beslutsprocessen rör den sig på en rent egoistisk basis.

Dessa sannolikhetsöväväganden spärras emellertid ut om man kräver att universaliserbarhetsresonemanget förlöper i enlighet med följande modell. I den första fasen sätter sig agenten in i var och en av de positioner som i alla relevanta avseenden är identisk med den något viljande väsen, vars intressen påverkas av handlingsalternativen i den faktiska situationen, befinner sig i. För var och en av dessa positioner frågar han sig: hur vill jag nu helst att jag blir behandlad om jag befinner mig i denna position? Givetvis får man inte föreställa sig att svaret nödvändigtvis blir detsamma för alla positioner, men vi kan bortse från denna komplikation eftersom den är parallell med en som uppstår i den andra fasen. Så antag att agenten når fram till ett favoriserat alternativ. Detta kan vara identiskt med det alternativ som favoriseras i den faktiska situationen – i så fall är beslutsprocessen avslutad. Men ofta är dessa preferenser oförenliga, varför det krävs en andra fas där man jämför värdeskillnaden mellan det favoriserade alternativet och de övriga i den hypotetiska situationen med värdeskillnaden mellan det favoriserade alternativet och de övriga i den faktiska situationen. Annorlunda uttryckt: man frågar beträffande den dominerande viljeattityden både i den hypotetiska och i den faktiska situationen vilken som har störst dominans över alternativen. Den viljeattityd som har den största dominansen är den universaliserbara viljeattityden (under förutsättning att ingen kompromiss är möjlig). De besvärande sannolikhetsövävägandena kan inte göra intrång i denna process. De kan inte tränga in i första fasen eftersom man här rör sig *inom* en viss hypotetisk situations ramar: man frågar sig vad som föredras bland sakförhållanden som alla delar samma hypotetiska villkor. Men de kan heller inte krypa in i den andra fasen som bara sammanställer värdeskillnaderna i den hypotetiska och faktiska situationen.

En illustration med ett tidigare exempel indikerar också detta. Den välbärgade har att jämföra den styrka med vilken han föredrar att behålla hela sitt välstånd i oreducerat skick framför att ge avkall på en ringa, obehövälig del av det, som krävs för att bistå en hungrande, med den styrka med vilken han, i en situation där han hungrar, sannolikt föredrar att han får sin hunger stillad genom en gåva från en välbärgad framför att han inte får något bistånd utan får fortsätta genomlida hungerns kval. Det råder inte någon tvekan om att, för en öväväldig-

de majoritet människor, den senare preferensens dominans är starkast, varför den universaliserbara attityden för dem blir att vilja bispringa hungrande – en konklusion som bättre harmonierar med den moraliska intuitionen än den tidigare slutsatsen.

Det står emellertid – för att återvända till ett exempel ännu längre tillbaka – fortfarande *A* fritt att, utan att kränka universaliserbarheten, bringa den handikappade *B* om livet, under förutsättning att han själv vill bli så behandlad i en situation där han är handikappad men har obruten livsvilja. (Det bör emellertid påpekas att, för de flesta människor, viljan att i dessa omständigheter deras vilja till fortsatt liv respekteras har större dominans över alternativ än en eventuell vilja i den faktiska situationen att bli kvitt ofärdiga människor har.) Innebär detta att vi alltså har ett förslag på moralkriterium som godkänner ett ogenerat trampande på andras intressen? Det är sant att den uppfattning av moral jag nu laborerar med tillåter en att strunta i andras intressen, hur allmänt förekommande de än må vara. Men argumentet 2) i föregående sektion visar att detta är ett nödvändigt krav på en definition av moral. Å andra sidan måste det finnas ett skydd mot missbruk av denna möjlighet att åsidosätta andras intressen, något som förhindrar att andra viljande varelsers intressen lättvindigt och skamlöst körs över. Denna garant har vi, tycks det mig, i klausulen att man har rätt att ignorera ett intresse endast om man (i beskriven mening) är villig att acceptera att det ignoreras även om man själv skulle komma att omfatta det. Att man är beredd att vilja att ett intresse inte tillmötesgås, även om man själv skulle komma att dela det, borgar för att ens negligering av intresset inte bara grundar sig på att det tjänar ens egoistiska syften för tillfället att köra över intresset. Om något förlänar en den moraliska rätten att inte respektera en annans intresse, så är det, förefaller det mig, ens villighet att acceptera att detta intresse heller inte respekteras om man själv skulle komma att omfatta det.

Så det kan kanske vara mödan värt att försöka formulera en definition som i kondenserad form återger denna andra explikation av den konfucianska och kristna idén:

(M) *A*'s vilja att utföra handlingen *H* (framför varje alternativ) är moralisk (= moraliskt oklanderlig) om och endast om *A*'s vilja att utföra *H* är universaliserbar i följande mening: *A* skulle vilja att *H* utförs hellre än varje alternativ även för varje hypotetisk situation som består i att *A* intar en position som, enligt en rimlig bedömning vid handlingstillfället, i alla relevanta hänseenden är identisk med den någon viljande varelse, vars intresse påverkas av *H* och/eller dess alternativ i den faktiska situationen, intar; eller, om *A* skulle

vilja en alternativ handling för dessa hypotetiska situationer, så har *A*'s vilja att utföra *H* större dominans över alternativ än denna eventuella vilja har.

En del punkter i (M) tarvar kommentar. (i) Jag definierar den moraliska karaktären hos viljeattityder snarare än hos handlingar, vilket kanhända är vanligare, enär jag anser att en handlings moraliska status är härledd från moraliteten hos den viljeattityd den är en intentionell manifestation av. (ii) Observera att definitionen fixerar innebörden hos termen "moralisk" när denna är synonym med "moraliskt oklanderlig (acceptabel etc)" och kontrasteras mot "omoralisk", inte i den (filosofiska) användning av "moralisk" som täcker både det moraliskt acceptabla och det moraliskt oacceptabla. (iii) Notera vidare att *A*'s viljeinställning kan vara moralisk utan att *A* *de facto* företagit tankeexperimentet att sätta sig in i de övriga inblandades situation: det talas inte om vad *A* "vill" utan vad han "skulle vilja", dvs skulle vilja om han företar det beskrivna tankeexperimentet. Det krävs inte att *A* faktiskt universaliserar sin attityd, bara att den är universaliserbar. (iv) Med ett "relevant" hänseende menas ett hänseende som påverkar *A*'s viljeattityder i positiv eller negativ riktning. (v) Personligen finner jag det naturligtast att tala om en moralisk valsituation endast om något viljande väsen utöver agenten påverkas av till buds stående alternativ; därför avser jag att frasen "för varje hypotetisk situation . . ." läses så att den implicerar att det finns dylika situationer. (vi) Frasen "enligt en rimlig bedömning vid handlingstillfället" är inskjuten för att poängtera att en agent som regel inte har någon exaktare information om de övriga inblandades intressen – ibland, som när det gäller beteende visavi offentliga serviceinrättningar, kan de inblandade inte ens identifieras. Allt som kan begäras av en agent är därför att han på basis av tillgänglig information gör en sund sannolikhetsbedömning av "insyltade" intressen.

En typisk filosofisk reaktion på en definition som (M) är att hävda något i stil med följande: termen "moralisk(t oklanderlig)" har primärt en (positiv) emotiv innebörd och bara i andra hand en viss beskrivande mening, vilken är beroende av den emotiva; därför blir en definition som (M) ovillkorligen persuasiv, dvs den blir ett förtäckt försök att rekommendera ett visst förhållningssätt. Som svar på detta är jag, för det första, beredd att tillstå att jag med (M), och denna uppsats som helhet, har velat göra reklam för anläggandet av ett visst betraktelsesätt när man formar sina attityder. Detta betraktelsesätt bör kunna vinna gehör hos alla och envar som förespråkar rationalitet och konsekvens i handlings- och viljelivet. Betraktelsesättet betonar rationalitet och konsekvens på så vis att det går ut på att man vill behandla en annan viljande varelse

annorlunda än man själv vill bli behandlad endast om man kan peka på någon egenskap, som denna varelse har till skillnad från en själv, men som det är logiskt möjligt att man själv kan vara utrustad med, i händelse av vilket man villigt accepterar samma behandling som man vill utsätta den andre för.

Den andra punkten i mitt svar är att jag önskar framhålla att detta medgivande inte utesluter att (M), fattad som en språkbruksbeskrivande definition, har ett betydande mått av sanning. Jag tror att termer som ”moralisk” i vardagsspråket också har en stark deskriptiv förankring, och jag vill förfäktat att (M) presenterar – kanhända i en preciserad och uppstramad form – åtminstone *en* dylik deskriptiv innebörd. Detta senare påstående bekräftas av att (M) kan ses som en explikation av en konfuciansk och kristen princip, som redan vid det etiska tänkandets begynnelse ställdes i centrum. Liknande tankegångar har sedan förts fram bla av Kant och R M Hare. (Den senare har myntat termen ”universaliserbarhet”; se *tex* hans *Freedom and Reason*, Oxford 1963. Att jag här undviker att redogöra för Kants och Hares doktriner beror inte på att jag försöker förtiga eller nedtona den inspiration jag hämtat från dem, utan på att en kartläggning av lik- och olikheter skulle föra alltför långt.) Låt emellertid inte detta skymma den rekommenderande eller propagerande aspekten hos (M): jag har lyft fram, och kanske hyfsat till, en bland förmodligen andra möjliga deskriptiva betydelse hos ”moralisk” emedan jag uppfattar den attityd som då betecknas som väsentlig och värd begrundan.

3. Några möjliga svårigheter för (M)

Jag har för avsikt att diskutera några svårigheter som man kan mena ansätter (M).

1) Först bör vi kanske försöka försäkra oss om att (M) går fri från invändning 1) som riktades mot intresse-utilitarismen i sektion 1 eller, *m a o*, om att (M) tillåter en plausibel analys av övermeriterande handlingar. Begrunda återigen *A* som står inför valet om han skall kasta sig över en exploderande granat eller söka skydd för den. Vi föreställer oss att *A* stannar för det senare alternativet – har han då, enligt (M) liksom enligt intresse-utilitarismen, handlat omoraliskt? Inte nödvändigtvis, svaret avhänger *A*'s viljeattityder beträffande de hypotetiska situationer där han ikläder sig de andra närvarandes relevanta karakteristika. Dessa personer, tänker vi oss, vill leva. Eftersom *A* vill detsamma, måste han tillmäta de andra personernas vilja att leva betydande vikt. Förmodligen

blir resultatet att, för dessa hypotetiska situationer, *A*'s dominerande vilja är att den, som här befinner sig i samma position som han själv i den faktiska situationen, utför offergärningen. Det finns således en konflikt mellan *A*'s dominerande viljeinställning i den faktiska situationen och den dominerande i de hypotetiska situationerna. När man undersöker vilken av dessa dominerande viljeattityder som är den starkaste, bör man göra klart för sig att det finns omständigheter som modererar den senare viljeattityden: med tanke på den situation *A* faktiskt befinner sig i, måste han vara medveten om hur svårt det är att offra sig för andra. Därför är det rimligt att anta att *A*'s vilja, för de hypotetiska situationerna, att bli räddad möter motstånd från en motvilja mot att ställa så stora krav på en annan människa som hans räddning kräver. Detta medför att hans vilja att rädda sitt eget skinn i den faktiska situationen kan få den största dominansen och följaktligen bli den universaliserbara (= moraliska) attityden.

Detta visar att (M), i kontrast till utilitarismen, inte fordrar att *A* (ens om han har genomsnittliga attityder) måste offra sig för att kunna sägas ha en moralisk inställning, men det belyser inte innebörden av termen "övermeriterande handling". För att klargöra denna för jag in en mer heroisk agent, *B*, som i en parallell situation väljer att rädda de andras liv. *B*'s attityd är också universaliserbar – den är ty o m lättare att universalisera än *A*'s, eftersom det här är mindre risk att den dominerande viljeattityden i den faktiska respektive hypotetiska situationen står i harnesk mot varandra. Jag vill nu hävda att *B*'s förhållningssätt inte är övermeriterande i förhållande till *hans egen* moral: han gör inte mer utan precis vad hans hjältemodiga hållning kräver. I stället är *B*'s handling övermeriterande *relativt A's moral*: *B*'s handling gör "mer" än *A*'s moral fordrar i den bemärkelsen att den tillgodoser viljor (nämligen de andra närvarandes livsviljor), som *A* själv håller för legitima, i större utsträckning än *A*'s egen universaliserbara handling gör. Min tes är alltså att en agents handlingar och attityder inte är övermeriterande relativt hans egen moralkodex utan relativt andra mindre hjältemodiga eller helgonlika individers moralkodex vilka bildar genomsnittet.

2) Någon tror sig kanske kunna ådagalägga att (M) är alltför snäv genom att visa på att (M) näppeligen kan välsigna de extrema åtgärder man kan vara nödsakad att tillgripa i nödvärn, åtgärder som dock intuitivt ter sig moraliskt berättigade. Antag att *A* angriper *B* i en uttalad avsikt att bringa honom om livet och att *B* kan avvärja detta anslag endast genom en handling som tillfogar *A* stort lidande, exempelvis handlingen att kasta frätande syra i ansiktet på *A*. Intuitivt förefaller *B*'s handling, trots dess grymhet, ha moralens bifall därför att den är nödvändig ur självförsvarysynpunkt, men det kan tyckas att den bara i

exceptionella omständigheter kan vara i samklang med (M). De flesta människor har en oerhört stark motvilja mot att få ansiktet sönderfrätt av syra, en motvilja så stark att det kan synas att dess dominans över andra alternativ i den hypotetiska situationen knappast kan överträffas av den dominans viljan att rädda det egna livet har i den situation *B* faktiskt befinner sig i. Risken verkar därför överhängande att *B*'s vilja paralyseras av en olöslig konflikt, om han försöker universalisera. Endast om *B* tillhör en fåtalig skara av moralatleter, hos vilka en motvilja mot att förgripa sig på en annan varelse effektivt motverkar motviljan mot att få syra i ansiktet, tycks dominansförhållandet bli sådant att hans vilja att rädda sitt liv i den faktiska situationen blir tydligt universaliserbar. Så det verkar som om (M) leder till att det moraliska berättigandet av extrema åtgärder företagna i nödvärn blir en alltför exklusiv vara.

Men skenet bedrar, ty i själva verket är det inte alls speciellt svårt för *B* att kunna samtycka till att få syra i ansiktet om han är ute i syfte att döda en annan person. Anledningen till detta är att en avsikt eller ett syfte som regel inte är något som hemsöker en utan ens egen förskyllan: en avsikt är normalt något man själv formar och som man därför har ansvar för. Vidare, *B* måste redan hysa en viss ovilja mot att forma intentionen att döda en annan person – annars skulle han aldrig undersöka sina attityders universaliserbarhet. Under dessa betingelser utgör motviljan mot att få syra i ansiktet inget hinder för att acceptera denna behandling om man gör sig skyldig till innehav av avsikten att döda en annan – *B*'s acceptering av avskyvärda följder för sin egen del i denna hypotetiska situation får bara till resultat att hans beslut, att se till att denna situation aldrig förverkligas, förstärks.

Man kan invända att fastän intentioner normalt inte uppkommer inom en utan ens egen förskyllan, så *kan* detta ske tex genom en ondskefull neurologs ingrepp. Detta måste medges, men om ens personlighet radikalt förändras genom ett dylikt ingrepp, så finns det inte mycket skäl att identifiera sig och sympatisera med den resulterande individen.

3) En invändning som slår i motsatt riktning är den som går ut på att (M) är otillräckligt restriktiv i det att den lämnar dörren öppen för moralisk relativism. Relativism blir en följd av (M) enär den inte ställer några innehållsliga krav på en moralisk attityd utan bara ett rent formellt krav på konsekvens: att en agent är beredd att vilja att den planerade handlingen exekveras även i de tänkbara situationer där han själv befinner sig vid den mottagande änden. Antag att *A* tilldelar kunskap högre värde i sig än lycka: han föredrar alltid att ta del av sanningen även om det innebär slutet på ett aldrig så intensivt lyckorus han befinner sig i. *B*'s värdering av kunskap och lycka är den omvända: han önskar sväva

vidare i rosenskimrande fåkunnighet framför den abrupta buklandning på ett skrovligt stycke verklighet som meddelandet av en obehaglig sanning innebär. Det är sannolikt att båda dessa attityder – och den behandling av andra varelser de ger upphov till – kommer att väcka anstöt hos den andre: *A* vill kanske brännmärka *B*'s attityd som "överdriven eftergivenhet gentemot animala föreställningar", medan *B* måhända fördömer *A*'s inställning som "hysterisk och hänsynslös förnuftsdyrkan". Men om (M) är korrekt, kan både *A* och *B* tvingas medge att den andres inställning är *moraliskt* oklanderlig, dvs att ingendera inställning kan kritiseras på *moraliska* grunder, fastän de därigenom naturligtvis inte nödgas överta den andres attityd som sin egen. Är detta övermåttan liberal: kan högst en av *A*'s och *B*'s attityder vara *moralisk*?

Jag har svårt att se vad som skulle kunna motivera ett jakande svar. Såvitt jag kan se, kan man inte stödja en karakterisering av en av attityderna som *den* moraliskt riktiga genom någon som helst argumentation. Det rör sig om värden i sig, om saker som är velade som yttersta ändamål, och sådana kan man knappast argumentera för eller mot eftersom argumentation förutsätter att man appellerar till andra värden. Vidare tänkte vi oss att varken *A* eller *B* kan anklagas för att göra godtyckliga undantag. Så hur kan man rättfärdiga klassifikationen av en attityd som den moraliska? Och vad kan denna klassifikation bli förutom ett gillandets sigill, dvs ett tecken på att talaren själv delar den så klassificerade attityden (något som kan uttryckas på mindre vilseledande sätt)? Det tycks mig därför vara ett utslag av förmyndarmentalitet och självrättfärdighet att beskriva den ena attityden (av en lycklig tillfällighet den man själv omfattar) som den moraliska. En ärlig människa måste hålla till godo med den relativism som härrör från det förhållandet att när alla sakförhållanden är utredda, när alla argument ligger på bordet, så kan en bitter bottensats av oöverbärliga viljemotsättningar kvarstå. Drömmen om ett absolut moralsystem är inget annat än en ren dröm. Den är ett stycke önsketänkande, liksom den inledningsvis nämnda tankegången att det moralen kräver sammanfaller med det det upplysta egenintresset kräver, att attityder som välvilja, medkänsla och sympati inte är nödvändiga för att motivera moraliskt beteende.

4) En annan möjlig anklagelsepunkt är att kriteriet (M) trots allt inte utgör en effektiv spärr mot nonchalans och åsidosättande av andras intressen. Roten till denna otillräcklighet ligger i att det är så vanligt förekommande att personer utmålar sig själva som mer heroiska, mer kapabla att utstå lidande och umbäranden i hypotetiska situationer än vad som faktiskt är fallet. Det finns dessa två felkällor i universaliseringsprocedurén: (i) agänten brister i självkänedom eller uppriktighet mot sig själv och tillskriver sig själv viljeattityder för hypotetiska situ-

ationer som han faktiskt inte kan hysa, och (ii) agenten är otillräckligt utrustad med fantasi och inlevelseförmåga: han åskådliggör förhållanden i hypotetiska situationer på ett inadekvat sätt. Båda dessa faktorer leder lätt till att en agent ställer alltför stränga krav på dem som befinner sig vid den mottagande änden och, vilket är sakens andra sida, tillåter sig själv alltför stora friheter.

Dessa iakttagelser är i stort riktiga: (M) är inget kriterium som kan appliceras mekaniskt och rutinmässigt; tvärtom, dess korrekta tillämpande tarvar som sagt självkänedom och inlevelseförmåga. Att inte universaliserbarheten utgör en mekaniskt tillämpbar avgörelsemetod kan beklagas av två anledningar: (a) det är angeläget att ge seriöst syftande agenter en så pålitlig vägledning som möjligt, och (b) det är önskvärt att beskära skrupelfria individers svängrum så långt det är möjligt. I realiteten väger emellertid (b) lätt, för även om den samvetlösse erbjuds liten chans till slingringsmanövrer vid *avgörandet* av vad som är moraliskt, så beror ändå *iscensättandet* av den nådda konklusionen på hans goda vilja, på hans sympati och medkänsla. Men kvar står att det av anledning (a) är beklagansvärt att (M) inte kan tillämpas med sömngångaraktig säkerhet. Jag tror emellertid att detta är en brist som varje moraliskt kriterium ofrånkomligen är behäftat med – åtminstone gäller, som jag tidigare antydde, detsamma för intresse-utilitarismen – ty det rör sig helt enkelt om en avspeglning av komplexiteten och svårgripbarheten hos etiskt relevanta fakta.

Däremot avslöjar dessa betraktelser hur olämpligt det vore att försöka lära ut universaliserbarhetskriteriet (M) direkt i grundläggande moralisk uppfostran. Här bör man i stället koncentrera sig på att inpränta vissa (tum)regler som man erhåller genom att applicera (M) på normala eller standardiserade viljeattityder. Först när en individ uppnått mognad blir han i stånd att själv hantera kriteriet på ett kompetent sätt och att testa om de regler, som han fått sig till livs, verkligen äger giltighet för honom och hans viljeattityder. Många individer tycks aldrig nå detta stadium utan betraktar livet igenom de inpräglade reglerna som etikens essens och som ägande en av viljeinställningar oavhängig bindande kraft.