

## BJÖRN PETERSSON

### *Hjärnklyvning – ett tveeggat argument*

---

I *Den personliga identiteten och livets mening* (Schildts, Lovisa 1986) ansluter sig Hans Rosing tämligen oreserverat till en tankegång som i olika former varit populär i medvetandefilosofins historia. (Exempelvis hos Thomas av Aquino i *Summa Theologica* II, Q77.) Den går ut på att det mänskliga medvetandet kan ses som ett slagfält där viljor och trosföreställningar tillhörande skilda delar av själslivet strider om att få styra handlandet. Dygden kämpar mot driften, förnuftet mot passionen osv. "Jaget är," säger Rosing med ett mer modernt språkbruk,

ett system som är uppbyggt av en mängd subsystem. Dessa kan i sin tur ha subsystem. Allt som oftast kommer olika system i konflikt med varandra. Det moraliska subsystemet kommer i konflikt med driftssystemet, plikten kommer i konflikt med lättjan osv. (s 170)

På 60-talet började man utföra experiment med personer som opererats mot svår epilepsi genom att förbindelsen mellan höger och vänster hjärnhalva avskurits. Resultaten av dessa experiment låter sig svårligen tolkas på annat sätt än som att den kropp som före operationen var förbunden med ett medvetande efteråt tycks vara säte för två helt skilda mentala nätverk. För Rosing är förekomsten av människor med två separata medvetanden det tyngsta argumentet för subsystemtanken. (kap 15)

Hjärnklyvningspatienterna har också anförts som belägg för flera andra filosofiska tankegångar: Bl a menar Derek Parfit i *Reasons and Persons* (Oxford U P, 1983) att den faktiska förekomsten av hjärnklyvning visar att det är teoretiskt möjligt att två personer skulle kunna vara "fullständigt psykiskt kontinuerliga med mig i mitt

nuvarande tillstånd” (s 255). I sin tur genererar antagandet om möjligheten av förgrenad fullständig psykisk kontinuitet en del radikala motivationspsykologiska hypoteser som haft stor inverkan på den moralfilosofiska debatten efter Parfits bok.

Hjärnklyvning är ett otvivelaktigt medicinskt faktum. Och, som Parfit säger: ”Ett faktum måste vara möjligt. Och det är ett faktum att människor med åtskilda hemisfärer har två separata medvetandeströmmar – två serier av tankar och erfarenheter som var för sig är omedvetna om varandra.” (Parfit 1983 s 247) Så långt är väl slutsatserna av hjärnklyvningen tämligen oomstridliga.

Men under vilka betingelser visar hjärnklyvning att medvetandet inte är enhetligt? Och hur är det möjligt för två personer att vara fullständigt psykiskt kontinuerliga med en person? Hjärnklyvningspatienterna är enligt min mening en större anomali för subsystemteorin än för synen på medvetandet som en konsistent enhet. Och idén om förgrenad fullständig psykisk kontinuitet förutsätter, tror jag, en kontroversiell syn på förhållandet mellan själsliga och materiella förlopp.

### *1. Visar hjärnklyvning att det mänskliga medvetandet normalt består av motstridiga mentala subsystem?*

Donald Davidsons teori om ”mentala avdelningar” utgör ett väl utarbetat försvar för möjligheten att förena subsystemtanken med modern medvetandeteori. (Se t e x ”Paradoxes of Irrationality” i Wollheim/Hopkins, utg, *Philosophical Essays on Freud*, Cambridge 1982 eller ”Deception and Division” i LePore/McLaughlin, utg *Actions and Events, Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford 1985.) Douglas Hofstadter och Daniel Dennett är två andra prominenta företrädare för teorin om en person ”som en förening av många subpersoner, som alla *har egna viljor*” (Hofstadter cit av Rosing s 184).

För det första bör det noteras att subsystemtanken sådan den tolkas av Rosing förmodligen låter sig förenas med en cartesiansk syn på jaget och beslutsfattandet. Det som i en cartesiansk modell skulle göra att alla dessa motstridiga viljor tillhörde samma medvetande vore det faktum att de hade samma ägare, att de var medvetandetillstånd som uppbars av samma cartesianska jag. (I *Om passionerna* tycks Cartesius avvisa subsystemtanken och inte bara kräva

odelbarhet, utan också enhetlighet. Men såvitt jag förstår skulle en extrem cartesian utan logisk inkonsistens faktiskt helt enkelt kunna förneka att jaget delas vid hjärnklyvning. Det som delas är medvetandeströmmen. Dess ägare, som ju enligt cartesianismen är något utöver de mentala tillstånden, kan fortfarande vara en.)

Den cartesianiska homunculusmodell av psyket som nyss beskrivits lider av andra teoretiska brister. Den främsta är kanske problemet att förklara jagets relation till de mentala tillstånden utan att hamna i cirkularitet eller i en infinit regress. Men ett medvetandes brist på enhetlighet är inte oförenligt med cartesianismen.

För icke-cartesianer är det däremot svårare att beskriva uppenbar logisk inkonsistens i ett medvetande. Eftersom vi då inte kan använda ägandeskap som instrument för avgränsning av jaget, som kriterium på *ett* medvetande, använder vi i vanliga fall just konsistens som ett sådant instrument. Ju svårare en handling är att inordna i det mönster av önskningar och föreställningar vi tillskriver en person, desto mer benägna är vi att betrakta handlingen som oavsiktlig, som orsakad av något annat än personens eget beslut. Och när vi undersöker huruvida så kallade multipla personligheter (som Rosing diskuterar utförligt) verkligen är skilda personer i en kropp eller om de är utslag av svårförenliga föreställningar och önskningar hos en och samma person försöker vi just hitta de upplevelser och minnen som gör de ytligt sett motstridiga personligheterna begripliga inom ett och samma logiska mönster, dvs de dolda premisser som upplöser den skenbara motsägelsefullheten.

Detsamma gäller, menar jag, för hjärnklyvningspatienter. Vad är det som gör att vi är benägna att hålla med Rosing och Parfit om att "det är ett faktum att människor med åtskilda hemisfärer har två medvetandeströmmar—två serier av tankar och erfarenheter som var för sig är omedvetna om varandra"? Det medicinska ingreppet i sig bevisar inte detta. Det är först när vi efter experiment och intervjuer kan slå fast att uppenbart motstridiga föreställningar och attityder är bundna till de olika hjärnhalvorna som vi blir övertygade om att en medvetandeström verkligen blivit två. Experimenten är oftast utformade just för att göra det möjligt att se om den ena hjärnhalvan kan bibringas en föreställning medan den andra tror att den är falsk.

I ett sådant experiment (som flera gånger visats i svensk tv) avskärmar man de båda hjärnhalvornas synfält, ger dem olika infor-

mation, och försöker sedan få hjärnklyvningspatienten att berätta vad den hjärnhalva som saknar talcentrum har sett. Patienten svarar fel. Sedan försöker man kommunicera med hjärnhalvan utan talcentrum, varpå den hand som är förbunden med denna hjärnhalva ritar en korrekt bild av det som visats. Patienten berättar då att den handen ofta tar ut ett helt annat plagg än det hon vill ha ur garderoben på morgonen. Först när vi ser att patienten tror både *p* och *icke p*, vill både *p* och *icke p*, blir vi övertygade om att hjärnklyvningen också var en medvetandeklyvning.

Hjärnklyvningsexperimenten ger alltså ytterligare belägg för att konsistens är anticartesianens huvudsakliga kriterium för avgränsning av ett medvetande. (Därmed inte sagt att logisk förenlighet är en tillräcklig förutsättning för medvetandets enhet; bara att ett visst mått av logisk förenlighet tycks vara nödvändigt för avgränsning av ett medvetande.) Därför är förekomsten av hjärnklyvning ett större problem för subsystemtanken än för den som hävdar att medvetandet är en konsistent enhet. I varje fall behöver subsystemtanken kompletteras med upplysningar om vad det är som gör att de olika subsystemen tillhör samma medvetande och inte är lika skilda som efter en hjärnklyvning (eller som hos ett extremfall av multipel personlighet).

Vilka drag i Donald Davidsons syn på medvetandet leder då honom till antagandet att medvetandet är indelat i "avdelningar"? I motsats till Wittgensteininfluerade medvetandefilosofier som exempelvis G H von Wright anser Davidson att det är meningsfullt att tala om kausala förbindelser mellan mentala tillstånd och handlingar. Samtidigt accepterar han att begrepp som 'trosföreställning', 'vilja' och 'avsiktlig handling' är logiskt relaterade till varandra. Att man ser en handling som avsiktlig innebär att man ser den som en begriplig del av ett logiskt mönster, en logiskt nödvändig konsekvens av vissa mentala tillstånd. Då en avsiktlig handling per definition är uttryck för bl a en viljeattityd så blir det (med vissa reservationer) en tautologi att säga att vi alltid gör vad vi vill göra.

Dessutom är Davidson övertygad om att fall av genuin viljesvaghet faktiskt förekommer, dvs att vi faktiskt ibland i beaktande av alla relevanta omständigheter gör saker som vi inte vill göra. Davidsons sätt att förklara viljesvaghet utan att överge idén om en begreppslig koppling mellan avsikter, viljor och trosföreställningar består i teorin om "mentala avdelningar": Avsikter är logiskt

förbundna med attityder i separata logiska nätverk som sinsemellan kan vara motstridiga. Och problemet för Davidson, liksom för Rosing, är, som jag ser det, att förklara vad det är som gör att inkonsistenta och separata logiska nätverk ingår i ett och samma medvetande.

Det är inte heller klart hur beslutsfattande går till i ett "avdelningsindelad" medvetande. Avgör slumpen vilken mental underavdelning som får styra handlandet när avdelningarna är motstridiga? I annat fall, vem är det som väljer och på vilka grunder sker valet? Här anas i värsta fall ett problem analogt med det som drabbar den cartesianska homunculusförklaringen.

För att travestera Davidsons eget argument mot den gamla bilden av mänskligt beslutsfattande som en batalj mellan motstridiga drivkrafter (Davidson, *Essays on Actions and Events*, London 1980, s 35): Antingen kämpar subsystemen mot varandra tills ett av dem vinner; i så fall blir det svårt att klandra agenten för det han gjort. Handlingen är ju bara resultatet av en kamp mellan två krafter inom honom. Vad skulle han kunna göra åt det? Eller också tänker man sig att agenten överblickar båda subsystemen och väljer vilket av dem som ska styra handlandet, men i så fall är vi tillbaka vid utgångspunkten. Vi har inte förklarat hur beslutsfattandet går till.

## *2. Visar hjärnklyvning att två personer kan vara fullständigt psykiskt kontinuerliga med en person?*

Min avsikt här är inte att ta ställning till frågan om personlig identitet är det vi bryr oss om. Men en av förutsättningarna för Parfits nekande svar på denna fråga är att hjärnklyvning öppnar för den teoretiska möjligheten att en person kan vara fullständigt psykiskt kontinuerlig med två personer. Eftersom två personer inte kan vara numeriskt identiska med en person kan numerisk identitet inte vara det vi bryr oss om enligt Parfit. En allmän intuition är nämligen den att vi bryr oss om vad som händer med den som är psykiskt kontinuerlig med oss: Skulle vi få erbjudandet att för en period placera vårt medvetande, t ex genom att byta hjärna, i en kropp vilken under denna period kommer att utsättas för ett smärtsamt ingrepp, så skulle vi förmodligen vara ytterst tveksamma. Så Parfit har förmodligen rätt i att vi bryr oss mer om vår psykiska kontinuitet än om vår personliga identitet. (Om man förutsätter att personlig identitet kräver

mer än psykisk kontinuitet.)

Är det omöjligt att ha mer än en fullständigt psykiskt kontinuerlig efterföljare, dvs om psykisk kontinuitet nödvändigtvis är en 1:1-relation liksom numerisk identitet, så blir emellertid konsekvenserna av detta inte speciellt långtgående. (Parfits moralpsykologiska slutsatser skulle exempelvis vila på en bräckligare grund.) Därför är det värt att undersöka under vilka förutsättningar "förgrenad fullständig psykisk kontinuitet" är möjligt.

För det första bör man lägga märke till att relationen "vara fullständigt psykiskt kontinuerlig med" inte kan tolkas som en symmetrisk relation om fullständig förgrening ska vara möjligt. Förgrenad kontinuitet innebär att A, som från början bara varit kontinuerlig med A, blir kontinuerlig både med B och med C, som inte är kontinuerliga med varandra. Eftersom kontinuitet nödvändigtvis är transitivt, vilket Parfit också poängterar (s 206), måste vi alltså förneka symmetrin för att inte B och C nödvändigtvis ska bli kontinuerliga med varandra. (Relationen R är transitiv om  $aRb$  och  $bRc$  implicerar  $aRc$ , den är symmetrisk om  $aRb$  implicerar  $bRa$ .) Eftersom termer som "vara fullständigt kontinuerlig med" i vanligt språkbruk antyder symmetri vore det kanske lämpligare att använda uttryck i stil med "efterföljas av" som ju beskriver icesymmetriska transitiva relationer.

Men det finns en annan och viktigare betingelse för att förgrenad fullständig psykisk kontinuitet ska vara teoretiskt möjligt. Fundera först över kontinuitetsrelationen i ett empiriskt möjligt fall av hjärnklyvning.

a) "Mitt exempel", säger Parfit, "förutsätter antagandet att båda mina hemisfärer har en komplett uppsättning färdigheter, vilket kan vara sant om en del existerande människor." Hans hemisfärer är "exakt likadana." (s 246) Hos vissa människor tycks det vara så att båda hjärnhalvorna är kapabla att utföra samma uppgifter. Detta innebär sannolikt inte att de faktiskt utför exakt samma uppgifter simultant. Om en sådan existerande person utsattes för hjärnklyvning och sedan fick sina hjärnhalvor transplanterade till två nya kroppar skulle förmodligen ingen av de resulterande personerna vara fullständigt psykiskt lik originalet. Deras minnen skulle förmodligen skilja sig åt vad beträffar innehåll eller tydlighet etc. De skulle förmodligen tänka och reagera lite olika, eftersom deras hjärnor

faktiskt hade utfört olika uppgifter före separationen. Ingen av dessa båda efterföljare är fullständigt psykiskt kontinuerliga med originalet. De psykiska processer som pågick i vänsterhalvan kommer att fortsätta vara förknippade med denna hjärnhalva och vice versa. Liksom hjärnan delats i två halvor har psyket delats i två halvor. Men för att hjärnklyvning ska ställa till problem för tanken att psykisk kontinuitet, liksom numerisk identitet, är en 1:1-relation krävs något mer: Samtliga psykiska processer i den odelade hjärnan vid hjärnklyvningstillfället måste fortsätta i var och en av de separata hjärnhalvorna. I varje hjärnhalva ska hela medvetandeprocessen fortsättas. Modifiera därför exemplet en aning (på bekostnad av empirisk trolighet):

b) Anta att de båda hjärnhalvorna före separationen inte bara har förmågan att utföra samma funktioner, utan också faktiskt hela tiden utför samma uppgifter simultant; lagrar likadana minnesbilder från samma händelser etc. Om hjärnan är Parfits och den delas för att placeras i två tvillingbröders kroppar, så skulle Parfit förmodligen ha rätt i att var och en av "de resulterande personerna tror att han är jag, har mina egenskaper och tycker sig minnas mitt liv", men skulle Parfit också sanningsenligt om sin efterföljare kunna hävda att han "på varje annat sätt är psykiskt kontinuerlig med mig" (s 255)?

Om psykisk kontinuitet bara bestod i likhet vore det förstås sant: Innehållet i de båda efterföljarnas minnen och föreställningar omedelbart efter klyvningen är ju likadant som innehållet i Parfits minnen och föreställningar omedelbart före klyvningen.

Men numeriskt sett skiljer sig inte detta exempel från det föregående. De mentala processer som före klyvningen var förknippade med vänstra hjärnhalvan fortsätter kontinuerligt i denna hjärnhalva efter klyvningen, och vice versa. Det är alltså inte frågan om förgrenad fullständig psykisk kontinuitet, utan om förgrenad halverad psykisk kontinuitet.

Parfits kriterium på psykisk kontinuitet är att det finns "överlappande serier av starka förbindelser" mellan psyken vid olika tillfällen. Han stipulerar vidare att "starka förbindelser" består i att minst halva antalet minnesförbindelser består från den ena dagen till den andra, jämfört med antalet minnesförbindelser "i verkliga normala liv". Har han då inte helt enkelt stipulerat att det räcker med att halva den mentala processen fortsätter för att vi ska tala om

fullständig psykisk kontinuitet? Av tre skäl tror jag inte att denna stipulation, även om den vore rimlig, skulle göra att exemplen ovan beskriver "förgrenad fullständig psykisk kontinuitet" i Parfits mening:

1. Halveringen i kriteriet tillåts p g a att antalet förbindelser från en dag till föregående dag normalt sett är betydligt mindre än antalet förbindelser från ett ögonblick till föregående ögonblick. Kriteriet på psykisk kontinuitet bör inte få den konsekvensen att vi inte normalt sett är psykiskt kontinuerliga från dag till dag. I det beskrivna hjärnklyvningsfallet halveras dock antalet mentala förbindelser bortsett från det normala bortfallet. Om vi antar att det normalt tillåtna bortfallet om 50% per dygn faktiskt också äger rum i varje hjärnhalva så kommer antalet förbindelser i varje separerad halva bara att vara 25% av förbindelserna i hjärnan dagen före separationen.
2. Om "förgrenad fullständig psykisk kontinuitet" bara innebär att den psykiska processen i ursprungshjärnan delas i två halvor om 50% skulle det inte bevisa att fullständig psykisk kontinuitet inte nödvändigtvis är en 1:1-relation.
3. Menade Parfit inte mer än detta skulle det också vara svårt att förstå varför han i sitt exempel på förgrenad fullständig psykisk kontinuitet kräver att de båda hjärnhalvorna är "exakt likadana": 50% av den psykiska processen fortsätter ju i varje hjärnhalva även om de skulle vara totalt olika.

Förgrenad fullständig psykisk kontinuitet, i den bemärkelse som förutsätts av Parfits argument om den personliga identitetens betydelse, kräver att hela den mentala processen hos originalpersonen omedelbart före klyvningen är fullständigt kontinuerlig med den mentala processen hos var och en av de båda efterföljarna. Detta måste innebära att den mentala processen i ursprungspersonen på något sätt är överdeterminerad, dvs att båda hjärnhalvorna var för sig simultant orsakar hela den mentala processen. (Ungefär som när två styranordningar av säkerhetsskäl används simultant för att styra ett flygplan.) Efter hjärnklyvningen fortsätter varje hjärnhalva var för sig att upprätthålla hela processen.

En teori som antar strikt identitet mellan psykiska och neurofysiologiska fenomen skulle därför implicera att förgrenad fullständig psykisk kontinuitet är omöjligt. Men jag misstänker att



även svagare teoretiska antaganden om kopplingen mellan fysiskt och psykiskt skulle vara svåra att förena med tanken på den sorts överdeterminering som krävs för fullt förgrenad psykisk kontinuitet. Anta exempelvis att Bernard Williams har rätt i att det är omöjligt att individuera mentala tillstånd utan referens till något materiellt. ("Personal Identity and Individuation", 1972) Dvs det är omöjligt att skilja kvalitativt identiska mentala egenskaper från varandra utan att se dem som egenskaper hos något som har rumslig utsträckning. Skulle det då vara möjligt att ett och samma (i numerisk mening) mentala tillstånd helt och hållet är en egenskap hos vänster hjärnhalva och samtidigt helt och hållet en egenskap i höger hjärnhalva?

### 3. *Slutsatser*

Hjärnklyvning visar att medvetandet inte är odelbart. Ett medvetande kan bli två medvetanden. Detta bevisar inte att ett odelat medvetande normalt sett är inkonsistent, dvs att det består av flera små löst sammankopplade medvetanden, subsystem, med sinsemellan oförenliga attityder och föreställningar. Tvärtom; den som är tveksam till idén om ett cartesianskt jag behöver antagandet om medvetandets enhetlighet för att kunna avgränsa ett medvetande. Hjärnklyvningsexperimenten demonstrerar detta kriteriums betydelse.

Hjärnklyvning ger starkt empiriskt stöd för teorin att det fysiska och psykiska är intimt sammankopplat. Om man dessutom antar att förbindelsen mellan mentalt och materiellt inte bara är kausal och kontingent, utan nödvändig av begreppsliga skäl, så tycks den sorts överdeterminering som krävs för att möjliggöra förgrenad fullständig psykisk kontinuitet bli svärförklarlig. Möjligheten att förklara det mentalas superveniens på det materiella så att relationen tillåter sådan överdeterminering kan inte uteslutas. Men såvitt jag förstår bör väl bevisbördan främst vila på den som vill bygga vidare på antagandet att en person kan ha två efterföljare och vara fullständigt psykiskt kontinuerlig med var och en av dem?