

Recension

Kathleen Wilkes: *Real People: Philosophy of Mind without Thought Experiments* (Oxford 1988).

Vad är ett filosofiskt tankeexperiment? Intresset för denna fråga är mindre utbrett bland filosoferna än den ständigt (och ofta okritiskt) brukade tankeexperimentella metoden själv. Ett undantag är emellertid Kathleen Wilkes i Oxford. I *Real People: Personal Identity without Thought Experiments* går hon i ett långt inledande kapitel till hårt angrepp mot vad hon betecknar som ett missbruk av tankeexperiment—inom filosofi i allmänhet och medvetandefilosofi i synnerhet.

Själv baserar Wilkes sin diskussion av personlig identitet på en imponerande bred exempelflora, hämtad ur den medicinska, psykiatriska och filosofiska litteraturen. Ibland hemfaller hon åt populärvetenskapliga miniessäer om kommissurotomi, hemisfärektomi, multipel personlighet, hypnos m fl fenomen. Det är kanske inte alltid så lätt att spåra den filosofiska relevansen hos dessa utläggningar—men de gör onekligen *Real People* till spännande läsning. Wilkes bok är ovanligt rik på stimulerande (och provocerande) teser. Det är omöjligt att inom ramen för en kort anmälan diskutera eller ens flyktigt beröra de flesta av dem. Jag skall därför begränsa mig till vad jag finner mest spännande, nämligen Wilkes kritik av tankeexperiment, samt till en ytterst summarisk kritik av hennes behandling av personlig identitet.

Vad Wilkes anklagar sina tankeexperimentörer kolleger för är helt enkelt slarv i den experimentella praktiken. Ett experiment syftar till att systematiskt pröva en tes eller teori genom att undersöka dess konsekvenser. Dessa konsekvenser följer som bekant inte ur teorin ensam; men poängen med de artificiella, noggrant utformade experimentsituationerna är just möjligheten att kontrollera och fixera den relevanta bakgrunden. Vi kan pröva en hypotes genom att variera den faktor den uttalar sig om—men bara genom att hålla andra relevanta faktorer konstanta. Detta gäller, menar nu Wilkes, även tankeexperiment. Ett tankeexperimentets mål är visserligen inte att pröva en teori empiriskt, utan i stället att undersöka dess konsekvenser i kontrafaktiska situationer, och tankeexperimentatorn har i motsats till kollegan i laboratoriet inte observationer utan intuitioner som måttstock på konsekvensernas hållbarhet. Men det befriar inte tankeexperimentet från kravet på att kontrollera och fixera den relevanta bakgrunden. Detta förutsätter emellertid att man har en välgrundad uppfattning om vilken bakgrunden är. Och det är här de filosofiska tankeexperimenten brister, enligt Wilkes. Alltför ofta underdeterminerar de förödande stora portioner av den kontrafaktiska situation som tesen prövas i, med påföljd att resultatet—dvs de intuitioner experimentet genererar—är intetsägande, på samma sätt som ett laboratorieexperiment där det visar sig att provrören varit smutsiga. Om vi inte vet *vilken* möjlig värld vi är på utflykt till har vi inte tillräckligt underlag för att säga vad som skulle vara fallet i den (t ex huruvida ett begrepp skulle vara applicerbart, en handling riktig, två individer identiska etc). Men då gör vi kanske klockast i att hålla oss till den faktiska världen—vilket också är vad Wilkes ivrigt uppmanar filosoferna att göra.

Wilkes är inte skeptisk mot alla tankeexperiment. De som används inom t ex fysik undgår ofta de problem som drabbar filosofiska tankeexperiment. Skälet är enligt Wilkes att de hypoteser som de prövar använder termer som betecknar *naturliga slag* ("natural kinds"), exempelvis "vatten", "ljus" och "elektron". Frågan om vad sådana termer egentligen har för status är ju omdiskuterad—men utmärkande för naturliga

slag tycks åtminstone vara att de är fruktbare både som vetenskapliga explananda och explanantia. I motsats till termer som "askfat" och "stängsel" är de djupt inflytade i vårt nätverk av teorier om världen; naturliga slag är vad lagar och generaliseringar uttalar sig om. Det gör att de logiska relationerna mellan satsor som handlar om naturliga slag är förhållandevis klara. Och detta gör i sin tur att det är relativt klart vad som är relevant i tankeexperiment som involverar naturliga slag. I synnerhet är det lätt att avgöra om de *omöjliga* antaganden som nästan alltid figurerar i tankeexperimentella scenarion är relevanta för hypotesen under prövning. Är de det måste experimentet avvisas, enligt Wilkes; men är omöjligheten irrelevant är experimentet legitimt. De tuktade logiska relationerna mellan termer för naturliga slag gör det också lätt att avgöra vad som skulle följa i den kontrafaktiska situation tankeexperimentet beskriver: utfallet följer mer eller mindre deduktivt. Fysiker är därför sällan oense om resultatet av tankeexperiment inom fysiken.

Skälet till att filosofiska tankeexperiment enligt Wilkes i regel inte är legitima är att de sysselsätter sig med begrepp som "person", "medvetande" och "mind"—termer som liknar "askfat" såtillvida som de inte betecknar naturliga slag. Eftersom sådana termer saknar klara lagmässiga förbindelser, blir det osäkert vad som är relevant för deras applicerbarhet i kontrafaktiska situationer. Därför är det också svårt att utesluta möjligheten att tankeexperimentet förutsätter sådant som är inte bara är omöjligt, utan dessutom relevant för den filosofiska hypotes som skulle prövas—vilket är förödande, hävdar Wilkes, eftersom det innebär att vi inte har något att grunda våra eventuella intuitioner i. En karakteristisk subjektivitet och oenighet vidlåder också intuitionerna inför filosofiska tankeexperiment.

Filosofier är ofta angelägna om att förläna sina tankeexperiment en viss minimal plausibilitet. Få är villiga att nöja sig med enbart *logisk* möjlighet; tankeexperimenten gör dessutom anspråk på *teoretisk* möjlighet (möjlighet "i princip"). Rågängen mot den rena fiktionen måste säkras. Wilkes utmaning består i att hon omklassificerar vad som trots och påstås vara teoretiskt möjligt och avfärdar det som på sin höjd logiskt möjligt. Åtskilliga filosofiska tankeexperiment hamnar därvid i samma kategori som *Alice i Underlandet*: de kan ses som roande, men filosofiskt irrelevanta, fantasier.

Antag exempelvis (Wilkes illustration) att järnstänger kunde flyta på vatten. Vi tror gärna att det finns en möjlig värld där de gör det—det tycks vara ett nog så harmlöst antagande. Men järn har en densitet mellan 7,3 och 7,8 och vatten har densiteten 1—och för att x skall flyta på y måste x ha en lägre densitet än y . Så antagandet tycks förutsätta att $7,3 < 1$. Men *det* är knappast möjligt. Alternativt förutsätter antagandet att järn har en densitet < 1 —men är det verkligen *järn* vi talar om då? Motsvarande svårigheter uppstår med de övriga sätt på vilka antagandet kunde försöka göras koherent. Försöken att fylla i skissen, att beskriva *hur* järn kunde flyta på vatten, undergräver vår ursprungliga tillit till att det finns en genuin teoretisk möjlighet här. Och på samma vis borde det vi vet om neurofysiologi undergräva tilliten till de flesta medvetandefilosofiska tankeexperiment. De representerar, menar Wilkes, inte ens teoretiska möjligheter. Det faktum att det är mycket vi inte vet om hjärnan och om medvetandet förvärrar endast saken, eftersom vi inte kan ta okunnighet till intäkt för spekulation:

To insist, 'But if this, this and this, then that, that and that' may be fun but is not helpful: the laws governing human development and function may *!...! forbid* 'this, this and this' from obtaining under any circumstances. Not knowing whether the 'this' could hold, we have no justification in concluding any 'that'; thought experiments that are the product of ignorance simply mislead. (s 39).

Epistemisk möjlighet utesluter alltså för Wilkes inte teoretisk omöjlighet.

Wilkes argumentation är uppfriskande, men inte helt övertygande. Framför allt är det oklart varför den omöjlighet som hon så skickligt blottlägger i många tankeexperiment också är relevant för dem. Relevans är, som Wilkes framhåller, relativ till ett bestämt syfte. Syftet hos många filosofiska tankeexperiment är att kartlägga ett begrepps gränser genom att se hur det uppstår sig i kontrafaktiska situationer. Derek Parfitts tankeexperiment i *Reasons and Persons*—för att ta ett av de medvetandefilosofiska verk som drabbas av Wilkes kritik—ledde honom bl a till resultatet att personlig identitet inte är en fråga om fysisk kontinuitet. Nu har Wilkes säkert rätt i att Parfitts tankeexperiment är omöjliga i en starkare betydelse än han antog: våra teorier om olika psykologiska funktioner och om hjärnan indikerar att personlig identitet är en relation som faktiskt inte kan hålla utan ett betydligt större mått av fysisk kontinuitet än den Parfitts teletransportpassagerare åtnjuter. Teletransport är förmodligen teoretiskt (och inte bara "praktiskt") omöjligt. Men detta ogiltigförklarar inte automatiskt Parfitts resultat. Att personlig identitet *faktiskt* förutsätter ordinär, fysisk kontinuitet är *förenligt* med det filosofiskt intressanta resultatet att teletransport, om det vore möjligt, skulle kunna utnyttjas för att skicka personer, men inte kroppar, till Mars—vilket säger något om personlig identitet. Att insistera att teletransport *är* omöjligt är att missuppfatta undersökningens syfte.

Låt oss kalla beskrivningen av den kontrafaktiska situationen i ett tankeexperiment för dess antecedent. Antag att en tankeexperimentell antecedent P faktiskt inte skulle kunna vara sann utan att Q också var sann, men Q nämns inte i presentationen av tankeexperimentet och kan inte antas vara underförstådd. Vi kan nu tänka oss två epistemiska situationer, A och B. I A vet tankeexperimentatorn och hennes läsare att P är beroende av Q; i B är detta helt okänt. Antag vidare att Q är teoretiskt omöjligt, på samma sätt som flytande järnstänger och teletransport (förmodligen) är omöjliga. Wilkes menar nu att Q är relevant för tankeexperimentet i såväl A som B och att dess omöjlighet därför är förödande. Men jag tror att hon har fel: jag tror att Q kan vara tankeexperimentellt irrelevant både i A och B. Anledningen är helt enkelt att tankeexperimentets syfte är mycket mer oberoende av P:s modala status än Wilkes förutsätter. Nomologiskt beroende medför inte, som Wilkes genomgående antar, automatiskt tankeexperimentell relevans. Detta illustreras i fall A av Parfitt, men också av flera av de tankeexperiment inom fysik som Wilkes är beredd att acceptera. Einsteins vederläggning av Maxwells elektrodynamiska teori skedde genom ett tankeexperiment där en observatör färdas med ljusets hastighet, ett antagande som naturligtvis är starkt omöjligt men inte relevant (eftersom experimentets resultat är oberoende av människors och farkosters förmågor). Fall B kan illustreras av en enkel kontrafaktisk villkorssats—ett tankeexperiment i miniatyr. Låt P = "Tändstickan t stryks mot plånet" och Q vara något nödvändigt men okänt villkor för P, samtidigt som Q faktiskt är teoretiskt omöjligt (t ex genom att världen är deterministisk och Q strider mot naturlagarna). Nu är P förvisso också omöjligt, men vi kan med viss tillförsikt uttala oss om vad som skulle vara fallet om P var sann: tändstickan skulle tändas.

Vi kan—förefaller det mig—känna till sanningsvärdet hos en kontrafaktisk villkorssats utan att kunna specificera alla satsen som inte är kotenabla med antecedenten; och vi kan, mer generellt, ofta bedöma utfallet av ett tankeexperiment utan att kunna specificera eller känna till *allt* som måste gälla för att antecedenten skulle kunna vara sann. Wilkes krav på tankeexperimentens bakgrund är helt enkelt alltför strängt.

Hon uppvisar dessutom en omotiverad bias. Många av de vetenskapliga tankeexperiment hon godtar skulle förkastas om de bedömdes på samma grunder som de filosofiska. Simon Stevin presenterade på 1600-talet ett berömt fysiskt tankeexperiment som involverar friktionsfria ytor. Wilkes säger att "The world as we know it does not allow for perfectly frictionless planes; but they would be a limiting case of the frictionlessness that we can achieve" (s 31) och menar att detta gör perfekt friktionslöshet till en teoretisk möjlighet. Men för det första kan en lång rad av de tankeexperimentella antaganden hon förkastar inom medvetandefilosofi konstrueras analogt, som extrapoleringar av kända fenomen. För det andra är försvaret inadekvat. Ren friktionsfrihet är lika svår att lokalisera till en *bestämd* möjlig värld som teletransport; att ackommodera den förutsätter oerhört djupgående transformationer av olika fysiska lagar. Men dess teoretiska omöjlighet är inte relevant, den upphäver inte tankeexperimentets nytta, trots att det handlar just om fysik. På samma sätt är den teoretiska omöjlighet som leder Wilkes till att förkasta en rad filosofiska tankeexperiment ofta irrelevant.

Wilkes bortser också från den normativa funktion tankeexperiment fyller. Det är sant att intuitionerna ofta går isär inför filosofiska tankeexperiment; men experimenten själva bidrar till att stabilisera och systematisera intuitionerna genom att tvinga den som kontemplerar dem att kritiskt ompröva sina uppfattningar. Våra intuitioner dikteras förvisso i stor utsträckning av vilka övertygelser vi redan har—men detsamma gäller observationer. Det ömsesidiga beroendet mellan experiment och teori är inte unikt för just tankeexperiment. Intuitionernas teoriavhängighet är därför inte så suspekt som Wilkes vill antyda.

Vidare bottnar tesen att de begrepp som figurerar i tankeexperiment saknar klara logiska relationer till andra begrepp—att de i motsats till naturliga slag är teoretiskt ofruktbara—i en underskattning av filosofins förmåga att *förläna* en teoretisk status åt begrepp som "person", "medvetande" etc. Mycket medvetandefilosofi kan ses som försök att upprätta en teoretisk, dvs logisk struktur åt dessa begrepp. Så om ett tankeexperimentets legitimitet beror på hur mycket teori som kringgärdar och regimenterar dess centrala termer, är det inte klart att filosofiska tankeexperiment är värre därän än naturvetenskapliga. När Wilkes hävdar att "person" nödvändigtvis är en ateoretisk och vag vardagsspråksterm, signalerar hon att hon inte erkänner *filosofiska* teorier om personbegreppet som relevanta. Men det är ad hoc. Att "person" obestriddligen är en term i vardagsspråket, precis som "ljus" och "vatten", hindrar inte att det också kan ges en teoretisk användning som delvis överlappar med och delvis avviker från den vardagliga. Inte heller kan filosofiska teorier diskvalificeras med hänvisning till att de är underutvecklade, att de har rivaler, eller att de är illa integrerade med naturvetenskapliga teorier. De två sista förhållandena gäller i ännu högre grad de sociologiska och psykologiska teorier som Wilkes är beredd att erkänna som konstitutiva för naturliga slag. Och vad underutvecklingen beträffar så var *fysiken* underutvecklad när flera av dess viktigaste tankeexperiment presenterades (och därigenom bidrog till dess utveckling!).

Wilkes snåla inställning till filosofin som normativ aktivitet återspeglas i de ståndpunkter hon intar senare i boken. De två frågor hon förelägger sig att besvara utan att tillgripa tankeexperiment rör identifikationen av personer: vad gör något till en person? och reidentifikationen: när är A och B genidentiska? Båda frågorna är etiskt relevanta. Wilkes svar är att något är en person när, och i den mån, det uppvisar egenskaper som är karakteristiska för vuxna, normala medlemmar av homo sapiens. Flera av dessa egenskaper är—liksom vad vi uppfattar som normalt—subjektiva: de är beroende av vår *attityd* till människor, som onekligen är en annan än

vår attityd till icke-människor. Följaktligen är personbegreppet vagt och subjektivt, och frågan huruvida exempelvis foster, psykopater, senila, och patienter med multipel personlighet skall räknas som (enhetliga och fullständiga) personer kommer ofta att sakna ett entydigt svar. Det är också med nödvändighet artistiskt: delfiner och datorer, för att ta två presumtiva kandidater till personstatus, göra sig icke besvär eftersom de är alltför främmande—*artfrämmande*—vid en jämförelse med paradigmala personer.

Wilkes ser dock inget problem i vare sig subjektiviteten, artismen eller vagheten, eftersom hon menar att vi intar en adekvat moralisk attityd mot gränfallen och icke-människorna oberoende av deras eventuella personstatus. Wilkes spårar denna attityd till tre faktorer. Den första är *potentialitet*: foster är potentiella vuxna och normala människor. Par av obefrukade ägg och spermier är, liksom provrörsembryon, också potentiella människor, men denna potentialitet är i motsats till vanliga emryons en *tvåstegspotentialitet*—befruktningen måste äga rum, resp provrörsembryot inplanteras, innan vi har något, en enhetlig referent, som under normala omständigheter blir en människa. Denna distinktion motiverar, hävdar Wilkes, också en moralisk distinktion. Den andra faktorn är *intressen*, vilket djur och foster har i lika hög grad som typiska personer. Den tredje faktorn är vad Wilkes kallar en "*aristotelisk princip*": "any member of kind K which is not a perfect instance of kind K is in some respect something to be pitied or deplored" (s 62). Vissa medlemmar av homo sapiens är varken potentiella normala människor eller medvetna om mer raffinerade intressen än de djur har; det gäller t ex svårt skadade foster och senila. Men enligt Wilkes motiverar det faktum att de lidit en "aristotelisk förlust" (i enlighet med principen ovan) att vi tillerkänner dem en *särskild* moralisk relevans—att vi har en prima facie förpliktelse inte endast att tillgodose deras faktiska, nuvarande intressen, utan dessutom att *kompensera* denna avvikelse från vad som är normalt för arten. (Wilkes aristoteliska princip skiljer sig alltså från sin namne i Rawls *A Theory of Justice* genom att vara en *normativ* princip; hos Rawls är principen en empirisk, psykologisk hypotes.)

Men en allvarlig fråga är naturligtvis *varför* vi skall tillmäta dessa faktorer en sådan betydelse. Wilkes argument för betydelsen av att vara en potentiell person undermineras i icke obetydlig mån av hennes slutsats att det ofta är oväsentligt, ur etisk synpunkt, huruvida något är en person. Att intressen och den aristoteliska principen är relevanta tycks för Wilkes följa ur att vi faktiskt räknar med dem i vår värderande praxis. Men att vi *har* en moralisk attityd är inte en tillfredsställande grund för att vi *bör* ha den.

På motsvarande sätt förutsätter Wilkes alltför lätt att en person med nödvändighet är vad vi spontant och i vardagslag faktiskt är beredda att kalla en person. I vårt vardagliga tal om personer utgår vi från att det går exakt en person per mänsklig kropp. Vi förlitar oss okritiskt på ett (grovt) fysiskt kriterium; härav vår perplexitet inför bl a multipel personlighet; och härav begreppets vaghet, enligt Wilkes. Denna diagnos av vagheten och dess grunder är säkert korrekt—men Wilkes förefaller mig alltför rask att förklara den för obotlig.

Slutligen förutsätter Wilkes diskussion även okritiskt ett bestämt svar på reidentifikationsfrågan. Vilka långsiktiga intressen någon har beror på vem (om någon) hon är genidentisk med i framtiden; men Wilkes böjer sig även här för språkbruket och utgår från att vi har att göra med samma person så länge vi har en och samma kropp, t o m i fall av grav senilitet eller kommissurotom (även om hon medger att språkbruket kollapsar inför multipel personlighet).

Filosofernas legitima uppgift är att förändra vårt faktiska tänkande och språkbruk genom kritiska argument; Wilkes nöjer sig genomgående med att beskriva det. Trots denna svaghet är hennes bok mycket läsvärd för den som intresserar sig för medvetandefilosofi och filosofins metodologi.

Sören Häggqvist

Notiser

I svensk översättning (med stöd från Erik och Gurli Hultengrens fond för filosofi) har nyligen utgivits Jacques Derrida, *Husserl och geometrins ursprung*, Thales 1991. I samma volym ingår också den skrift av Husserl som Derrida utgår från, alltså "Geometrins ursprung" från 1936. Översättningen är gjord av Hans Ruin, som är doktorand i teoretisk filosofi i Stockholm.

En paradox som diskuterats en hel del av filosofer är den som brukar kallas "The Unexpected Examination" (eller "The Unexpected Hanging" eller "The Hangman's Paradox"). Den mycket beläste Jan Österberg har hittat uppgifter som tyder på att denna paradox faktiskt härstammar från Sverige. Det tycks vara så att Radiotjänst någon gång 1943 eller 1944 förmedlade en uppgift om att en civilförsvarsövning skulle äga rum någon gång under den kommande veckan, men att den exakta dagen skulle hållas hemlig, så att deltagarna skulle vara oförberedda. Det paradoxala ligger förstås i att övningen då aldrig kan inträffa. Om lördagen är sista dagen i veckan, så kan övningen inte inträffa på lördagen, ty då skulle ju alla veta det dagen innan (eftersom då endast en dag återstår av veckan). Men då kan övningen inte heller inträffa på fredagen, ty då skulle man ju veta det på torsdagen. Och av samma skäl kan den inte inträffa på torsdagen, och så vidare. Det tycks ha varit så, att den förste som insåg det paradoxala i Radiotjänst budskap var en matematiklärare på Östermalm vid namn Lennart Ekbohm.

Om lärjungar.

Året var 1976. Milton Friedman hade fått nobelpriset i ekonomi. Jag deltog som ung lärare för första gången vid fakultetens sammanträde, som ägde rum i det traditionstyngda Spökslottet i Stockholm. Fakulteten var på den tiden fortfarande ett kollegium, och den bestod av en illuster samling herrar och någon dam. Med mig till sammanträdet hade jag ett förslag till protestuttalande mot att Friedman tilldelats nobelpriset. I min skrivelse framhöll jag särskilt det omoraliska i hans och hans lärjungars stöd till juntan i Chile.

Förste talare var, sedan min skrivelse delats ut och föredragits, professor Sven Ulric Palme från historiska institutionen. Denne höll ett kort men koleriskt anförande, som avslutades med den bitande kommentaren att han, för sin del, aldrig skulle kunna tänka sig att underteckna en skrivelse, där ordet lärjunge stavats med "lj".

Jag stirrade i min skrivelse och blygdes. Ordet var stavat med "lj", precis som i ordet "ljung" och inte alls som i tyskans "Junge".