Filosofisk tidskrift

Årgång 38 Nr 3 September 2017

Filosofisk tidskrift

Utges av Stiftelsen Filosofisk tidskrift och Stiftelsen Bokförlaget Thales

Redaktör och ansvarig utgivare: Lars Bergström

Filosofisk tidskrift utkommer med fyra nummer per år

Prenumeration per år inom landet: 200 kr (inkl. moms); utom landet: 300 kr
Lösnr: 55 kr (inkl. moms)
För prenumerationsärenden kontakta:
Nätverkstan ekonomitjänst, Box 31120, 400 32 Göteborg
Telefon 031-743 99 05 Fax 031-743 99 06
ekonomitjanst@natverkstan.net

Förlagsadress: Bokförlaget Thales, Box 50034, 104 05 Stockholm
info@bokforlagetthales.se www.bokforlagetthales.se

Grafisk form och produktion: Ulf Jacobsen
Satt med Arnhem Pro Blond
Tryck: Carlshamn Tryck & Media AB, Karlshamn 2017
ISSN 0348-7482

ETT TEMANUMMER OM FILOSOFI I GÖTEBORG

Den filosofiska forskningen vid institutionen för filosofi, lingvistik och vetenskapsteori vid Göteborgs universitet spänner över många områden och inkluderar såväl praktisk som teoretisk filosofi, logik, och filosofihistoria. Särskilda satsningar har gjorts på forskning om moraliskt ansvar samt kognition och perception i den aristoteliska traditionen. Föreliggande temanummer ger ett smakprov på olika forskningsinriktningar vid institutionen i Göteborg och representerade av forskare på olika akademiska nivåer, från doktorander till professorer.

Flera av medarbetarna i filosofiämnena vid institutionen har tvärvetenskapliga samarbeten med andra forskningsområden såsom lingvistik, vetenskapsteori, medicin, statsvetenskap, juridik och psykologi. Men även mera strikt inomfilosofisk forskning inom t.ex. metafysik och metaetik finns också starkt representerad i Göteborg.

Vid sidan om de ordinarie områdena praktisk och teoretisk filosofi respektive logik finns två stora forskningsprojekt förlagda vid institutionen. Ett tvärvetenskapligt program med titeln ”Representation och verklighet: Historiska och nutida perspektiv på den aristoteliska traditionen” som förenar filosofihistoria med klassisk filologi och fokuserar på medvetandefilosofiska frågor i den aristoteliska traditionen förvaltas här. En av forskarna inom programmet – Jakob Leth Fink – finns representerad i detta nummer av *Filosofisk tidskrift*. Leth Fink bedriver forskning som fokuserar på Platon och Aristoteles och presenterar och diskuterar här ett fenomen som beskrivs av Aristoteles i *Den nikomachiska etiken* som han kallar ”moralisk blindhet” (Aristoteles om moralisk blindhet).

Ett annat stort forskningsprojekt vid Göteborgs universitet handlar om moraliskt ansvar. Projektet behandlar frågor relaterade till moraliskt ansvar såsom kollektivt ansvar, fri vilja, berättiganden, ursäkter och implikationer för t.ex. straffrätt och global politik. I detta nummer finns även en forskare från detta projekt representerad – Sofia Jeppsson – postdoktor vid Gothenburg Responsibility Project. I sin artikel diskuterar och kritiserar Jeppsson idén om att brottslingar ska straffas för sina gärningar för att de förtjänar det (Förtjänst och straff).

Den praktiska filosofin vid institutionen representeras här vidare av David Brax, postdoktor vid Centrum för Europaforskning vars forskning fokuserar på hatbrott och den straffrättsliga regleringen runt dessa brott, samt Dorna Behdadi, doktorand i praktisk filosofi som skriver en avhandling om moraliskt agentskap hos icke-mänskliga varelser. Brax resonerar i det här numret av *Filosofisk tidskrift* kring hur begreppet hatbrott egentligen ska definieras (Vad gör hatbrott värre än andra brott?) och Behdadi företar en kritisk diskussion av Peter Carruthers medvetandefilosofiska teori med avseende på dess implikationer för djurs förmåga till fenomenellt medvetna upplevelser.

Logikforskningen vid institutionen har en bred profil med bl.a. kopplingar till filosofi, lingvistik och matematik. Paul Gorbow som är doktorand i logik skriver här om Stephen Yablos formella system för att modellera essens och kontingent identitet (Essens, att sammanfalla och karaktärisering).

Inom teoretisk filosofi bedrivs forskning framförallt inom områdena metafysik och medvetandefilosofi. I föreliggande nummer skriver Alla Choifer som är doktorand i teoretisk filosofi om barns förmåga att förstå andra människor (Att förstå sig själv och att förstå den andre – ett filosofiskt perspektiv på utvecklingspsykologi) och Anna-Sofia Maurin, professor i teoretisk filosofi diskuterar hur lika (eller olika) metafysik och vetenskap är eller bör vara.

Vi hoppas att numret ger upphov till stimulerande läsning och tankar och att det ger en god bild av några av de forskningstraditioner som finns representerade i Göteborg.

ANNA-SOFIA MAURIN OCH SUSANNA RADOVIC

DORNA BEHDADI

”THEORY OF MIND” OCH FENOMENELLT MEDVETANDE

En kritisk granskning av Peter Carruthers dispositionella HOT-teori

1. INTRODUKTION

Ur etiskt hänseende är innehav av fenomenellt medvetande av största vikt. Förmåga till subjektiva upplevelser utgör ofta själva grundkriteriet för att en varelse ska tillskrivas någon slags moralisk status. För medan en entitet som helt saknar upplevelseförmåga också brukas anses sakna moralisk status, medför ett ”inre ljus” att en annan varelse kvalificerar sig för eventuell hänsyn (för en genomgång av teorier om moralisk status se Warren 1997).

I flera diskussioner och utredningar om icke-mänskliga djurs moraliska status brukar kontentan dessutom vara att det *räcker* med basal upplevelseförmåga, för att inneha åtminstone någon grad av moralisk ställning. Det är tillräckligt att musen kan lida för att det ska gå att invända mot att hen får utsättas för vad som helst på labbet. På samma sätt krävs inte någon extraordinär kognitiv förmåga hos ett ungt barn för att hen ska anses vara i behov av bedövning vid kirurgiska ingrepp. Idén uttrycks i Jeremy Benthams citat: ”Frågan är inte Kan de resonera? Och inte heller Kan de tala? Utan i stället Kan de lida?” (Bentham 1789; 1970 fotnot kapitel sju, min översättning). Denna ”sentientistiska” (från engelskans ”sentience”: förnimbarhet eller upplevelseförmåga) premiss rörande moralisk status kan även ses reflekterad i europeisk djurskyddslagstiftning: ”Unionen och medlemsstaterna ska, eftersom djur har upplevelseförmåga, ta största hänsyn till deras välfärdsbehov…” (TFEU 2009 II, artikel 13 (min översättning)).

I ljuset av denna relation mellan upplevelseförmåga och moralisk ställning, är det uppenbart att teser som rör huruvida en grupp kan tillskrivas medvetna mentala tillstånd, måste vara väl underbyggda. Inte enbart för att rena intuitioner och grundlösa antaganden riskerar att leda till felaktig kunskap, utan för att den etiska insatsen helt enkelt är så stor. Medan sentientistiska etiska teorier, som Peter Singers utilitarism (1975) eller Tom Regans rättighetsetik (1983), och en rad andra syn-

Filosofisk tidskrift 2017 nr 3, 5–18

sätt som har en sentientistisk bas, därmed inkluderar långt fler grupper än enbart typiska vuxna människor i den grupp som tillskrivs upplevelseförmåga och moralisk status, finns *medvetandefilosofiska* teorier som menar annorlunda. Vissa teorier om medvetande hävdar att avancerade kognitiva förmågor möjliggör mer basala egenskaper. Vi bör enligt dessa utgå från innehav av ”högre” kognitiva funktioner för att kunna tala om medvetna upplevelser, och i förlängningen moralisk ställning. I denna text kommer en sådan typ av medvetandeteori (Carruthers 1998, Carruthers 2000) att kritiseras. Detta för att den, på otillräckliga grunder, utesluter icke-mänskliga djur från den krets av varelser som kan tillskrivas fenomenellt medvetande. Kritiken kommer dels att peka på att empirisk forskning talar emot ett sådant antagande och dels belysa att förnekandet av fenomenellt medvetande hos icke-mänskliga djur leder till ett underminerade av den ultimata förklaringskraften hos teorin.

2. BAKGRUND

Higher order (HO)- eller monitor-teorier om medvetande menar att vissa tillstånd är medvetna i kraft av ett slags metakognitiva tillstånd. Dessa andra ordningens mentala tillstånd kan sammanfattas som ett slags högre *representationer*. Som sådana kan de, beroende på teori, vara perceptioner, tankar eller inbyggda självrepresentationer. Idén går ut på att dessa mentala tillstånd gör vissa, första ordningens, mentala tillstånd (perceptioner) medvetna (Carruthers 2016). Detta medan de lämnar andra perceptioner omedvetna. Själva relationen mellan dessa båda, i sig omedvetna, tillstånd gör alltså den första ordningens tillstånd medvetna.

HO-teorier kan på så vis förklara till exempel subliminal perception (Persaud, McLeod et al. 2007), ett fenomen som innebär att sinnesintryck som ligger strax under den perceptuella tröskeln, och därmed subjektets fenomenella medvetande, ändå registreras och ger utslag i försök. HO-teorier verkar också kunna förklara varför vissa tillstånd i hjärnan och psyket förblir omedvetna, vilket illustreras hos patienter med blindsyn, ett tillstånd där personen ifråga är kortikalt, och fenomenellt, blind men i försök ändå lyckas lokalisera visuella objekt (Celesia 2010). Introduktionen av moderna HO-teorier (Armstrong 1968, Armstrong och Malcolm 1984, Lycan 1995) innebar därför ett synsätt som kunde förklara existensen av sådana icke-medvetna mentala tillstånd.

HO-teorier om medvetande menar därmed att första ordningens mentala tillstånd (som perceptioner) behöver särskilda metakognitiva

mentala tillstånd (högre perceptioner, tankar m.m.) för att bli medvetna. Deras generella tes kan beskrivas som:

> *Ett mentalt tillstånd M är fenomenellt medvetet om och endast om det finns ett högre ordningens mentalt tillstånd M', så att M' är en representation av M.*

Med medvetet tillstånd eller fenomenellt medvetande syftas här på sådana tillstånd som kan tillskrivas en subjektiv karaktär. Tillstånd där det ”är på ett visst sätt” att vara subjekt (Nagel 1974). I den här texten kommer begrepp som fenomenellt medvetande, qualia, medvetna upplevelser och subjektiva upplevelser att användas synonymt.

2.1. CARRUTHERS HOT-TEORI

En särskild variant av HO-teorin är Higher Order Thought-teorin (HOT-teorin). Den företräds bland annat av Peter Carruthers, som menar att medvetna mentala tillstånd (M) är det i kraft av att ha en *disposition* att vara åtföljda av högre ordningens tankar (HOT:s) (1998, 2000, 2005). Hans tes ser därför ut enligt följande:

> *Ett M är fenomenellt medvetet om det har möjlighet att orsaka HOT:s om sig självt.*

Carruthers är som nämnt också en stark förnekare av intuitionen att andra djur har fenomenellt medvetande. I sin argumentation (1998, 2000) refererar han till tre empiriska artiklar (Byrne och Whiten 1988, 1998; Povinelli 1996) som stöd för sitt resonemang. Med tanke på att studiet av djurs tänkande är en förhållandevis ung disciplin vilar Carruthers argument idag dock på svag grund. De senaste decennierna har forskning inom djurkognition producerat många resultat som visar på avancerade kognitiva förmågor hos andra djur. Dessutom är det av vikt att undersöka hur väl ett sådant avfärdande av rön går ihop med Carruthers evolutionära berättelse.

3. CARRUTHERS NATURALISTISKA TEORI

Hur ska Carruthers slutsats om andra djurs oförmåga till fenomenellt medvetande bäst förstås? Vad är det ickemänskliga djur saknar för att kunna ha subjektiva upplevelser? Och hur väl står sig denna HOT-teori mot resultat från studier av andra djurs tänkande?

3.1. NATURALISM

Carruthers (1998, 2000) menar att den dispositionella HOT-teori som han anför är *naturalistisk* (boken från 2000 har titeln *Phenomenal Consciousness: A Naturalistic Theory*). Teorin ska alltså kunna förklara varför vissa mentala tillstånd är medvetna, medan andra förblir omedvetna, med hänvisning till principer, relationer och element som finns och studeras inom naturvetenskapen.

Enligt Carruthers synsätt uppstår fenomenellt medvetande som en relation mellan första ordningens (perceptuella) mentala tillstånd och metakognitiva tillstånd. Det är förhållandet mellan metakognitionen (HOT:en) och perceptionen som möjliggör att den senare får en subjektiv karaktär. Relationen till en slags överordnade tankar förvandlar alltså luktperceptionen (till exempel registrering av doftmolekyler från ros på luktslemhinnan) till ett tillstånd som är på ett visst sätt för subjektet (upplevelsen av rosendoft). På så vis menar Carruthers att hans redogörelse för fenomenellt medvetande inte kräver referenser till något mystiskt eller övernaturligt. Vi vet redan allt vi behöver veta för att kunna förklara varför vissa mentala tillstånd är medvetna.

3.2. DEN KRITISKA EGENSKAPEN

Så varför saknar andra djur fenomenellt medvetande? Enligt Carruthers (2000) har den evolutionära utvecklingen av fenomenellt medvetande skett i två huvudsakliga steg. Först uppstod system som krävs för första ordningens sensorisk representation. Via särskilda korttidsminnesbuffertar är informationen tillgänglig för begreppsligt tänkande och bildandet av trosföreställningar (första ordningens mentala tillstånd). Detta medan andra korttidsminnesbuffertar utgör en fylogenetiskt äldre väg som gör den perceptuella informationen tillgänglig för handlingsbeslut. Det andra steget i utvecklingen av fenomenellt medvetande skedde genom tillkomsten av en tankeläsnings-komponent, som kunde använda samma perceptuella information från den första typen av korttidsminnesbuffert som möjliggör första ordningens-representationer.

Fenomenellt medvetande har, enligt Carruthers, därmed uppstått som en biprodukt. För medan *theory of mind* (ToM), och därmed förmågor som att förutsäga andras intentioner och tolka beteenden, har selekterats för *direkt*, har denna tankeläsningsförmåga även medfört att (många) perceptuella tillstånd upplevs på ett medvetet plan. Carruthers

poäng här är att tankeläsningskomponenten behöver ha tillgång till hela spektrumet av perceptuella representationer. Det hade inte räckt om ToM byggde på abstrakta begrepp, då tankeläsning ofta tycks ta hjälp av subtila skillnader, som små förändringar i ansiktsmimik. Skillnader som det ofta saknas begrepp för. ToM-komponenten är därmed riktad nedströms, mot samma system som ursprungligen utvecklades för bildandet av första ordningens-representationer. Genom att ha tillgång till visuell, auditiv och sensorisk input kan ToM-funktionen bland annat tolka rörelser och handlingar, vokalisationer och tal samt beröring. ToM har alltså utvecklats hos människor på grund av de fördelar som kan fås genom s.k. tankeläsning. Denna ”machiavelliska intelligens” (se t.ex. Byrne och Whiten 1988 för tankar om hur komplex social samvaro kräver avancerade kognitiva förmågor och ökad hjärnstorlek) möjliggör att människor kan förstå och förutsäga andras handlingar samt manipulera varandra.

När det kommer till fenomenellt medvetande i sig, nämner Carruthers (2000) en fördel, nämligen att kunna reflektera över sina egna perceptuella tillstånd. När denna anpassning väl hade skett, krävdes det enligt Carruthers, inte något mer för att människor skulle ha medvetna fenomenella upplevelser. Tillkomsten av tankeläsningskomponenten medförde därmed ett skifte från första ordningens begrepp såsom *rött, bittert, mjukt* till andra ordningens sådana som *ser rött ut, smakar bittert* och känns mjukt.

Fenomenellt medvetande är sannolikt, enligt Carruthers (2000), därmed en exaptation (Gould och Vrba, 1982), dvs. en anatomisk struktur eller ett beteende som först uppstått för ett syfte, men som kommit att få en ny funktion eller användningsområde. Sådana ”på köpet”-förmågor är långt ifrån ovanliga. Som exempel kan nämnas att fjädrar först troligtvis uppstod som en hjälp att reglera kroppstemperatur (Hickman et al. 2014). Senare kom de dock även att möjliggöra flygförmåga. På samma sätt hävdar Carruthers att tankeläsningsförmågan uppstod som en direkt följd av selektionstryck och med denna nya egenskap följde även fenomenellt medvetande. Och som flygförmåga kom att vara adaptivt, menar Carruthers att fenomenellt medvetande i allmänhet (även när det inte rör sig om ToM) främjar individens fitness. Att kunna reflektera över sina egna perceptuella tillstånd kan till exempel leda till lärdomen att vissa omständigheter leder till illusoriska perceptioner. Därmed kan subjektet lära sig att inhibera första-ordningens omdömen och påföljande olämpliga handlingar. Fenomenellt medvetande möjliggör även

att man kan informera andra om sina mentala tillstånd, något som också förefaller nyttigt, särskilt för sociala djur.[1]

Carruthers HOT-teori kräver därmed att subjektet har ToM. Det är förmågan till ToM som evolutionärt har lett till att människor har andra ordningens mentala tillstånd. Genom att kunna tolka andra individers beteenden och dra slutsatser om deras tankar, trosföreställningar och känslor, vanns många sociala och strategiska fördelar. Så förmågan till HOT:s uppstod för att ToM är adaptivt. Och eftersom HOT:s i sin tur är helt nödvändiga för fenomenellt medvetande, så behöver en varelse ha ToM för att kunna vara fenomenellt medveten. Carruthers menar att eftersom forskare inte ens är överens om huruvida schimpanser har ToM, så drar han slutsatsen att det inte är troligt att andra icke-mänskliga djur har det.

4. BRISTER I CARRUTHERS TEORI

Hittills kan åtminstone följande brister identifieras i Carruthers resonemang:

1. För det första tycks Carruthers implicit vila sitt resonemang på ett föråldrat fylogenetiskt synsätt, där utveckling betraktas som en funktion av härstamning och tid. Det går, enligt ett sådant perspektiv, att dela in arter i mer eller mindre komplexa beroende på hur de har använt tiden att utvecklas (jämför med den hierarkiska idén hos ”scala naturae”). Carruthers skriver ”eftersom, det finns en kraftfull debatt kring om ens schimpanser har theory of mind [...] verkar det osannolikt att hundar, katter eller fladdermöss är kapabla till de HOT:s som krävs” (s. 194, 2000, min översättning). Han tycks mena att en egenskap som människor har, mest sannolikt återfinns hos vår närmaste släkting och att sökande på annat håll är meningslöst. Men det är en bristfällig förståelse av hur egenskaper uppstår och utvecklas. Utveckling är en process som bör ses i ljuset av den miljö som en art har behövt anpassa sig till (Persson och Lindblom 2012). Enligt detta ”adaptiva synsätt” är det just miljön, och artens anpassning till denna, som är intressant. Inte släktskap i sig. Det kan i praktiken innebära att människoapor, till sina kognitiva förmågor, är mer lika sociala kråkfåglar än många andra primater (se till exempel Osvath et al. 2014).

1. Carruthers nämner i sin argumentation (1998, 2000) inte några andra fördelar av fenomenellt medvetande, såsom att integrera information och göra den tillgänglig för beslutsfattande eller att förmågan skulle kunna möjliggöra en större flexibilitet eller kreativitet (Flanagan, 2000; Bringsjord och Noel, 2002).

Detta leder till att hänvisning till tre relativt gamla schimpansstudier inte utgör en särskilt bra grund att vila sin argumentation på. Carruthers tes är trots allt en delvis empirisk sådan och med tanke på att studiet av djurs tänkande är en relativt ny disciplin, bör i stället en grundligare genomgång av studieresultat göras. Både med avseende på andra arter än schimpanser, men också med en ödmjukhet inför att studiet av djurs tänkande är ett mycket ungt forskningsområde där de största upptäckterna gjorts under de senaste få decennierna.

2. För det andra verkar förklaringskraften i Carruthers naturalistiska HOT-teori försvagas då den klargör människans förmåga till fenomenellt medvetande som en följd av selektion av en adaptiv tankeläsningsförmåga. Den lämpligaste kognitiva rollen för ett högre ordningens tillstånd, är enligt Carruthers (1998, 2000) de förmågor som följer av ToM. Dessa är något som icke-mänskliga djur, enligt honom, inte har, vilket visar sig i att de saknar ToM, andra metakognitiva tillstånd och därmed fenomenellt medvetande överhuvudtaget. Frågan är dock om Carruthers ”just so”-berättelse om de adaptiva fördelarna hos ToM och fenomenellt medvetande verkligen kan betraktas som den *bästa* förklaringen. Varför uppstod egenskapen endast hos ett, av tusentals sociala, ryggradsdjur? Det finns mängder med andra arter där flockstorlek, interna grupperingar, födostrategier och kommunikationssystem m.m. borde innebära att avancerade sociala förmågor som ”tankeläsning” vore oumbärlig?

4.1. THEORY OF MIND OCH METAKOGNITION I DJURRIKET

Eftersom ToM, enligt Carruthers, är nödvändigt för innehavet av HOT:s och därigenom fenomenellt medvetande, är det av vikt att undersöka distributionen av denna förmåga. ToM kan definieras som förmågan att förklara beteende genom att referera till mentala tillstånd. Det gör subjektet i stånd att tillskriva andra individer mentala tillstånd (till exempel trosföreställningar, intentioner, kunskap och viljeattityder) och möjliggör därigenom slutledningar om dessas beteenden (Ward 2015). Att ha ToM gör det dessutom möjligt för subjektet att påverka andras beteenden genom att manipulera deras mentala tillstånd (Pearce 2013). Om en individ är utrustad med ToM kommer hen därför att, utifrån till exempel en artfrändes varningsrop, kunna tolka in det mentala tillståndet rädsla samt en trosföreställning om annalkande fara. Samma individ kan kanske sedan också utnyttja denna insyn genom att själv utföra ett varningsrop och därefter snabbt äta upp den mat som den flyende

flockmedlemmen lämnar efter sig. Ett sådant vilseledande beteende är redan dokumenterat hos flera arter, däribland capuchinapa (Wheeler och Hammerschmidt 2013) och schimpans (Osvath och Karvonen 2012).

Det finns många studier som undersökt specifikt om andra djur har ToM (se till exempel Shettleworth 2010 och Lurz 2010 för en bred överblick). Exemplen nedan utgör blott några få exempel.

Schimpanser tycks kunna göra slutledningar om andras intentioner. I ett försök (Buttelmann et al. 2007) testades sex schimpanser i ett imitationsförsök. De fick se en människa hantera apparatur så att intressanta saker hände (till exempel ljud- och ljusförändringar), och hade sedan till uppgift att härma dennes beteende. Det naturliga för schimpanserna var att manipulera apparaturen med sina händer, men detta demonstrerades aldrig för dem. I försökssituationen fick de antingen (a) se en människa tända en lampa med foten då dennes händer var upptagna med att bära en låda, eller att (b) tända lampan med foten trots att händerna var fria. I det första fallet valde schimpanserna att använda händerna. Detta medan de i det andra fallet imiterade hela handlingen och slog på strömbrytaren med foten. Resultaten tycks bäst förklaras i samma termer som när försöket utförs med små barn; nämligen att schimpanserna inte bara förstod vad människan försökte göra (målet för handlingen), utan också varför hen gjorde som hen gjorde. Enligt (Tomasello et al. 2005) innebär en förståelse av handlingen som ett sätt att nå ett visst mål, också en förståelse för intentionen (se även Hare et al., 2006 och Melis et al. 2006 för ToM hos schimpanser). Positiva resultat finns även bl.a. för bonoboer (Buttelmann et al. 2012) och makaker (Santos et al. 2006).

Sammantaget visar resultat som dessa, enligt Call och Tomasello, att ”schimpanser förstår både andras mål och intentioner, såväl som deras perception och kunskap om andra. De förstår också hur dessa psykologiska tillstånd fungerar ihop för att leda till en intentional handling; dvs. de förstår andra i termer av en relativt koherent perception-mål-psykologi, där den andra handlar på ett visst sätt för att hon ser världen på ett visst sätt och har särskilda mål för hur hon vill att världen ska vara” (2008, s. 5, min översättning).

En av flera fågelarter som tycks kunna tillskriva artfränder kunskap, är snårskrikan. När en skrika upptäcker en ny födokälla så kan hen börja samla och gömma den för senare konsumtion. Om en annan skrika ser var maten göms kommer hen att stjäla den innan ägaren återvänder (Emery och Clayton 2004). I ett experiment undersökte Emery och Clayton (2007) vad som händer om en skrika får gömma maten medan en

annan skrika tittar på. När den första fågeln såg att observatören flugit iväg valde hen ofta att gömma maten på ett nytt ställe. Det var i första hand skrikor som själva stulit andras föda som också omgrupperade sina födosamlingar till nya platser. Det verkar alltså som att skrikorna både mindes att de blivit observerade, men också att de förutom att tillskriva kunskap, om de själva förut stulit, kunde tillskriva artfränder samma intention.

Det är dock viktigt att vara medveten om att den allra största majoriteten av planetens däggdjurs- och fågelarter, för att inte tala om vertebrater och sociala evertebrater, överhuvudtaget inte har studerats med avseende på ToM. Om ToM, enligt Carruthers själv (2000, 2008) dessutom tycks nästintill, eller helt, omöjlig att bevisa hos andra djur (för att deras tankeläsning egentligen handlar om associativ inlärning), kan då inte en lösning vara att försöka finna bevis för fenomenellt medvetande i sig? Enligt Carruthers är troliga adaptiva fördelar av fenomenellt medvetande, som nämnt, att man kan reflektera över sina egna mentala tillstånd samt kommunicera om dem.

Ett relevant forskningsfält borde därför vara studiet av metakognition rörande epistemiska tillstånd (studiet av förmågan att reflektera över vad man vet, se Shettlewort och Sutton 2006 för en genomgång), en förmåga som verkar involvera just de adaptiva fördelar som Carruthers ser med fenomenellt medvetande. I en studie av Hampton (2001) undersöktes förmåga till metaminne hos två rhesusapor. I försöket fick de först se en slumpmässigt vald bild på en pekskärm. Efter att denna inledande bild försvunnit, dök en annan bild upp. Denna andra bild utgjorde ett s.k. *forced-choice trial*, och aporna var tvungna att klicka på denna för att komma vidare. I den tredje och sista bilden visades flera olika bilder, och aporna hade nu till uppgift att klicka på den bild som matchade den första slumpmässiga som de först blev visade. Om de gissade rätt fick de en attraktiv matbit.

Men försöket kunde gå en annan väg. Efter att den första slumpmässiga bilden visats kunde aporna ibland få välja mellan två bilder. Den ena figuren var likadan som den i ”forced-choice trial” och ledde till samma matchningsuppgift. Men om aporna valde den andra figuren kom de till en ny bild som alltid gav belöning om den pekades på. Denna belöning var dock inte lika uppskattad. Resultaten visar att aporna lyckades matcha bilderna i 65 procenet av fallen om de tvingades att gissa, medan de hade rätt i 90 procenet av fallen när de fick välja mellan den svåra och den enkla uppgiften. Detta talar för att de, innan de valde, verkade ha en

medvetenhet om hur väl de mindes den första slumpmässiga bilden (se även Beran et al. 2014 för liknande studie på capuchinapor). Det verkar tala för att aporna kan vara medvetna om sina egna mentala tillstånd, som ju enligt Carruthers är en av fördelarna med fenomenellt medvetande.

Avslutningsvis tycks dessa studier visa att det finns andra vägar att gå för den som vill visa att djur har fenomenellt medvetna upplevelser. Det tycks därmed inte uppenbart varför Carruthers evolutionära ”just so”-berättelse om tankeläsningsförmågan är den bästa tillgängliga förklaringen till varför fenomenellt medvetande har uppstått och hur det fungerar.

4.2. ”JUST SO”-BERÄTTELSER OCH EPIFENOMENALISM

Trots detta tycks Carruthers, om möjligt, än mer skeptisk till att andra egenskaper går att påvisa hos djur. När det kommer till studier av metaminne har Carruthers (2008) kritiserat tolkningarna som alltför välvilliga. Metaminne behöver enligt Carruthers inte beskrivas i termer av medvetna subjektiva tillstånd där aporna värderar hur väl de minns, utan kan förklaras som en omedveten process där styrkan mellan övertygelse och begär samspelar. Men Carruthers hävdar konsekvent nog att ”varje gång en människa formulerar meta-kognitiva förklaringar, kommer en meta-kognitiv tanke finnas (dvs. själva tanken som finns i förklaringen)” (2008, s. 28, min översättning). Så metakognitiva tillstånd realiseras först när människan formulerar en rapport om sitt beteende. Viktigt att notera dock är att Carruthers inte anser att ToM och metakognition är språkberoende egenskaper, han skriver ”tvärtom, utgår jag från att sådana förmågor [ToM och metakognition] är oavhängiga naturligt språk” (2008, s. 28, min översättning).

Carruthers tycks alltså både implicit (1998, 2000) och explicit (2008) uttrycka skepsis över om det är möjligt att undersöka mentala tillstånd som inbegriper ToM och andra metakognitiva tillstånd och därmed påvisa fenomenellt medvetande. En möjlig tolkning av en sådan kraftfull skepsis är att denne menar att det helt enkelt inte går att bevisa förekomst av ToM eller fenomenellt medvetande genom hänvisning till beteende. Åtminstone när det kommer till varelser som saknar (mänskligt) språk. Mot bakgrund av detta och det empiriska forskningsläget inom djurs tänkande kan Carruthers argumentation leda in i en av två möjliga riktningar:

(A) ”Just so”-idéer är som bäst ett slags hypoteser. Och som sådana behöver de prövas. Carruthers behöver beakta den forskning som finns om

andra djurs tänkande och revidera sin modell. Om fler arter har ToM, är fenomenellt medvetande vidare distribuerat i djurriket än vad Carruthers hittills velat göra gällande. Kanske föregår ToM evolutionärt heller inte fenomenellt medvetna. Om så är fallet, faller den ultimata och proximata[2] förklaringskraften i Carruthers tes, nämligen att endast människor har fenomenellt medvetande och att endast ett slags andra ordningens mentala tillstånd kan göra perceptioner fenomenellt medvetna.

Det kan därför vara intressant att återvända till Carruthers resonemang om att en högre nivå av mentala tillstånd måste tillskrivas en ”lämplig roll” (1998, s. 13, min översättning). Det är i och med denna förklaring som Carruthers vill passa in idén om ToM som en basal funktion, med fenomenellt medvetande som en exaptation. Denna ”just so”-berättelse, hur elegant teorin än förefaller, tar inte hänsyn till de senaste decenniernas forskning om djurs tänkande. Den kan därmed inte utgöra den bästa förklaringen till existensen av en andra ordningens mentala nivå. En sådan berättelse bör i stället baseras på ett fylogenetiskt synsätt, där teorin dels kan förklara varför sociala djur i komplexa sociala miljöer, *generellt*, har nytta av s.k. machiavellisk intelligens som ToM. Men också varför fenomenellt medvetande, i sig, inte kan ha utvecklats utan att förutsätta metakognitiva förmågor som tankeläsning. Detta med tanke på att fenomenellt medvetande, om det har kausal kraft (vilket Carruthers antar), verkar ha fler fördelar än att kunna reflektera över och kommunicera om sina mentala tillstånd. Därför kan det mycket väl ha utvecklats vid fler än ett tillfälle och därmed också kunna realiseras via fler än en enda kognitiv mekanism. Carruthers behöver alltså se över sin proximata modell för hur perceptioner blir fenomenellt medvetna via HOT:s.

(B) Alternativet till revidering av den proximata delen av teorin, är att lämna den nuvarande ultimata förklaringen. Om Carruthers vill fortsätta hävda att positiva resultat rörande ToM, kategoribildning, självigenkänning, metaminne m.m. går att förklara i termer av första ordningens-mentala tillstånd, *samtidigt* som denne uttrycker stark skepsis till att det överhuvudtaget går att utforma experiment som kan påvisa ToM, metakognition eller fenomenellt medvetande hos icke-mänskliga individer, leder idén farligt nära tanken om filosofiska zombier (Chalmers 1996). Om andra djur uppvisar beteenden som liknar mänsklig

2. Av de olika biologiska förklaringsnivåerna, där en ultimat förklaring försöker svara på varför en egenskap har utvecklats medan en proximat förklaring försöker ge svar på hur egenskapen i fråga fungerar hos den enskilda individen.

självigenkänning, metaminne, episodiskt minne och ToM, men utan att ha fenomenellt medvetande, tycks teorin tangera epifenomenalism. Utan kausal kraft att påverka individens beteende, blir medvetandet blott en s.k. *nomological dangler*, ett bihang till ett redan komplett mentalt och fysikaliskt system. Idén om fenomenellt medvetande som en exaptation faller därför.

5. SUMMERING

Carruthers ståndpunkt är alltså inte uppdaterad enligt aktuell relevant forskning. Dessutom medför hans evolutionära förklaring till fenomenellt medvetande, tillsammans med utgångspunkten att andra djurs (och människors) beteenden kan förklaras i icke-fenomenella termer, att teorin tycks bestå av två motstridiga delar: en där fenomenellt medvetande betraktas vara adaptivt, samt en del där fenomenellt medvetande enbart kan uttryckas i *mänskliga* rapporter om metakognition. De senaste decenniernas resultat om andra arters förmåga till ToM och metakognitiva tillstånd borde leda Carruthers till att omvärdera sitt förnekande om icke-mänskligt fenomenellt medvetande, eller till ett accepterande av att teorin förefaller betrakta fenomenellt medvetande som ett epifenomen, och inte en exaptation.

Carruthers teser och argument bör, med ovanstående slutsats som bakgrund, betraktas som tvivelaktiga när det kommer till frågan om icke-mänskliga djurs innehav av fenomenellt medvetande. Teorin riskerar att, felaktigt, leda till att andra djur betraktas som känslolösa entiteter, utan förmåga vare sig till ”högre” kognitiva funktioner eller mer ”basala” medvetna upplevelser av lidande och välbehag. Utan sådana förmågor kan djur reduceras till rena ting. Med de senaste decenniernas rön om icke-mänskliga djurs tänkande och upplevelseförmåga, är Carruthers ståndpunkt, inte bara faktiskt, utan även etiskt, problematisk.

LITTERATUR

Armstrong, D. M. 2002. *A Materialist Theory of the Mind*. London: Routledge.

Armstrong, D. M. och Norman Malcolm. 1984. *Consciousness and Causality: A Debate on the Nature of Mind*. Oxford: Basil Blackwell.

Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation*. London: Athlone Press. 1970.

Beran, Michael J., Bonnie M. Perdue och J. David Smith. 2014. ”What Are My Chances? Closing the Gap in Uncertainty Monitoring between Rhesus Mon-

keys (Macaca Mulatta) and Capuchin Monkeys (Cebus Apella)”. *Journal of Experimental Psychology: Animal Learning and Cognition* 40.3, s. 303–16.
Bringsjord, Selmer, Ron Noel och David Ferrucci. 2002. ”Why Did Evolution Engineer Consciousness?”. *Advances In Consciousness Research* 34, s. 111–38.
Buttelmann, David, et al. 2007. ”Enculturated Chimpanzees Imitate Rationally”. *Developmental Science* 10.4, s. F31–F38.
Buttelmann, David, et al. 2012. ”Great Apes Infer Others’ Goals Based on Context”. *Animal Cognition* 15.6, s. 1037–53.
Call, Josep och Michael Tomasello. 2008. ”Does the Chimpanzee Have a Theory of Mind? 30 Years Later”. *Trends in Cognitive Sciences* 12.5, s. 187–92.
Carruthers, Peter. 1998. ”Natural Theories of Consciousness”. *European Journal of Philosophy*, 6, s. 203–222.
———. 1999. ”Sympathy and Subjectivity”. *Australasian Journal of Philosophy* 77.4, s. 465–82.
———. 2000. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
———. 2005. *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Clarendon Press.
———. 2008. ”Meta-Cognition in Animals: A Skeptical Look”. *Mind & Language* 23.1, s. 58–89.
———. 2016. ”Higher-Order Theories of Consciousness”. *The Stanford Encyclopedia of Philosophy*, utg. Edward N. Zalta.
Celesia, Gastone G. 2010. ”Visual Perception and Awareness”. *Journal of Psychophysiology* 24.2: 62–67.
Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
Clayton, Nicola S., Joanna M. Dally och Nathan J. Emery. 2007. ”Social Cognition by Food-Caching Corvids. The Western Scrub-Jay as a Natural Psychologist”. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362.1480, s. 507–22.
Emery, Nathan J., Joanna M. Dally och Nicola S. Clayton. 2004. ”Western Scrub-Jays (Aphelocoma Californica) Use Cognitive Strategies to Protect Their Caches from Thieving Conspecifics”. *Animal Cognition* 7.1, s. 37–43.
Flanagan, Owen. 2000. *Dreaming Souls: Sleep, Dreams and the Evolution of the Conscious Mind*. Oxford: Oxford University Press.
Gould, Stephen Jay och Elisabeth S. Vrba. 1982. ”Exaptation – a Missing Term in the Science of Form”. *Paleobiology* 8.01, s. 4–15.
Hare, Brian, Josep Call och Michael Tomasello. 2006. ”Chimpanzees Deceive a Human Competitor by Hiding”. *Cognition* 101.3, s. 495–514.
Hickman Jr, Cleveland. 2014. *Animal Diversity*. McGraw-Hill Higher Education.
Lurz, Robert W. 2010. ”Belief Attribution in Animals: On How to Move Forward Conceptually and Empirically”. *Review of Philosophy and Psychology* 2.1, s. 19–59.
Lycan, William G. 1995. ”Consciousness as Internal Monitoring”. *Philosophical Perspectives* 9, s. 1–14.

Melis, Alicia P., Josep Call och Michael Tomasello. 2006. ”Chimpanzees (Pan Troglodytes) Conceal Visual and Auditory Information from Others”. *Journal of Comparative Psychology* 120.2, s. 154.
Nagel, Thomas. 1974. ”What Is It Like to Be a Bat?” *The Philosophical Review* 83.4, s. 435–50.
Osvath, Mathias och Elin Karvonen. 2012. ”Spontaneous Innovation for Future Deception in a Male Chimpanzee”. *PloS one* 7.5, s. e36782.
Osvath, M., C. Kabadayi och I. Jacobs. 2014. ”Independent Evolution of Similar Complex Cognitive Skills”. *Animal Behavior and Cognition* 1:3, 249–64.
Pearce, John M. 2013. *Animal Learning & Cognition: An Introduction*. Hove: Psychology Press.
Persaud, Navindra, Peter McLeod och Alan Cowey. 2007. ”Post-decision wagering objectively measures awareness.” *Nature neuroscience* 10, s. 257–61.
Persson, Thomas och Jessica Lindblom. 2012. ”Djurkognition”. I *Kognitionsvetenskap* (utg. J. Allwood och M. Jensen), s. 325–36.
Povinelli, Daniel. 1996. ”Chimpanzee Theory of Mind? The Long Road to Strong Inference”. *Theories of Theories of Mind* 18, s. 293–329.
Regan, Tom. 1983. *The Case for Animal Rights*. Berkeley: University of California Press.
Santos, Laurie R., Aaron G. Nissen och Jonathan A. Ferrugia. 2006. ”Rhesus Monkeys, Macaca Mulatta, Know What Others Can and Cannot Hear”. *Animal Behaviour* 71.5, s. 1175–81.
Shettleworth, Sara J. 2010. *Cognition, Evolution, and Behavior*. Oxford: Oxford University Press.
Shettleworth, Sara J. och Jennifer E. Sutton. 2006. ”Do Animals Know What They Know?” I *Rational Animals?*, red. Susan Hurley och Matthew Nudds, s. 235–46. Oxford: Oxford University Press
Singer, P. 1975. *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: New York Review / Random House.
TFEU. 2009. Fördraget om Europeiska unionens funktionssätt. II, Article 13: http://eur-lex.europa.eu/legal-content/SV/TXT/?uri=CELEX:12012E/TXT
Tomasello, Michael, Malinda Carpenter och R. Peter Hobson. 2005. ”The Emergence of Social Cognition in Three Young Chimpanzees”. *Monographs of the Society for Research in Child Development* 70:1, s. 1–152.
Ward, Jamie. 2015. *The Student’s Guide to Cognitive Neuroscience*. Hove: Psychology Press.
Warren, Mary Anne. 1997. *Moral Status: Obligations to Persons and Other Living Things*. Oxford: Clarendon Press.
Wheeler, Brandon C. och Kurt Hammerschmidt. 2013. ”Proximate Factors Underpinning Receiver Responses to Deceptive False Alarm Calls in Wild Tufted Capuchin Monkeys: Is It Counterdeception?”. *American Journal of Primatology* 75.7, s. 715–25.
Whiten, Andrew och Richard W. Byrne. 1988. "The Machiavellian Intelligence Hypotheses". I *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, red. R. W. Byrne och A. Whiten. Oxford: Oxford University Press.

DAVID BRAX

VAD GÖR HATBROTT VÄRRE ÄN ANDRA BROTT?

> Begreppet hatbrott används allt oftare som ett samlingsbegrepp för rasistiska, främlingsfientliga och homofobiska brott. Internationellt används begreppet ”hate crime”. Det finns dock inte någon vedertagen definition av begreppet. […] Hatbrotten kännetecknas dock alltid av att de utgör ett angrepp på de mänskliga rättigheterna och strider mot grundläggande samhällsvärderingar om alla människors lika värde. (Riksåklagaren 2002)

1. VAD ÄR ETT HATBROTT?

För en analytisk filosof som närmar sig ämnet hatbrott är den uppenbara angreppspunkten begreppslig: Vad är ett hatbrott? Vad *räknas* som ett hatbrott, och hur skiljer de sig från andra brott? Det är uppenbart nog att vissa brott bör räknas dit, som t.ex. Dylan Roofs massaker i en kyrka i Charleston, South Carolina, sommaren 2015, eller skolattacken i Trollhättan i oktober samma år. Dessa brott låter sig lätt kategoriseras som hatbrott: de var rasistiskt motiverade brott utförda med avsikten att skada eller skrämma en viss grupp, och som samtidigt uttryckte en negativ värdering av denna grupp. Det råder alltså ingen tvekan huruvida dessa brott kan beskrivas som hatbrott. Men det finns också mindre uppenbara fall. I själva verket består en majoritet av de hatbrott som anmäls och redovisas i den officiella statistiken från Brottsförebyggande rådet (BRÅ) av mindre uppenbara och mer svårutredda fall. Av de mellan 4000–7000 anmälningar som varje år kategoriseras som hatbrott (se BRÅ 2016:15) leder en mycket liten andel till åtal där den s.k. hatbrottsregeln tas upp, och ännu färre till domar där den tillämpas. Enligt en tillsynsrapport som åklagarmyndigheten publicerade 2016 rör det sig om ca 30 sådana domar sedan 2013 (ÅM 2016:1). Det finns en rad skäl till detta, av vilka jag kommer att beröra två: (1) Det är oklart exakt vad det är som gör ett brott till ett hatbrott, och därmed *vad* det är som ska bevisas i rätten, och (2) Rättfärdigandet av hatbrottslagstiftningen har inte klarlagts.

Som brukligt är i filosofiska sammanhang visar det sig snart att den begreppsliga frågan inte är helt enkel att besvara. Det finns ingen officiell

Filosofisk tidskrift 2017 nr 3, 19–32

definition av ”hatbrott”. Eller snarare: ett antal delvis överlappande definitioner förekommer, och dessa är besvärande ofta formulerade i vaga termer. Detta innebär att de ger mycket begränsad vägledning gällande när och hur begreppet ska tillämpas, särskilt i mindre uppenbara fall. Detta i sin tur leder till att regeln inte tillämpas annat än i uppenbara fall, vilka alltså är få. Åklagare, i vars uppdrag det ligger att driva sådant som de har goda skäl att tro kommer leda till fällande dom, har därför begränsat incitament att driva hatbrottsvinkeln i svårare fall, vilket i sin tur innebär att vi inte får några avgöranden från högre instans i sådana fall, och den elakartade cirkeln fortsätter.

Hatbrott är en företeelse av stort filosofiskt intresse. Det är ett ovanligt tydligt uttryck för vissa sociala problem och motsättningar som till större delen består av mer svåridentifierade uttryck för fördomar t.ex. i form av indirekt diskriminering på bostads- och arbetsmarknaden. Dessa senare former av diskriminering är svårare att komma åt eftersom det där rör sig om ett *undvikande* av den utsatta gruppen, och dessutom ett missgynnande som framförallt blir synligt som en aggregerad effekt. Hatbrott är en mer *uppsökande/aggressiv* form av diskriminering, och därför lättare att identifiera och fördöma. Eftersom hatbrott faller inom straffrättens område är det en i huvudsak individorienterad form av diskriminering, vilket enklare passar med våra intuitiva uppfattningar om moraliskt ansvar.

Mitt fokus i den här texten är på frågan *varför hatbrott är värre än andra brott.* Hatbrott är en form av brottslighet som betraktas som särskilt allvarlig. ”Hat”-elementet är en *försvårande omständighet* enligt Brottsbalkens 29 kapitel § 2 punkt 7. Varför är det så? Innebär de ett större moraliskt övertramp? Är de mer klandervärda än liknande brott som saknar detta element? Det finns en tydlig koppling mellan den begreppsliga och den normativa frågan: det som *gör* ett brott till ett hatbrott bör också ha en tydlig och pålitlig koppling till vad som *gör dem värre* än andra brott. Det måste dessutom vara en normativ relevans som det finns skäl att särskilt peka ut i en straffrättslig kontext. Detta är avgörande för att straffskärpningen ska vara förtjänt.

I den här artikeln försöker jag erbjuda svar på frågorna vad som är utmärkande för hatbrott och varför de kan betraktas som särskilt allvarliga. Jag demonstrerar att båda dessa frågor har *flera* svar, och att den vaghet officiella dokument innehåller är förenlig med flera olika lösningar. Mitt *modus operandi* är att presentera alternativ, och peka på behovet att fatta ett beslut angående vad en viss användare avser. Men

jag ger inga rekommendationer gällande vilka av dessa alternativ vi bör välja. Det står dock bortom tvivel att en mer enhetlig och transparent tillämpning vore önskvärd.

2. HATBROTTSBEGREPPET

Ordet ”hatbrott” förekommer i olika sammanhang. Inom juridiken: Även om ordet inte förekommer i lagtexten så förekommer det i riktlinjer för åklagare samt i ett antal domskäl. Inom policy-området: i regeringens instruktioner och planer, i styrdokument för t.ex. polisen och brottsförebyggande rådet. Inom kriminologin (Tiby 1999), och inom samhällsdebatten i stort. Det är inte ovanligt att ord som på detta sätt används i olika kontexter också varierar något i mening. Ur ett kriminologiskt perspektiv är vi i första hand intresserade av vad som *förklarar* förekomsten av en viss typ av brottslighet, och den mest bokstavliga tolkningen av ”hatbrott” är just i form av ett förklarande begrepp: brott i någon mening orsakade av hat/fördomar. Ett motivorienterat begrepp är fullt naturligt i denna kontext. I den straffrättsliga kontexten är huvudintresset åtminstone delvis ett annat. Där handlar det om att identifiera en faktor som har relevans för brottets *allvar*, dvs. någonting vi är rättfärdigade att bestraffa en gärningsperson för. I sin inflytelserika bok *Making Hate a Crime* från 2002 påpekade Jennes och Grattet att hatbrott är ”Ett socialt problem som kräver en juridisk lösning”. Detta sociala problem erkändes i Sverige i mitten av nittiotalet som något som behövde bemötas med särskilt utformade straffrättsliga medel. Liknande resonemang har också förts på europeisk nivå (Brax 2014). Ett socialt problem överfördes därmed till straffrättens område, där särskilda villkor råder för att en åtgärd ska vara rättfärdigad. Detta kan jämföras med andra brottsrelaterade problem, t.ex. brott som sker nattetid, eller på offentliga platser. Det finns ett uppenbart intresse av att bevaka utvecklingen av olika sorters brottslighet och att utveckla strategier för att förebygga och motverka trender, men det betyder inte att brottsplatsen eller tidpunkten bör vara relevant vid straffvärdesbedömningen. Det är uppenbarligen något utöver den oroväckande frekvensen av hatbrott som gjorde att de ansågs kräva specifika straffrättsliga åtgärder.

Anledningen till att definitionsfrågan är komplex är att hatbrott har ett antal utmärkande egenskaper som kan behandlas som centrala antingen var för sig, eller som ett kluster som tillsammans utgör begreppets innehåll. Det faktum att ingen myndighet eller lagstiftare

lyft denna frågas komplexitet eller behandlat den innebär att oklarhet fortfarande kvarstår. Ett ”typiskt” hatbrott, dvs. ett brott som är enkelt identifierat som ett sådant är ett brott som har *samtliga* av dessa egenskaper: Ett offer väljs av gärningspersonen p.g.a. sin grupptillhörighet i en av de skyddade grupperna, som gärningspersonen ”hatar”, detta ”hat” uttrycks i samband med eller genom brottet, som också har som avsikt att sätta skräck i eller på annat sätt negativt påverka den utsatta gruppen, vilket också har en icke negligerbar chans att lyckas. Inbakat i denna beskrivning kan vi urskilja fem distinkta element (se Brax 2016a).

(1) *Motiv* – Gärningspersonens (GP) skäl för att utföra brottet, t.ex. en fientlig inställning till en viss grupp.
(2) *Uttryck/Mening* – Rasistisk eller liknande fientlig inställning som GP uttrycker i anknytning till eller genom brottet.
(3) *Avsikt* – Vad GP hade för avsikt att åstadkomma med brottet, t.ex. att sprida skräck i gruppen och påverka dess beteende.
(4) *Diskriminering* – GPs val av offer ur en skyddad grupp.
(5) *Konsekvenser* – Effekten brottet har på offret, den utsatta gruppen och samhället.

Problemet med att ingen rättskälla indikerat hur hatbrottsregeln ska tolkas mer exakt innebär att det saknas vägledning om hur polis och åklagare bör resonera kring svåra fall, dvs. fall som bara stämmer in på en eller några av dessa kriterier. Det finns brott som räknas som hatbrott enligt vissa av dessa men inte enligt andra. Om t.ex. ett brott uttrycker en fientlig inställning till en utsatt grupp, men inte kan knytas till något särskilt motiv eller någon särskild avsikt, bör det då räknas som ett hatbrott?

Låt oss för enkelhetens skull besvara frågan hur ”hat” ska definieras helt kort: det rör sig om en gruppbaserad negativ inställning eller attityd. Detta mentala element kan vara *motivet,* dvs. *skälet* i förklarande och/eller rättfärdigande mening, till att brottet äger rum. Det vill säga ”hatet” är svaret på frågan *varför* brottet ägde rum. Men det kan också resultera i en *avsikt,* dvs. *vad gärningspersonen försökte åstadkomma*. Men avsikten, att kränka en person på grund av grupptillhörighet kan förekomma utan koppling till just detta motiv. För att ta ett exempel: låt säga att en person vill sälja sitt hus. Nu planeras ett boende för asylsökande i området. Personen attackerar boendet med avsikten att skrämma dem därifrån. Anledningen är att han misstänker att asylboendet skrämmer bort potentiella köpare. Motivet är således ekonomiskt. Är

detta ett hatbrott? Om *avsiktsmodellen* är korrekt så är det ett hatbrott. Om vi kräver hat*motiv* är det inte det. På samma sätt förbehåller det sig med (2) och (4). Jag kan uttrycka en värdering jag inte har, t.ex. med avsikten att skapa uppmärksamhet. Jag kan dessutom välja ett offer på grund av grupptillhörighet för att jag misstänker att dessa är mindre benägna att vända sig till polisen. Skälet till att jag väljer detta offer är inte ”hat” mot gruppen, utan instrumentell opportunism. Det är givetvis klandervärt, men är det ett hatbrott? Svaret beror på vilken av de ovanstående modellerna vi väljer. Frågan vi behöver besvara för att avgöra vilken av dessa tolkningar som är rimligast är vilken som bäst motsvarar de skäl som finns att betrakta dessa brott som särskilt allvarliga.

3. STRAFFSKÄRPNINGSREGELN

Den juridiska kontexten är av lätt insedda skäl central för hatbrottsfrågan – även om ordet inte förekommer i lagtexten så är det i grund och botten en juridisk konstruktion. Statistiken som finns är baserad på anmälningar, och anmälningar är beroende av vad som är brottsligt, vilket i sin tur är baserat på hur lagstiftningen ser ut.

”Hatbrottsregeln” är en straffskärpningsregel i Brottsbalkens 29 kapitel 2§, en paragraf som behandlar försvårande omständigheter. Vid bedömning av straffvärdet ska hänsyn tas till om ”ett motiv för brottet varit att kränka en person, en folkgrupp eller annan sådan grupp av personer på grund av ras, hudfärg, nationellt eller etniskt ursprung, trosbekännelse, sexuell läggning eller annan liknande omständighet”. De brott som bedöms som hatbrott enligt denna regel är mer allvarliga än liknande brott som saknar denna egenskap. Paragrafen innehåller en precisering av mer allmänna regler för straffvärdesbedömningar som införts några år tidigare. §1 slår fast att dessa ska beakta ”den skada, kränkning eller fara som gärningen inneburit, vad den tilltalade insett eller borde ha insett om detta samt de avsikter eller motiv som han eller hon haft. Det ska särskilt beaktas om gärningen inneburit ett allvarligt angrepp på någons liv eller hälsa eller trygghet till person”. Redan före ”hatbrottsregelns” införande fanns det alltså möjlighet att ta hänsyn till rasistiska (och andra) motiv, men lagstiftaren ansåg att det fanns anledning att särskilt lyfta fram denna grupp av särskilt klandervärda motiv.

Trots att regeln existerat i Sverige i över 20 år så är den svenska rättsvetenskapliga litteraturen mycket begränsad (se Wennberg 1996, Granström 2008). Även förarbetena uttrycker sig mycket sparsamt om hur

regeln ska tolkas (SOU 1991:75, Prop. 1993/94:101), och rättspraxis är i princip obefintlig. Än så länge har bara ett enda fall där hatbrottsregeln tillämpas behandlats i Högsta domstolen (NJA 1996, s. 509), och där är den mycket sparsmakat behandlad. Med en handfull undantag är de tings- och hovrättsdomar där regeln behandlats också kortfattade i sin redovisning av den argumentation som beaktats. Det saknas således avgöranden i svåra fall, dvs. fall där vissa hatbrottsbegrepp är tillämpliga men inte andra.

Motiv. Det begrepp som oftast förekommer i samband med hatbrottslighet och som också förekommer i straffskärpningsregelns formulering är *motiv*. Åklagarmyndigheten och polisen använder dessutom uttrycket ”hatbrottsmotiv”, och i brottsbalken står det att straffvärdesbedömningen ska ta hänsyn till ”motiv”. Hatbrottsregeln komplicerar bilden genom att använda formuleringen ”motiv att kränka […] på grund av ras, hudfärg …”. Vilket av de ovan listade begreppen motsvaras bäst av denna formulering, och sammanfaller det med den rimligaste tolkningen av hatbrottsbegreppet så som det tillämpas i rätten? ”Motiv att kränka” använder visserligen begreppet ”motiv”, men märk väl att vad som följer är en ”att”-klausul. Det låter sig alltså lika gärna tolkas som en form av *avsikt*: dvs. *avsikten* att kränka. I den amerikanska rättsvetenskapliga litteraturen skiljer man ofta på motivets ”varför” och avsiktens ”vad”. Avsikten är *vad* gärningspersonen försökte göra, motivet *varför* GP försökte göra det. Som vi såg ovan är det inte svårt att föreställa sig situationer då en människa av instrumentella skäl avser att kränka en grupp, utan att därför hysa några särskilda inställningar till gruppen som sådan. Bör detta räknas som ett fall då straffskärpningsregeln ska tillämpas?

Motivbegreppet är anmärkningsvärt nog dåligt utrett i den straffvetenskapliga litteraturen. Asp och von Hirsch berör det t.ex. inte i sin artikel om straffvärde (1999). Än mer anmärkningsvärt är att begreppet inte förekommer i fängelsestraffkommitténs slutbetänkande, det som ligger till grund för påföljdsreformen och där straffvärdesbegreppet införs. Där lyder förslaget på formulering i stället så här:

> Ett brotts straffvärde bestäms av brottets svårhet med särskild hänsyn till 1. den skada eller fara som gärningen inneburit, 2. gärningsmannens skuld sådan den kommit till uttryck i gärningen. (SOU 1986:13, s. 76)

Formuleringen ”gärningsmannens skuld så som denna kommit till uttryck i gärningen” förekommer också i senare läroböcker i straffrätt (se

t.ex. Asp, Ulväng och Jareborg, 2013), dvs. efter det att den alternativa formuleringen som innehåller begreppet ”motiv” etablerats i brottsbalken. Formuleringen tyder också på att vad som avses redan i sig självt är ett normativt begrepp: ”skuld”. Inget av detta klargör hur ”motiv” ska tolkas, eller vilken uppfattning som ligger bakom dess konstruktion som relevant för straffvärdet. Detta är anmärkningsvärt med tanke på att den internationella (läs: amerikanska) diskussionen kring hatbrottslagstiftning till stor del cirkulerat just kring frågan huruvida motiv bör förstås som relevant för den typ av skuld som har straffrättslig relevans (se Hurd och Moore 2004, Kahan 2001, Brax 2016b). Flera inflytelserika rättsteoretiska kommentatorer menar att motiv korrekt förstått *inte* ingår bland de faktorer (*mens rea*) som bestämmer gärningspersonens skuld. Detta är en diskussion som helt uteblivit i den svenska kontexten, vilket bidragit till regelns oklarhet.

Frågan om hur hatbrott ska definieras har varit uppe för behandling vid flera tillfällen sedan regeln infördes, men med otillfredsställande resultat. Polisen presenterade 2015 en återredovisning av ett regeringsuppdrag angående hatbrottsarbetet. I uppdraget ingick att uppnå en ”enhetlig tillämpning av begreppet hatbrott” gemensam för polismyndigheten, åklagarmyndigheten och brottsförebyggande rådet. Det enda myndigheterna lyckades enas om var dock att använda ”hatbrott” som paraplybegrepp för (1) Hets mot folkgrupp: (BrB 16 kap. 8§), (2) Olaga Diskriminering (BrB 16 kap. 9§) samt (3) alla andra brott som faller under ”hatbrottsregeln” (BrB 29 kap. 2 § 7). Ingen vidare analys presenterades av hur straffskärpningsregeln skulle tolkas, vilka framgångsfaktorer som identifierats vid åtal och ingen vägledning gavs kring hur polis och åklagare bör arbeta för att säkra bevisning och förbättra hanteringen av hatbrott i rättskedjan.

Brå var den enda myndighet som visade sig medveten om definitionsproblemet. I sin kommentar till rapporten skrev de att den gemensamma ”definitionen” utgjorde en grund, men att mer arbete krävdes.

> Brå ser ett behov för Polisen, Åklagarmyndigheten och Brå att tränga djupare in i ordalydelsen och att ytterligare definiera vad som bör avses med olika benämningar. […] Brå menar att det krävs ett mer omfattande arbete för att målet med uppdraget, en likartad redovisning av hatbrott inom myndigheterna, ska kunna uppfyllas och att den föreslagna beskrivningen även fortsättningsvis lämnar mycket öppet för tolkning.

Det är naturligtvis rättsinstanserna som ytterst avgör hur juridiska defi-

nitioner ska förstås, men med tanke på bristen på vägledande praxis när det gäller gränsfall vore det önskvärt om brottsutredande och statistikförande myndigheter delar syn på vilka typer av fall som bör betraktas som hatbrott under utredningens gång fram till åtal.

Ytterligare ett tillfälle att klargöra vad som avsågs med hatbrott gavs i Åklagarmyndighetens (ÅM) tillsynsrapport 2016:1, en granskning av åklagarnas handläggning av hatbrott. Denna rapport samlade bland annat information om när hatbrottsregeln prövats i domstol och identifierade 30 domar där den tillämpats sedan 2013 (varav 20 utlästs ”mellan raderna” i domskälen). Rapporten presenterar dock inga exempel ur dessa domar och ger ingen information om hatbrottsdefinitionen utöver den som angetts ovan, dvs. en referens till den svårtolkade straffskärpningsregeln. Det lilla antal mer utförliga tings- och hovrättsdomar som finns där regeln behandlas, och som visar på hur olika instanserna uppfattar lagrummet, tas inte alls upp i rapporten (se Brax och Mellgren 2016).

Utifrån det tillgängliga materialet förefaller ”motiv” vara det centrala begreppet. Åklagarmyndighetens granskning påpekar bland annat att det är relativt svårt att bevisa *varför* en gärningsperson begår ett brott. I praktiken tycks det vara fall där gärningspersonen gjort ett yttrande som kan styrka ”hatbrottsmotiv” som leder till åtal (”om gärningsmannen inte sagt något eller vittnen saknas blir det genaste svårare att styrka motivet även om tillräcklig bevisning för gärningen i sig föreligger”). Uttalandet tas här som *evidens* för motivet, snarare än som konstitutivt för vad som utgör ett hatbrott. Det tycks också som om ÅM menar att det avgörande när det gäller hatbrott är att besvara frågan ”varför”. Men det är fortfarande inte uppenbart vad som här avses med ”motiv”, eftersom distinktionen mellan ”vad” och ”varför” inte uppmärksammas. Det är också möjligt att tolka detta i enlighet med diskrimineringsmodellen ovan. ”Motiv” kan tolkas som ”syfte” eller ”för vads skull”, dvs. *vad* GP försökte åstadkomma genom brottet, snarare än som ”skäl” i förklarande/rättfärdigande mening. Båda tolkningarna är också konsistenta med vad som skrivs i Riksåklagarens Promemoria från 2002:

> Syftet eller motivet bakom ett hatbrott skall klarläggas. Det är en viktig uppgift för åklagaren att leda förundersökningen på ett sådant sätt det klarläggs om ett motiv för brottet har varit att kränka offret på det sätt som anges i 29 kap. 2 § 7 brottsbalken (sic).

I brist på vidare vägledning kring hur regeln ska tolkas vänder vi oss nu

till de moraliska övervägningarna. Vilket av dessa begrepp motsvarar de skäl som finns för att betrakta hatbrott som värre än andra brott?

4. MORALISKA ÖVERVÄGNINGAR: VAD GÖR HATBROTT VÄRRE?

Hatbrott är en deskriptiv term, den handlar om en typ av brottslighet som har en viss koppling till gärningspersonens attityder. Det är som sådant det används inom forskningen, inom det fält som sedan ett par år tillbaka går under beteckningen ”hate studies”. Men det är samtidigt ett begrepp med symbolisk och normativ vikt. Det används för att särskilt fördöma en handling och i samband med förslag på särskilda åtgärder. I den straffrättsliga kontexten, som vi sett, använts begreppet för en regel som föreskriver straffskärpning. Dessa brott är alltså särskilt allvarliga. Men varför ska de betraktas som särskilt allvarliga? Här är det värt att skilja på två normativa frågor: (1) är hatbrott *moraliskt sett* värre än andra brott, dvs. brott som saknar detta inslag? (2) Är de värre på ett sätt som det är rimligt att straffa människor för?

Som nämnts finns det skäl att behandla de begreppsliga och normativa frågorna parallellt. I policydokument, riktlinjer och förarbeten förekommer uttryck som försvar för mänskliga rättigheter och allas lika värde, men det framgår sällan varför hatbrott anses utgöra ett särskilt hot mot dessa principer. Lagstiftningen beskylls därför ofta för att vara ”symbollagstiftning” och emedan den symboliska och expressiva betydelsen av lagstiftning inte är att förakta (se t.ex. Kaupinnen 2015) så kan denna i princip bara utföras effektivt när lagstiftningen tillämpas, vilket i sin tur kräver begreppslig precision. I polisens återredovisning (2015) står t.ex. att hatbrott är ”resultatet av bristande respekt för mänskliga rättigheter och lika värde”. Riksåklagarens ”Promemoria och riktlinjer för bekämpning av hatbrott” (2002) anger, efter ett erkännande att definitionen inte är uppenbar

> Hatbrotten kännetecknas dock alltid av att de utgör ett angrepp på de mänskliga rättigheterna och strider mot grundläggande samhällsvärderingar om alla människors lika värde. […] Ett sätt att bedöma om det rör sig om ett hatbrott kan vara att försöka fastställa om brottet begåtts mot bakgrund av gärningsmannens uppfattningar, värderingar eller ideologier som strider mot principen om alla människors lika värde. […] Ett hatbrott kan också motiveras av hat eller illvilja mot vad en person anses representera. […] utgör brott mot grundläggande fri- och rättigheter och därmed

> brott mot det fria, demokratiska samhället och de värderingar som ligger till grund för samhällsordningen. (s. 22–23)

I den proposition där straffskräpningsregeln läggs fram för riksdagen står bland annat att det inte råder någon tvekan om att ”brott med rasistiska och liknande motiv har ett särskilt högt straffvärde”.

> Vårt samhällsskick bygger på att alla människor har ett lika värde oavsett ras, hudfärg och etniskt ursprung. Rasism och liknande yttringar som tar sig uttryck i förakt eller förtryck av utsatta grupper är oförenliga med grundläggande värderingar och kan därför aldrig accepteras. Samma enighet råder om att det över huvud taget är av största vikt att alla innevånare i vårt land är trygga mot brott. (Prop. 1993/94:101, s. 20–21)

Dessa vaga hänvisningar till alla människors lika värde hjälper oss dock inte mycket för att förstå de normativa grunderna för hatbrottslagstiftningen. *Hur* och *varför* innebär det en större kränkning av människors lika värde att utföra ett brott av rasistiska motiv? Vad jag kan se finns huvudsakligen sex distinkta grunder för påståendet att hatbrott är moraliskt sett värre än andra (”parallella”) brott. Dessa kan användas för att rättfärdiga straffskärpningar för dessa brott, förutsatt att vi accepterar att straffbestämmelser vilar på moralisk grund.

Hatbrott gör mer skada (*Individ, grupp, samhälle*). Påståendet att hatbrott gör mer skada än andra brott har länge använts som rättfärdigande för hatbrottslagstiftning. Hatbrott gör mer skada på individer (Iganski och Lagou 2015) på gruppen (Noelle 2001) och på samhället (Lawrence 1994). Skadeprincipen är central för straffrätten, och det är förhållandevis okontroversiellt att straffa människor i proportion till den skada de gör eller avser att göra. Dock kan forskningen bara etablera att hatbrott *tenderar* att göra mer skada än andra brott. Hatelementet används som en *proxy* för särskilt omfattande skada (Hurd och Moore 2004). Är det en pålitlig nog proxy för att grunda straffrättsligt ansvar?

Hatbrottsmotiv är värre än andra motiv. Hatbrott som begås av t.ex. rasistiska motiv strider mot principen om alla människors lika värde. De behandlar människors ursprung, religion, hudfärg etc. som ett skäl att begå ett brott, dvs. behandla någon illa. Hatbrott innebär enligt denna tolkning ett större moraliskt ”*misstag*” än brott begångna av andra motiv (t.ex. ekonomisk vinning). (Se Kahan 2001, Brax 2016b.) Detta är, som vi sett, en kontroversiell position att anföra inom straffrätten. Även om vi köper att hatbrottsmotiv är värre än andra motiv, är det den sortens

element som vi bör straffa för? Innebär det ett avsteg från en handlingsorienterad uppfattning gällande vad som bör vara straffbart?

Hatbrott sker med särskilda onda avsikter. Straffskärpningsregelns ”motiv att kränka” kan tolkas som uttrycket för en avsikt. Att avse att orsaka särskilt omfattande skada eller kränkning är en oproblematisk grund för skuld, så som detta förstås inom straffrätten. Hatbrott är enligt denna tolkning värre än andra brott av samma skäl som terror-brott är det, dvs. med hänvisning till vad gärningspersonen avsåg att åstadkomma genom brottet.

Diskriminering, orättvisa. Det faktum att hatbrott innebär ett *diskriminerande val av offer* (Lawrence 1994) kan i sig själv betraktas som moraliskt relevant. Det finns dock olika uppfattningar om *när* och *varför* diskriminering är fel. Enligt en sammanfaller det med (2) ovan, enligt annan med (5). Det diskrimineringsbaserade begreppet ovan menar att diskriminering sker när offrets grupptillhörighet på ett eller annat sätt förklarar gärningspersonens val. För att göra (4) distinkt från (2) och (5) kan vi därför säga att det är fel att välja offer på detta sätt, oavsett *varför* detta val görs. Oavsett om jag attackerar mina mörkhyade grannar för att jag hatar mörkhyade, eller för att jag tror att de då kommer att flytta vilket skulle ha en positiv inverkan på värdet av min bostad, så skulle detta vara ett hatbrottselement vilket ökar min skuld.

Utsatta/sårbara offer. Hatbrott drabbar framför allt de grupper som redan befinner sig i en utsatt situation. Dels just på grund av risken för den här typen av brottslighet, men också av andra skäl. Detta kan vara moraliskt relevant av flera skäl. Dels att skadan som sker (se punkt 1) då riskerar att bli större. Men också för att det är klandervärt att ”sparka på de som ligger”. Ytterligare ett skäl att betrakta detta som en förvärrande omständighet att den skada som sker, oavsett om den är större än skadan vid annan brottslighet, drabbar de som har det sämst i samhället. Denna tolkning är förenlig med straffskärpningsregelns förarbeten där den åtminstone indirekt beskrivs som en form av minoritetsskydd. Det är dock nämnvärt att regeln är neutralt formulerad, och att det i princip tycks möjligt att tillämpa den när en medlem av ”majoritetsbefolkningen” drabbas. Detta är en kontroversiell fråga, som ofta dyker upp i samhällsdebatten, men som aldrig behandlats på ett tillfredställande vis. Vi kan betrakta den debatten som en konflikt mellan tolkningarna (4) och (5).

Uttryck. Det budskap som hatbrott uttrycker är särskilt klandervärt. Hatbrott gör en ”symbolisk” skada som påverkar offrets sociala ställning, om det fortgår utan åtgärd (se Kauppinen 2015). Hatbrott är en

form av hets mot folkgrupp som sker med en brottslig handling som medium. Denna tolkning av vad som gör hatbrott mer fel än andra brott väcker frågor om yttrandefrihet, dvs. huruvida straffskärpningen innebär en bestraffning av vad människor säger. I USA, som saknar ”hate speech” av konstitutionella skäl, har det gjorts klart att det *inte* är på denna grund som hatbrottslagstiftningen vilar. Också här kan man påpeka att medan förnedrande och kränkande symboliska uttryck mycket väl kan betraktas som moraliskt relevanta, är det ytterligare ett steg att konstruera detta som ett skäl att tillämpa straff.

Som vi sett ovan finns det kopplingar mellan den normativa och den begreppsliga frågan. Skälen vi har att betrakta dessa brott som särskilt allvarliga har implikationer för vad som är den rimligaste tolkningen av hatbrottsbegreppet, åtminstone så som detta används i den rättsliga kontexten. Om det är på grund av att vissa motiv är särskilt klandervärda som straff är förtjänt så är också en ”motiv”-tolkning av regeln särskilt lämplig. Även om det ytterst är avsikter, diskriminerande val av offer och uttryck som utgör bevisningen för sådana motiv. Enligt åklagarmyndigheten (2016:1) är det i praktiken bara brott där gärningspersonen yttrat sig på ett sätt som stödjer hatmotiv som leder till fällande domar där regeln tillämpas. Den praktiska definitionen av hatbrott kan således vara uttrycksorienterad även om den teoretiskt och på normativa grunder är motiv-orienterad. Den normativa grunden har också betydelse för huruvida och i vilka fall straffskärpningar är rättfärdigade. Om det handlar om den skada brotten gör eller riskerar att göra är det fullt möjligt att detta varierar mellan grupper (se Iganski och Lagou 2015). Det vore i så fall värre att attackera en grupp där skadan tenderar att spridas i särskilt hög utsträckning, t.ex. på grund av utsatthet i övrigt, bristande förtroende för polisen etc.

Som jag nämnt finns en skillnad mellan att säga att en faktor är normativt relevant och att säga att denna bör vara relevant för straffvärdesbedömningar. Detta är värt att hålla i minnet när hatbrottslagstigning diskuteras: den kan diskuteras på moralisk grund (är hatbrott värre än andra brott?) och på rättsteoretisk grund (är de värre på ett sätt som bör tillmätas straffrättslig relevans?). Jag hoppas att uppräkningen ovan åtminstone kan underlätta denna diskussion, och öppna för en djupare rättsfilosofisk diskussion om straffvärdesbedömningar i allmänhet, och om hatbrottslagstiftningen och dess tillämpning i synnerhet.

5. AVSLUTANDE ANMÄRKNINGAR

Det är av avsevärd praktisk betydelse hur begreppet ”hatbrott” tolkas, och hur rättsväsendets aktörer förstår de normativa grunderna för detta begrepp. I den här texten har jag inte gett några rekommendationer kring vilka alternativ som är att föredra. Av rättssäkerhetsskäl är det viktigt att tillämpningen är enhetlig, vilket kan ske oavsett vilket av begreppen som tillämpas. Av legitimitetsskäl är det viktigt att vi bestämmer vilken eller vilka de normativa grunderna är, och att dessa stämmer överens med etablerade grunder för straffrättsliga sanktioner i allmänhet.

Brottskoder och straffvärdesbedömningar är av stort intresse ur moralfilosofisk synpunkt, vilket gör det relativt svaga intresset för rättsfilosofi inom svensk akademisk filosofi svårbegripligt. Inom straffrätten nedlägger samhället vissa principer och uttrycker samtidigt att vissa intressen är särskilt skyddsvärda. Med tanke på den oenighet som råder i moraliska frågor är detta område där dessa frågor faktiskt måste avgöras, och dessutom kodifieras på ett enhetligt och rättssäkert sätt, av stort intresse. Vilka moraliska principer är det straffrätten inkorporerar? Vad rättfärdigar överhuvudtaget ett straff? Vilken funktion har straffrätten, och vad kan lagstiftare legitimt använda den till? Hatbrottsregeln är, tror jag, ett nyckelfenomen här, och en fördjupad diskussion om dess grunder har potential att avslöja djupgående meningsskiljaktigheter i straffrättsteoretiska frågor.

LITTERATUR

Asp, Petter, Magnus Ulväng och Nils Jareborg. 2013. *Kriminalrättens grunder*. Uppsala: Iustus.

Asp, Petter och Von Hirsch, Andrew. 1999. ”Straffvärde”. *Svensk Juristtidning*, häfte 2, s. 151–75.

Brax, David. 2014. ”Hatbrottslagstiftning i Sverige och i Europa”. I *Förhoppningar och Farhågor: Sveriges första 20 år i EU*, red. Linda Berg och Rutger Lindahl, s. 181–96. Göteborg: Centrum för Europaforskning vid Göteborgs universitet.

———. 2016a. ”Hate Crime Concepts and Their Moral Foundations: A Universal Framework?” I *The Globalization of Hate: Internationalizing Hate Crime?*, red. Jennifer Schweppe och Mark Walters. Oxford: Oxford University Press.

———. 2016b. ”Motives, Reasons, and Responsibility in Hate/Bias Legislation”. *Criminal Justice Ethics* 35, nr 3, s. 230–48.

Brå. 2016:15. *Hatbrott – Statistik över polisanmälningar med identifierade hatbrottsmotiv och självrapporterad utsatthet för hatbrott*. Stockholm: Brottsförebyggande rådet.

Granström, Görel. 2008. ”Rättsväsendets prioritering av hatbrott”. *Retfærd. Nordisk Juridisk Tidsskrift* 31, nr 4, s. 83–101.
Hurd, Heidi M. och Michael S. Moore. 2004. ”Punishing Hatred and Prejudice.” *Stanford Law Review* 56, nr 5, s. 1081–1146.
Iganski, Paul och Spiridoula Lagou. 2015. ”Hate Crimes Hurt Some More Than Others: Implications for the Just Sentencing of Offenders”. *Journal of Interpersonal Violence* 30, nr 10, s. 1696–1718.
Jenness, Valerie och Ryken Grattet. 2002. *Making Hate a Crime: From Social Movement Concept to Law Enforcement Practice*. New York: Russell Sage Foundation.
Kauppinen, Antti. 2015. ”Hate and Punishment”. *Journal of Interpersonal Violence* 30, nr 10, s. 1719–37.
Kahan, Dan, M. 2001. ”Two Liberal Fallacies in the Hate Crimes Debate”. *Law and Philosophy* 20, nr 2, s. 175–93.
Lawrence, Frederick, M. 1994. ”The Punishment of Hate: Toward a Normative Theory of Bias Motivated Crimes”. *Michigan Law Review* 93, nr 2, s. 320–81.
NJA 1996, s. 509, nr 84.
Noelle, Monique. 2001. ”The Ripple Effect of the Matthew Shepard Murder: Impact on the Assumptive Worlds of Members of the Targeted Group”. *American Behavioral Scientist* 46, nr 1, s. 27–50.
Polisen, Utvecklingsavdelningen. 2015. Återredovisning av regeringsuppdraget beträffande hatbrott dnr A121.608/2014. Prop. 1993/94:101 Åtgärder mot rasistisk brottslighet och etnisk diskriminering i arbetslivet.
SOU 1991:75 *Organiserad rasism*.
SOU 1986:13 *Fängelsestraffkommitténs huvudbetänkande*.
Tiby, Eva. 1999. *Hatbrott? Homosexuella kvinnors och mäns berättelser om utsatthet för brott*. Diss. Stockholm: Kriminologiska institutionen, Stockholms universitet.
Åklagarmyndigheten. 2016:1. *Hatbrott – en granskning av åklagarnas handläggning*, tillsynsrapport.
Riksåklagaren. 2002. *Promemoria och riktlinjer för bekämpning av hatbrott*, dnr 2002/0025.
Wennberg, Suzanne. 1996/97. ”Klippanmålet och behandlingen av rasistiska motiv.” *Juridisk Tidskrift*, nr 3.

ALLA CHOIFER

ATT FÖRSTÅ SIG SJÄLV OCH ATT FÖRSTÅ DEN ANDRE

Ett filosofiskt perspektiv på utvecklingspsykologi

1. INTRODUKTION

Vi är sociala varelser och vår förståelse för (eller vårt tänkande om) vem vi är hänger oundvikligen samman med vår förståelse av (vårt tänkande om) de andra. Vilket samband dessa två typer av förståelse (tänkande) har – självförståelsen och förståelsen av den andre – kommer jag att undersöka här. Filosofer brukar skilja mellan olika typer av självuppfattning men något förenklat kan man särskilja två ytterligheter: den omedelbara, oreflekterade och direkt givna upplevelsen och den objektifierande, reflekterande relationen till egna mentala tillstånd och känslor. Ett exempel på den förra är en känsla av upplevd glädje eller ilska, medan den senare ofta uttrycks som någon form av propositionell kunskap som, till exempel, ”Jag är glad” eller ”Jag är nog arg därför att jag har sovit för lite idag”. Beroende på hur självförståelsen eller självkunskapen förstås så brukar olika antaganden göras om det sätt på vilket andra människor blir involverade i vårt uppfattande av oss själva. Det finns olika teorier om hur detta går till och hur de två olika sätten att uppfatta (tänka) hänger samman. De två övergripande paraplyteorierna på området är den s.k. simuleringsteorin och teori-teorin. Simuleringsteori-teoretiker utgår från att vi för att förstå oss på den andre ”simulerar” hur *den andre* tänker eller känner. Självkunskapen om det simulerade blir därmed primär för vår förståelse av de andra. Teori-teorins anhängare hävdar dock att självförståelsen och förståelsen av de andra utvecklas parallellt och med hjälp av en teori-applicerande mekanism. Ett viktigt begrepp i detta sammanhang är ”mentalisering”. Begreppet används ofta för att beteckna vårt tänkande om våra egna och andras mentala akter.

Forskningen på området har också intresserat sig för hur förmågan att mentalisera utvecklas och hur den relaterar till människans övriga kognitiva och sociala utveckling. Utvecklingspsykologin – vetenskapen om människans psykologiska utveckling: förändringar i beteenden, reaktionsmönster och psykiska egenskaper under barndomen

Filosofisk tidskrift 2017 nr 3, 33–42

och ungdomsåren – erbjuder ett rikt empiriskt underlag för att studera mentaliseringsförmågan. Inom ramarna för detta perspektiv har utvecklingspsykologerna utarbetat ett viktigt testförfarande (det s.k. *false-belief*-testet eller i svensk tappning falsk-tro-testet) som anses fånga den kognitiva förmåga som ligger till grund för mentalisering.

Syftet med denna artikel är att utifrån ett filosofiskt perspektiv problematisera relationen mellan förståelsen av oss själva och förståelsen av andra. Inom ramarna för det utvecklingspsykologiska perspektivet kommer jag att kritiskt granska implikationerna av falsk-tro-testet, vilket för med sig en ny tolkning och en ny förståelse av detta test. Den nya tolkningen bär med sig ett viktigt budskap: vår förståelse av den andre föregår och är en nödvändig förutsättning för självförståelsen. Denna slutsats strider dock mot de två i dagens forsknings mest populära förklaringsmodellerna för mentalisering: simuleringsteorin och teori-teorin. En annan viktig konsekvens av den nya tolkningen är att den ifrågasätter en viss idé om barnens tidiga (före framgång i falsk-tro-testet) mentaliseringsförmåga.

Jag börjar med att inom det utvecklingspsykologiska perspektivet redogöra för det falsk-tro-test som fick en så stor genomslagskraft i diskussionerna kring mentaliseringen (avsnitt 2). Jag kommenterar vidare en viktig konsekvens av detta test, vilket leder till en ny förståelse av det (avsnitt 3). Den nya tolkningen innebär att falsk-tro-testet kan ses som ett test av barns mentala förmåga att differentiera sig från människor runt omkring sig – en viktig förutsättning för mentaliseringen. Denna tolkning överensstämmer med Piagets lära om egocentrism. Piagets teoretiska resonemang erbjuder också en möjlighet att spekulera kring sambandet mellan självförståelsen och förståelsen av den andre: förståelsen av den andre är en nödvändig förutsättning för förståelsen om en själv (avsnitt 4). I avsnitt 5 reflekterar jag slutligen över på vilket sätt det framlagda sambandet strider mot de vanliga förklaringsmodellerna för vår förmåga att mentalisera.

2. Det utvecklingspsykologiska perspektivet och falsk-tro-testet

Inom modern kognitiv psykologi och neurovetenskaplig forskning har man diskuterat den s.k. *theory of mind* (ToM) – en beteckning på den uppfattning (den teori) som ett subjekt antas ha om sitt eget och andra personers själsliv (mind). ToM står alltså inte för någon specifik teori

om hur medvetande (mind) fungerar, utan är en generell beteckning på vår förmåga att tolka och förstå våra egna och andras avsikter och beteenden. Att det föreligger ett samband mellan de två olika typerna av förståelse – självförståelsen och förståelsen av den andre – antyds redan i den allra första beskrivningen av fenomenet då termen myntades av Premack och Woodruff (1978, s. 515): ”När vi säger att en individ har en theory of mind [ToM] menar vi att individen tillskriver sig själv och andra mentala tillstånd” (min översättning; i det följande är alla citat från engelskspråkig litteratur mina egna översättningar till svenska). En liknande idé uttrycks också i befintliga definitioner av begreppet. Abu-Akel (2003, s. 29) refererar till exempel till ToM med följande ord: ”Theory of mind (ToM), som ibland används omväxlande med termer som mentaliserande kapacitet, är förmågan att representera egna och andras mentala tillstånd, exempelvis avsikter, trosföreställningar, önskningar, begär och kunskap”. Happé, Brownell och Winner (1999) är också av samma uppfattning när de refererar till fenomenet som ”förmåga att tillskriva sig själv och andra tankar och känslor”.

En viktig aspekt som diskuteras i detta sammanhang är också *när* i människans utveckling man förvärvar självförståelsen och förståelsen för de andra. Man skiljer här mellan en explicit och en implicit förmåga att mentalisera. ”Det mentaliserande (theory of mind) systemet fungerar troligen från cirka 18 månaders ålder och tillåter ett implicit tillskrivande av avsikter och andra mentala tillstånd”, skriver Uta Frith och Christopher Frith (2003, s. 459). Forskarna talar här om en *implicit* mentaliseringsförmåga. De refererar till en studie av Leslie (1987) som visar att ett 18 månader gammalt barn verkar kunna förstå moderns intentioner och börja skratta när modern lekfullt tar en banan och låtsas prata i telefon. Man menar då att barnen har en förmåga att uppfatta ett leksammanhang, vilket man i sin tur tolkar som en otvetydig manifestation av förmågan att mentalisera. I kontrast till en implicit ToM talar man också om en *explicit* förmåga att mentalisera: ”Mellan 4 och 6 års ålder blir explicit mentaliserande möjligt, och från den åldern kan barnen förklara de vilseledande anledningar som givit upphov till en felaktig tro” (Frith och Frith, 2003, s. 459). Det stringenta testet på existensen av en ToM går ut på att man testar barns förmåga att förutsäga en annan persons beteende på basis av denna persons falska trosuppfattningar. Därav namnet, det s.k. falsk-tro-testet. Det experimentella paradigmet har genomgått olika modifikationer men utvecklades från början av Wimmer och Perner (1983).

I det ursprungliga testet (Wimmer och Perner 1983) presenterade man följande scenario för ett barn: Maxi har en chokladbit som han lägger i ett blått skåp. Maxi går sedan ut. Maxis mamma kommer in och flyttar chokladbiten till ett grönt skåp. Maxi kommer tillbaka för att få sin choklad. Barnet ombes sedan att svara på frågan om var någonstans Maxi kommer att leta efter chokladbiten. Det finns två olika sätt för barnet att svara på denna fråga: antingen pekar barnet på det blå skåpet där chokladbiten var när Maxi var närvarande, eller pekar barnet på det gröna skåpet (dit chokladbiten förflyttades efter att Maxi lämnat rummet). Från ungefär 4,5 års ålder börjar barn förstå detta scenario och kan förstå den andres uppfattning om verkligheten (det blå skåpet) även om denna uppfattning strider mot de (från barnets perspektiv) verkliga förhållandena (det gröna skåpet). Yngre barn kan inte lösa denna uppgift och genom att peka på det gröna skåpet (där chokladbiten finns just nu sedd från barnets perspektiv) projicerar dessa barn på den andre sin egen uppfattning om de verkliga förhållandena. Att barn vid 4–6 års ålder kan förstå de missledande omständigheter som ger upphov till den andres uppfattningar ser man som ett tecken på en *explicit* mentaliseringsförmåga. Man har fortsatt att studera denna förmåga i en mängd modifikationer av falsk-tro-testet. Wellman, Cross och Watson (2001) har utvärderat 178 olika ToM- studier som sammanfattar testresultaten för flera tusen barn. Slutsatsen av forskarnas meta-analys är att barnens prestationer i olika varianter av falsk-tro-testet inte har något systematiskt samband med vilken variant som använts. Barnens framgång i falsk-tro-testet ligger vidare på ungefär samma åldersnivå oberoende av barnets kulturella och sociala bakgrund. Falsk-tro-testet visar sig vara ett robust mått på en viktig form av kognitiv mognad i barns tidiga utveckling.

Det finns dock en viktig konsekvens av falsk-tro-testet som inte har fått någon större uppmärksamhet i forskningen på området. Barnets misslyckande med att lösa falsk-tro-uppgiften är ett tecken på att barnet upplever alla andra i sin omgivning som om dessa andra tänker och känner på samma sätt som barnet själv gör. Utifrån barnens prestation i falsk-tro-test kan man således dra slutsatsen att de barn som tillskriver den andre sin egen uppfattning om de rådande förhållandena inte riktigt klarar att differentiera sig från människor runt omkring sig. Barnens förmåga att skilja sig från andra är dock avgörande för att vi alls ska kunna tala om deras förmåga att mentalisera. Att kunna avgöra barnets kognitiva mognad i fråga om att skilja sig själv från de andra är också av vikt för våra spekulationer om sambandet mellan självförståelsen och

förståelsen av de andra. Detta moment är vidare väsentligt för hur den implicita mentaliseringen, det vill säga, mentaliseringen innan barnens framgång i falsk-tro-testet ska förstås. Jag reflekterar kring detta mer utförligt i nästa avsnitt och börjar med att närmare beskriva den viktiga konsekvensen av falsk-tro-testet.

3. EN KONSEKVENS AV FALSK-TRO-TESTET

Som nämnts ovan så finns det två möjliga utfall av falsk-tro-testet. Antingen besvarar barnet frågan om vart den andre tror att det gömda objektet finns genom att peka på det gröna skåpet, eller så pekar barnet på det blå skåpet. I det förra fallet (det gröna skåpet) så klarar inte barnet testet – det överför till den andre sin egen uppfattning om de verkliga förhållandena. I det senare fallet (det blå skåpet) visar barnet att det förstår den andre som en individ med en egen (från barnets perspektiv skild) uppfattning om världen – en uppfattning som strider mot barnets kunskap om världen.

Barnets prestation i falsk-tro-testet bör, med andra ord, tolkas som att ett barn som inte kan klara testet tillskriver den andre sin egen uppfattning om världen. Medan de barn som löser falsk-tro-uppgiften visar att de accepterar den andre som en individ med sina egna (och från barnets perspektiv kontrafaktiska) uppfattningar om världen. Den direkta slutsats som följer är att innan den kognitiva mognad som signaleras med framgång i falsk-tro-testet så sammanfaller barnets förståelse för sitt eget och andra människors sätt att tänka, känna och handla. Annorlunda uttryckt, falsk-tro-testet sätter barns förmåga att skilja sig själva från människor runt omkring dem på prov. Framgång i falsk-tro-testet visar därmed att barnet förstår den andre som en från barnet skild individ; en individ med ett eget mentalt liv. Om detta är riktigt så innebär det att falsk-tro-testet kan förstås som en experimentell procedur som avgör huruvida ett barn är kognitivt moget att skilja sig själv från människor runt omkring.

Denna förståelse av testet ligger delvis i linje med den vanliga uppfattningen om falsk-tro-testet som ett test av barns förmåga att mentalisera. Dock fokuserar den nya tolkningen på en viktig förutsättning för mentaliseringen: barnets förmåga att skilja sig själv från andra. Ett barn som inte känslomässigt och tankemässigt skiljer ut sig själv från de andra kan inte heller attribuera några mentala tillstånd till sig själv eller till andra.

Detta resonemang går på tvärs mot en vanlig idé om barns implicita mentalisering, det vill säga den mentalisering som anses fungera redan

från och med 18 månaders ålder. I vilken mening mentaliserar ett barn som inte kan differentiera sig från människor i sin omgivning?[1] Vad man till nöds kunde kalla ”mentaliseringen” hos ett barn som inte riktigt drar en skiljelinje mellan sig själv och den andre kommer till uttryck till exempel då barnet täcker för sina ögon och tror att det inte syns. Ett sådant barn lever fortfarande i Piagets egocentriska värld – alla som bebor barnets värld tänker och upplever *automatiskt* världen på det sätt som barnet själv tänker och upplever den. Det är just detta förhållande som kommer till uttryck i och med barnets oförmåga att lyckas med falsk-tro-uppgiften.

Piagets lära om barns ursprungliga egocentriska relation till sin omvärld är här i samklang med den nya tolkningen av falsk-tro-testet. I barns egocentriska värld finns det överhuvudtaget inte någon ”annan” som behöver förstås. Piagets teoretiska reflektioner lägger också en grund för hur sambanden mellan mentaliseringen om en själv och mentaliseringen om det andre ska förstås. I nästa avsnitt kommer jag att med Piagets teori om barnets kognitiva utveckling som kontext diskutera dessa två nära relaterade färdigheter: självförståelsen och förståelsen av den andre. Jag kommer att visa att förmågan att upptäcka en annan skild från en själv (framgång i falsk-tro-testet) är en viktig förutsättning för barnets förmåga att referera till sig själv och därmed uppnå självförståelse.

4. Piaget, egocentrism och barns kognitiva utveckling

Små barns oförmåga att referera till sig själva uppmärksammades för första gången på 1920-talet av Jean Piaget, en schweizisk psykolog och epistemolog. Detaljerade iakttagelser av förändringar i ett barns kognitiva organisation från födelsen till tonåren ledde Piaget till slutsatsen, att barn under en viss ålder inte kan se sig själva och världen runt omkring sig ur andra personers perspektiv. Piaget tolkade det han observerade som ett resultat av barnets outvecklade kapacitet att fullständigt differentiera sig själv från andra personer i sin omgivning. En konsekvens av denna bristande differentiering är, menade Piaget, att barnet visar en tendens att uppfatta och tolka världen endast utifrån sitt eget

1. Hur ska vi förstå ett 18 månader gammalt barn då det skrattar när hans/hennes mor tar en banan och låtsas prata i telefon? Det barn som känslomässigt och tankemässigt identifierar sig med sin mor uppfattar kanske skämtet med en banan på ett mycket enklare sätt än att det tillskriver modern en ovanlig mental intention. Det barnet kan, till exempel, helt enkelt skratta på grund av att två för barnet vardagliga och mycket bekanta objekt nu har förväxlats så att *prata i telefon* blev *prata i banan* i stället.

perspektiv – ett fenomen som Piaget döpte till ”egocentrism”. En annan konsekvens av barnets egocentriska referensram är att det är omöjligt för barnet att se sig själv från en annan persons perspektiv och som ett objekt för sina egna eller andra personers betraktelser.

Med tiden blev dock Piaget mindre nöjd med beteckningen ”egocentrism” och lade ner möda på att förklara hur det utvecklingspsykologiska fenomenet skiljer sig från egocentrism i den allmänna betydelsen av narcissistisk hållning. I vardagsspråk syftar ju ordet ”egocentrisk” ofta på en självupptagenhet, självcentrering och överdriven fokusering på den egna personen, de egna upplevelserna och de egna känslorna. Det som Piagets term ”egocentrism” betecknade innebär ju precis motsatsen, nämligen, en oförmåga hos barnet att referera till sig själv. Detta är själva kärnan i egocentrism, eller som Piaget uttryckte det ”egocentrism när den är självmedveten, är inte längre egocentrism” (Piaget 1959, s. 268).

I sin egocentriska värld är barnet centrum för universum, ett centrum som är omedvetet om sitt egocentriska predikament. För barnet i det odifferentierade själv-andra tillståndet förefaller alla människor runt omkring barnet tänka och uppleva världen på samma sätt som barnet tänker och upplever den själv. Just detta förhållande fångas av falsk-tro-testet – ett barn som inte klarar testet tillskriver automatiskt människor runt omkring sig en uppfattning om världen som barnet själv har. Till skillnad från detta visar de barn som klarar testet att de kan erkänna de andra som tänkande människor med sina egna uppfattningar om världen. Förmågan att förstå sig på den andre blir därmed avgörande för att barnet har en möjlighet att ta denna andres perspektiv på sig själv och därigenom uppnå självreferens.

Med andra ord, för att kunna referera till sig själv, i meningen att ta den andres perspektiv på sig själv som ett objekt, så måste det finnas en sådan annan skild från barnet självt. Barnet måste klara falsk-tro-testet för att kunna använda den andre som en utsiktsplats för att ta en titt på sig själv. Den nya tolkningen av falsk-tro-testet som ett test för barns mentala mognad i att förstå den andre som en från barnet skild individ erbjuder därmed en möjlighet att spekulera över hur sambandet mellan mentaliseringen om sig själv och mentaliseringen om den andre ska se ut. Framgången i falsk-tro-testet (mentaliseringen om den andre) blir därmed den nödvändiga *förutsättningen* för att barnet ska kunna mentalisera om sig självt. Med andra ord: mentaliseringen om en själv blir därmed en senare historia än mentaliseringen om den andre.

De förklaringsmodeller som idag gör anspråk på att förklara hur

mentaliseringen om en själv och den andre hänger samman – simuleringsteorin och teori-teorin – tar inte hänsyn till själv-andra-differentieringsfrågan. Detta är av allt att döma en anledning till varför ingen av dessa två teorier drar slutsatsen att förståelsen av en själv är sekundär till förståelsen av den andre.

5. TEORI-TEORIN OCH SIMULERINGSTEORIN

Som redan nämnts så finns det två övergripande modeller som försöker förklara hur barnets förståelse av falsk-tro-testscenariot och lösningen av uppgiften går till: den s.k. simuleringsteorin och teori-teorin. Enligt teori-teorin har ett litet barn (i analogi med teorier om den fysiska världen) en teori om hur andra människor tänker och upplever världen. För att förstå den andre applicerar barnet denna teori på den andres agerande och testar de förutsägelser som teorin medför. Teori-teori-teoretiker menar också att vår kunskap om oss själva och vår kunskap om andra utvecklas i tandem, och att mentaliseringen om oss själva och de andra stöds av samma slags teori. Teoretiker som förespråkar simuleringsförklaringen föreslår dock i stället en modell enligt vilken det som ligger till grund för mentaliseringen är vår förmåga att leva oss in i – simulera – den andres sätt att vara och känna. Vi använder sedan självkunskapen om det simulerade för att förstå oss på den andre. Inom simuleringsteorin har man också föreslagit olika varianter av hur simuleringsprocessen går till: med eller utan hjälp av introspektion. Både teori-teorin och simuleringsteorin är familjer av teorier och inkluderar en mängd olika varianter av hur förståelsen av den andra och en själv går till.

Teori-teorin och simuleringsteorin skiljer sig, som det framgår, med avseende på hur sambandet mellan mentaliseringen om en själv och mentaliseringen om den andre ska förstås. Antingen är mentaliseringen om en själv en färdighet som möjliggör och därför är primär i relation till förmågan att förstå sig på andra (simuleringsförklaringen), eller utvecklas förmågan att mentalisera om en själv parallellt med förmågan att mentalisera om den andre (teori-teori-teoretiker).

Både teori-teorins och simuleringsteorins förklaringsmodeller är, utifrån vad som sagts ovan, otillfredsställande vad gäller sambandet mellan de två typerna av förståelse. Detta först och främst eftersom de inte skiljer mellan förmågan att se ett objekt från den andra personens perspektiv och förmågan att se sig själv som objekt från detta perspektiv. Att se ett objekt från den andra personens perspektiv (falsk-tro-testet, mentaliseringen

om den andre) är inte på något sätt detsamma som att kunna se *sig själv* som ett objekt från detta perspektiv (självreferens, mentaliseringen om en själv). För att kunna uppnå det senare behöver dock barnet ha förmåga att lyckas med det förra. I detta avseende förefaller självkunskapen vara en mer avancerad kognitiv prestation än det som testas av falsk-tro-testet. Om förmågan till självreferens förutsätter barnets goda prestation i falsk-tro-testet så innebär det att självreferensen varken föregår (simulationsteorin) eller utvecklas parallellt med (teori-teorin) uppfattandet av den andre. Om detta resonemang är riktigt behöver dessa teoretiska konstruktioner – teori-teorin och simulationsteorin – ses över.

6. AVSLUTNING

”I reflektionen”, hävdade Sartre (2008, s. 379), ”omfattar jag den Andres perspektiv på min kropp; jag försöker uppfatta den som om jag var den Andre i förhållande till den”. Husserl talade om en fundamental förändring i inställningen till en själv som föranleds av den andre: ”Det är den andre som lär mig att uppfatta mig själv från ett tredjepersonsperspektiv” (Zahavi 2008, s. 94). Självreferens i denna betydelse är en förmåga att se sig själv som objekt från den andres perspektiv. Denna reflekterande och i det vardagliga livet ofta självklara inställning till en själv innebär dock en kognitiv ansats och prestation. Den självrefererande kapaciteten har inte alltid varit närvarande under vår utvecklings gång.

Inom barnpsykologisk forskning utgår man gärna från att barn som klarar ett väletablerat test, det s.k. falsk-tro-testet, redan besitter förmågan till självreferens (simulationsteoretikerna). Man har alternativt argumenterat för att förmågan till självreferens utvecklas i tandem med vår förmåga att förstå de andra (teori-teori-teoretiker). Detta är dock inte alls så självklart. I denna text uppmärksammade jag en viktig konsekvens av falsk-tro-testet, nämligen att ett barn som inte klarar testet inte heller skiljer mellan sig själv och den andre. Barnets framgång i falsk-tro-test är ett viktigt tecken på att barnet kan uppfatta den andre som en självständig (från barnet skild) tänkande individ. Mentaliseringen om den andre (lyckad utgång av falsk-tro-test) blir därmed en viktig förutsättning för att kunna ta denne andras perspektiv på en själv och uppnå självreferens (mentalisera om en själv). I motsats till vad som påstås av både teori-teorin och simuleringsteorin är således självförståelsen sekundär i relation till förståelsen av den andre. Men självförståelsen följer inte automatiskt av förståelsen av den andre.

I denna artikel föreslår jag en ny tolkning av en mycket använd experimentell procedur – falsk-tro-testet – som anses testa barns förmåga att mentalisera. Den nya tolkningen bygger på en viktig observation: de barn som inte klarar testet och tillskriver andra sina egna mentala tillstånd särskiljer inte riktigt sig själva från människor runt omkring dem.

Konsekvensen att de barn som inte klarar det klassiska falsk-tro-testet befinner sig i ett slags odifferentierat själv-andra-tillstånd har inte fått någon större uppmärksamhet i litteraturen på området. Denna konsekvens är dock av en avgörande betydelse för hur vi ska förstå oss på den implicita mentaliseringen, det vill säga, den mentalisering som anses ske före barnet klarar falsk-tro-testet. Slutsatsen av diskussionen i denna uppsats är att barnet inte har en förmåga till mentalisering innan den kognitiva mognad uppnåtts som signaleras med framgång i falsk-tro-testet (vid cirka 4,5–5 års ålder). Vill man framhålla att små barn ändå implicit mentaliserar så är det viktigt att förklara på vilket sätt det implicit mentaliserande barnet uppfattar den andre som den *andre*. Utan tydlighet på denna punkt har man inte uteslutit att mentaliseringen om den andre sammanfaller med mentaliseringen om en själv och vice versa.

LITTERATUR

Abu-Akel, Ahmad. 2003. ”A Neurobiological Mapping of Theory of Mind”. *Brain Research Reviews* 43, nr 1, s. 29–40.

Frith, Uta och Christopher D. Frith. 2003. ”Development and Neurophysiology of Mentalizing”. *Phil. Trans. R. Soc. Lond. B*. 358, s. 459–73.

Leslie, Alan M. 1987. ”Pretence and Representation: The Origins of ’Theory of Mind’”. *Psychological Review* 94, s. 412–26.

Piaget, Jean. 1959. *The Language and Thought of the Child*. Tredje uppl. London: Routledge & Kegan Paul.

Premack, David och Guy Woodruff. 1978. ”Does the Chimpanzee Have a ’Theory of Mind’?”. *Behavioral and Brain Sciences* 4, s. 515–26.

Sartre, Jean-Paul. 2008. *Being and Nothingness: An Essay on Phenomenological Ontology*. London: Routledge.

Wellman, Henry M., David Cross och Julanne Watson. 2001. ”Meta-Analysis of Theory of Mind Development: The Truth about False Belief”. *Child Development* 72, nr 3, s. 655–84.

Wimmer, Heinz och Josef Perner. 1983. ”Beliefs about Beliefs – Representation and Constraining Function of Wrong Beliefs in Young Children’s Understanding of Deception”. *Cognition* 13, nr 1, s. 103–28.

Zahavi, Dan. 2008. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass.: MIT Press.

PAUL GORBOW

ESSENS, ATT SAMMANFALLA OCH KARAKTÄRISERING

1. INLEDNING

Låt oss föreställa oss en metallsked. Metallskeden kan också ses som en metallbit. Men är Metallskeden identisk med Metallbiten? Utifrån synvinkeln att ting har essens, som karaktäriserar dem, är det naturligt att särskilja mellan Metallskeden och Metallbiten. Det tillhör Metallskedens essens att ha en viss typ av form, och denna egenskap tillhör inte Metallbitens essens. Om Metallskeden existerar, då måste den ha denna form, men Metallbiten kan existera utan att ha denna form. Utifrån denna essentialistiska synvinkel har alltså Metallbiten och Metallskeden olika egenskaper, så enligt lagen om identiska tings oskiljaktighet kan vi dra slutsatsen att skeden inte är identisk med Metallbiten.

Om nu Metallskeden och Metallbiten inte är samma ting, så står de ändå i en intim relation till varandra. Jag ska använda uttrycket att Metallskeden och Metallbiten *sammanfaller*. Historiskt har även begreppet *kontingent identitet* använts för samma relation, och jag kommer använda dessa båda uttryck som synonyma. I denna artikel ska vi gå igenom Yablos formella system för att modellera essens och kontingent identitet, så som det framställts i dennes (1987). Jag kommer att härleda nya satser från systemet, som jag menar är oönskade vid modellering av essens, samt föreslå en modifikation av systemet som undviker dessa konsekvenser.

Bakgrunden och motivet till Yablos system presenteras i avsnitt 2, i termer av en problematisering av kontingent identitet som lyfts fram av Kripke. Yablo accepterar Kripkes slutsats, men försöker visa att vi ändå kan ha en koherent teori om kontingent identitet. Här redogörs också för Yablos syn på essens.

Yablos formella system presenteras i avsnitt 3. Först ser Yablo till att hans modell hanterar egenskaper på ett sätt som stämmer överens med en grundläggande intuition om essens. I modellen är ett tings essens den mängd egenskaper som tinget nödvändigtvis har. Yablo låter sig guidas av idén att Metallskedens essens inkluderar och är större än Metallbitens essens. Det krävs mer för att vara Metallskeden än för att

Filosofisk tidskrift 2017 nr 3, 43–57

vara Metallbiten. Men här uppstår ett problem med vissa egenskaper, till exempel egenskapen ”att vara identisk med Metallbiten”. Om denna egenskap tillhörde Metallbitens essens, så skulle inte skedens essens kunna inkludera Metallbitens essens. Därför inför Yablo begreppet *kumulativ egenskap*, med vilket han avser egenskaper som kan tillhöra en essens utan att hindra andra egenskaper från att tillhöra essensen. Yablo designar sitt system så att endast kumulativa egenskaper kan ingå i ett tings essens.

I avsnitt 4 genomförs en kritisk granskning av Yablos system. Jag härleder en sats som jag menar bör betraktas som icke önskvärd. Det visar sig exempelvis att i modeller av verkligheten vari essentiella egenskaper är intrinsikala, så tvingar Yablos system samtliga ting att vara kontingent identiska med varandra. Mer generellt, så menar jag att systemet lider av en svårkontrollerad instabilitet: modelleringen av kontingent identitet kollapsar fullständigt såvida inte ett svårtolkat kriterium om vissa extremt extrinsikala egenskaper uppfylls.

I avsnitt 5 föreslår jag att vi ersätter ett av Yablos axiom med ett nytt axiom, vilket kopplar essens till karaktärisering. Syftet med denna modifikation är att undvika kollapsen som påvisats i avsnitt 4, samtidigt som idén med det gamla axiomet vidhålls i det nya axiomet.

2. BAKGRUND

En hörnsten till det system Yablo utvecklar i sin (1987) är argumentet nedan för att sanna identitetspåståenden är nödvändigt sanna. Argumentet formulerades först av Kripke (1971, s. 136), och utgår från att vi accepterar två konventionella lagar om identitet: dels att varje ting är identiskt med sig självt, dels att identiska ting är oskiljaktiga.

1. Anta att $\alpha = \beta$.
2. β har egenskapen att nödvändigtvis vara identisk med β.
3. Enligt 1, 2 och lagen om identiska tings oskiljaktighet, så har α egenskapen att nödvändigtvis vara identisk med β. Det vill säga, det är nödvändigt sant att $\alpha = \beta$.

Detta innebär att det inte finns något identitetspåstående som endast är kontingent sant. Yablo accepterar Kripkes argument. Därför försöker han ge en teori för den relation som exemplifierades med Metallskeden och Metallbiten i inledningen, utan att motsäga Kripkes argument. Denna relation hade tidigare kallats för kontingent identitet. Kanske

är det därför som Yablo bibehåller detta namn på relationen, trots att terminologin går stick i stäv med Kripkes insikt. Så här formulerar Yablo syftet med sin teori:

> Syftet med detta papper är att förklara, för det första, varför kontingent identitet erfordras av essentialism och, för det andra, hur kontingent identitet tillåts av essentialism. (Yablo 1987, s. 294)

Det är Yablos andra fråga, alltså ”hur”-frågan, som besvaras med hjälp av ett formellt system. Han vill visa hur kontingent identitet kan tillåtas inom ramen för essentialism. Yablo förklarar inte explicit vad han menar med essentialism, men det framgår att han åtminstone inbegriper ståndpunkten att det finns många fysiska ting, händelser, med mera, vilka har en essens som karaktäriserar vad det tinget är (1987, s. 294–300).

För att ge ett exempel med händelser, tänk dig en joggingtur som sammanfaller med en naturvistelse. Då har vi både att Naturvistelsen kunde ha existerat utan att Joggingturen hade det, och vice versa att Joggingturen kunde ha existerat utan att Naturvistelsen hade det. Från essentialistens synvinkel tillhör det Joggingturens, men inte Naturupplevelsens, essens att den sker springande. Därmed skiljer sig dessa två händelser från varandra vad gäller egenskapen att nödvändigtvis ske springande. Av lagen om identiska tings oskiljaktighet följer nu att Joggingturen och Naturvistelsen inte är identiska. Eftersom det är uppenbart att Naturvistelsen och Joggingturen ändå i någon mening är samma händelse, menar Yablo att det behövs en ny identitetsliknande relation som förklarar detta (1987, s. 295). Han använder, kanske av hänsyn till traditionen, uttrycket ”kontingent identitet” för denna nya relation, som alltså är distinkt från identitetsrelationen.

Det är i sin (1987, s. 296–98) som Yablo introducerar sin syn på essens. Mer exakt menar Yablo att essenser bör uppfylla två funktioner. Dels att ett tings essens ska vara en mängd egenskaper i kraft av vilka det är tinget i fråga. Dels att essensen ska vara ett mått på vad som krävs för att vara tinget i fråga. Den första funktionen innebär för Yablo att en egenskap som ”att vara identisk med Metallbiten”, inte bör inbegripas i Metallbitens essens. För det vore en trivialisering att påstå att Metallbiten är Metallbiten i kraft av att den är identisk med Metallbiten. Den andra funktionen exemplifieras för Yablo av att skedens essens är större än Metallbitens, vilket kan ses som ett mått på att det krävs mer för att vara skeden än för att vara Metallbiten.

En viktig konsekvens av Yablos system är att om ett tings essens inkluderar ett annat tings essens, och det första tinget existerar i en möjlig värld, då existerar även det andra tinget i den världen, och de två tingen sammanfaller med varandra i den världen. Det är alltså inte bara så att Yablo modellerar essens och kontingent identitet i samma system, begreppen relateras också till varandra på ett intressant sätt.

3. YABLOS SYSTEM FÖR KONTINGENT IDENTITET

Yablo gestaltar sina idéer om kontingent identitet i ett formellt system. Detta avsnitt syftar till att presentera och utförligt förklara de definitioner och härledningar som systemet utgörs av, i enlighet med specifikationen i (Yablo 1987, s. 300–3, 310–11). Systemet kan ses som en modifikation av den vanliga mängdteoretiska semantiken för modallogiken, och existensmodeller utgör grundstommen i systemet. För att definiera dem, låt Pow(A) beteckna mängden av alla delmängder av A, det vill säga Pow(A) = $\{X \mid X \subseteq A\}$.

Definition 1. En *existensmodell*, (**W**,**O**,**D**), består av

1. en mängd *världar*, **W**;
2. en mängd *ting*, **D**, som kallas för *diskursdomänen*;
3. en *ontologisk funktion*, **O**: **W** → Pow(**D**), sådan att för varje värld W är **O**(W) mängden av alla ting som *existerar* i W, och som uppfyller att för varje $\alpha \in$**D**, så finns det en värld $W \in$**W**, sådan att $\alpha \in$**O**(W).

Ta exemplet med Metallskeden. Metallskeden är ett ting som ska ingå i diskursdomänen **D**, skrivet Metallskeden ∈ **D**.[1] Låt säga att F är en egenskap, att vara formad på ett visst vis, som tillhör Metallskedens essens. Då ska Metallskeden ha egenskapen F i alla världar där den existerar. Men Yablo vill att uttrycket ”Metallskeden har nödvändigtvis egenskapen F” ska vara synonymt med ”Metallskeden har essentiellt egenskapen F”. Yablo tänker sig med andra ord att Metallskeden har egenskapen F i alla världar om, och endast om, den har egenskapen F i alla världar där den existerar. För att få detta att gå ihop särskiljer Yablo först mellan begreppen *attribut* och *egenskap* enligt följande definition:

Definition 2. Låt (**W**,**O**,**D**) vara en existensmodell. P är ett *attribut* (med avseende på en existensmodell (**W**,**O**,**D**)) om P är en funktion P: W → Pow(**D**).

1. Rent formellt behöver förstås inte **D** vara en mängd av fysiska ting, utan kan lika gärna bestå av formella objekt som representerar fysiska ting.

Om P är ett attribut, $\alpha \in \mathbf{D}$, $\mathbf{W} \in \mathbf{W}$ och $\alpha \in P(\mathbf{W})$, då säger vi att α *har* P i $\mathbf{W}$. P är en *egenskap* om P är ett attribut, sådant att om α har P i alla världar vari α existerar, då har α P i alla världar.

Låt mig illustrera definitionen med hjälp av Joggingturen, från exemplet ovan, som essentiellt sker springande. Om P är attributet som representerar *att ske springande*, då har vi att Joggingturen $\in P(V)$ för varje värld V vari Joggingturen existerar. Om P dessutom är en egenskap, då följer att Joggingturen $\in P(V)$ för varje värld V, alltså även de världar vari Joggingturen inte existerar. Ett typiskt attribut som i regel inte är en egenskap, enligt denna definition, är att existera.

En direkt konsekvens av definitionen är att identiska ting har samma egenskaper och samma attribut. Det följer också att modalt ekvivalenta egenskaper är identiska. Att vara identisk med sig själv, och att vara sådan att $2 + 2 = 4$, är därmed samma egenskap. Alla ting har ju båda dessa egenskaper, så de modelleras båda av funktionen som tar det konstanta värdet **D** (dvs. hela diskursdomänen) för varje värld.

Det ska visa sig i vad som följer (se bl.a. Definitionerna 6 och 8) att Yablo behöver kunna begränsa sig till endast en viss typ av egenskaper. Därför inför han en till parameter som kan användas för att specificera en begränsad mängd egenskaper:

Definition 3. Låt **X** vara en mängd egenskaper med avseende på en existensmodell (**W**,**O**,**D**). Då kallar vi (**W**,**O**,**D**,**X**) för en *egenskapsmodell*.

Fixering 4. *Fixera en egenskapsmodell* $\mathbf{\Omega} = (\mathbf{W},\mathbf{O},\mathbf{D},\mathbf{X})$, *för resten av denna artikel.*

För att underlätta framställningen av systemet, inför Yablo förkortande notation. Till exempel, om P är ett attribut så står $P^{\Box}$ för attributet ”nödvändigtvis P”, därav box-notationen som traditionellt hör samman med nödvändighet.

Notation 5. Låt P vara ett attribut, låt $\mathbf{Y}$ vara en mängd attribut och låt $W \in \mathbf{W}$.

1. $P^{\Box}: = \{\alpha \in \mathbf{D}: (\forall W \in \mathbf{W})(\alpha \in P(W))\}$.
2. $\mathbf{Y}(W): = \{\alpha \in \mathbf{D}: (\forall P \in \mathbf{Y})(\alpha \in P(W))\}$.
3. $\mathbf{Y}[W]: = \mathbf{Y}(W) \cap \mathbf{O}(W)$.
4. $\mathbf{Y}^{\Box}: = \{\alpha \in \mathbf{D}: (\forall P \in \mathbf{Y})(\forall W \in \mathbf{W})(\alpha \in P(W))\}$.
5. $\mathbf{Y}^{\Box}[W]: = \mathbf{Y}^{\Box} \cap \mathbf{O}(W)$.

Nu kommer vi till den centrala definitionen av essens.

Definition 6. Låt $\alpha \in \mathbf{D}$. α:s *essens*, $\mathbf{E}(\alpha)$, är mängden av alla egenskaper $P \in \mathbf{X}$, som α har i alla världar vari α existerar.

I denna definition används parametern **X** som infördes i Definition 3. **X** är mängden av de egenskaper som tillåts att ingå i essenser. Yablo ger aldrig någon formell definition som stipulerar vilka egenskaper som ska ingå i **X**, men Definition 8 kan ses som en stipulation av hur egenskaperna i **X** ska bete sig i förhållande till egenskapsmodellen i stort.

Följande sats visar att ett tings essens består av precis de egenskaper som tinget nödvändigtvis har.

Sats 7. Låt $\alpha \in \mathbf{D}$ *och låt* $P \in \mathbf{X}$. $P \in \mathbf{E}(\alpha)$ *om och endast om* $\alpha \in P^{\square}$.

Bevis. Vi börjar med riktningen $\Leftarrow$. Anta att $\alpha \in P^{\square}$. Då är $\alpha \in P(W)$ för varje $W \in \mathbf{W}$. Eftersom $\alpha \in P(W)$ gäller i alla världar, så gäller det i de världar vari α existerar. Enligt Definition 6 har vi alltså att $P \in \mathbf{E}(\alpha)$, som önskat.

För riktningen $\Rightarrow$, anta att $P \in \mathbf{E}(\alpha)$. Då är $\alpha \in P(W)$, för varje $W \in \mathbf{W}$ vari α existerar. Eftersom P är en egenskap, så får vi enligt Definition 2 att $\alpha \in P(W)$ för varje $W \in \mathbf{W}$. Det vill säga, $\alpha \in P^{\square}$. QED.

Beviset av denna sats (riktningen $\Rightarrow$) är beroende av Definition 2 ovan. Egenskaper definierades så att om ett ting har en viss egenskap i alla världar där tinget existerar (dvs. om egenskapen ingår i essensen), då har tinget egenskapen i alla världar över huvud taget (dvs. den har egenskapen med nödvändighet).

Låt oss konstruera en enkel egenskapsmodell av ett exempel som Yablo använder genomgående. Skruden i Turin är den skrud i Turin som sägs ha tjänat som Jesus begravningsskrud. Tyget i Turin är det tyg som rent fysiskt utgör denna skrud. För att vara Tyget i Turin krävs det inte att ha tjänat som Jesus begravningsskrud, men Skruden i Turin existerar inte om inte Tyget i Turin faktiskt tjänat som Jesus begravningsskrud. Här är en egenskapsmodell som fångar detta metafysiska förhållande:

$\mathbf{\Omega}_T = (\mathrm{W}_T, \mathrm{O}_T, \mathbf{D}_T, \mathbf{X}_T)$
$\mathbf{W}_T = \{\mathrm{V}, \mathrm{W}\}$
$\mathbf{D}_T = \{\text{Skruden, Tyget}\}$
$\mathbf{O}_T(V) = \{\text{Skruden, Tyget}\}$
$\mathbf{O}_T(W) = \{\text{Tyget}\}$

$\mathbf{X}_T = \{J\}$
$J(V) = \{\text{Skruden, Tyget}\}$
$J(W) = \{\text{Skruden}\}$

Egenskapen ”att ha tjänat som Jesus begravningsskrud” representeras här av *J*. Det följer av definitionerna att E(Skruden) = {*J*} och att **E**(Tyget) = ∅, det vill säga att *J* tillhör Skrudens essens, men inte Tygets essens. Det finns en intuition här att kraven för att vara Skruden i Turin är entydigt striktare än kraven för att vara Tyget i Turin, som speglas i att Skrudens essens inkluderar Tygets essens som en delmängd. Men så blir inte fallet om vi utökar **X** till att även innehålla egenskapen *K*, ”att vara identisk med Tyget” (dvs. den konstanta funktionen med värde {Tyget}), för *K* ingår då i Tygets essens men ingår inte i Skrudens essens. De funktioner som Yablo anser att essens bör fylla – att ange de egenskaper i kraft av vilka tinget är tinget i fråga och att vara ett mått på vad som krävs för att vara tinget i fråga (se Avsnitt 2) – får honom att vilja begränsa essenser till att endast inkludera egenskaper som inte blockerar andra egenskaper på det här sättet, och dessa egenskaper kallar han kumulativa (Yablo, 1987, s. 299):

> … det finns egenskaper som endast kan ”bygga upp” essenserna i vilka de figurerar. Eftersom att inkludera sådana egenskaper i en essens inte är (förutom trivialt) att hålla någon annan egenskap ute, så kommer de att kallas *kumulativa*. (1987, s. 299)

Yablo uttrycker relationen att Skrudens essens inkluderar Tygets essens, som att Skruden *förädlar* Tyget, och han försöker fånga kumulativiteten formellt genom att kräva av modellen $\mathbf{\Omega} = (\mathbf{W},\mathbf{O},\mathbf{D},\mathbf{X})$ att den är *sluten*, enligt följande definition.[2]

Definition 8. Låt $\alpha,\beta \in \mathbf{D}$. *β förädlar α*, skrivet $\beta \geq \alpha$, om $\mathbf{E}(\beta) \supseteq \mathbf{E}(\alpha)$.[3]
Ω är *sluten uppåt* om för varje $\alpha \in \mathbf{D}$, $\mathbf{Y} \subseteq \mathbf{X}$ och $W \in \mathbf{W}$,
$\alpha \in \mathbf{Y}[W] \rightarrow (\exists \beta \geq \alpha)(\beta \in \mathbf{Y}^{\square}[W])$.
Ω är *sluten nedåt* om för varje $\alpha \in \mathbf{D}$, $\mathbf{Y} \subseteq \mathbf{X}$ och $W \in \mathbf{W}$,
$(\exists \beta \geq \alpha)(\beta \in \mathbf{Y}^{\square}[W]) \rightarrow \alpha \in \mathbf{Y}[W]$.
Ω är *sluten* om den är sluten uppåt och nedåt.

2. För att underlätta framställningen har jag avvikit marginellt från Yablos definition. Den intresserade läsaren kan lätt verifiera att min definition är ekvivalent med den som Yablo ger (1987, s. 302–3).

3. Notera att förädlingsrelationen är reflexiv, antisymmetrisk och transitiv (eftersom delmängdsrelationen är det).

I naturligt språk: Att $\mathbf{\Omega}$ är sluten uppåt innebär att om ett ting α existerar och har vissa egenskaper $\mathbf{Y} \subseteq \mathbf{X}$ i en värld W, då existerar det också ett ting β i W, som har egenskaperna $\mathbf{Y}$ essentiellt. Att $\mathbf{\Omega}$ är sluten nedåt innebär att om ett ting α förädlas av ett ting β, som existerar i en värld W och som har vissa egenskaper $\mathbf{Y} \subseteq \mathbf{X}$ essentiellt, då existerar även α i W och α har egenskaperna $\mathbf{Y}$ i W.

Givet att $\mathbf{\Omega}$ är sluten, så definierar Yablo en *kumulativ* egenskap, P, som en egenskap sådan att den är medlem i $\mathbf{X}$. $\mathbf{X}$ är alltså mängden av alla kumulativa egenskaper. Poängen är att om vi kräver att $\mathbf{\Omega}$ är sluten, då hindrar vi i viss mån[4] icke-kumulativa egenskaper som ”att vara identisk med Tyget” från att ingå i $\mathbf{X}$, enligt intuitionen som presenterats ovan. Låt oss titta närmare på hur detta hänger ihop med slutenhet. Anta att β existerar i en värld W, och att β förädlar α. Då följer av slutenhet nedåt att även α existerar i W. Detta speglar att det krävs mindre för att vara α än för att vara β.

För att förstå hur kravet om slutenhet uppåt är kopplat till kumulativitet, betrakta den enkla modellen $\mathbf{\Omega}_T$ ovan (om Tyget och Skruden i Turin). Som sagt anser Yablo att vissa egenskaper, t.ex. ”att vara identisk med Tyget”, inte är kumulativa och därför inte bör ingå i Tygets essens. Det skulle nämligen förstöra förädlingsrelationen, vilken syftar till att fånga intuitionen att ett tings essens är ett mått på vad som krävs för att vara tinget i fråga. Att Skruden förädlar Tyget, dvs. att det krävs mer för att vara Skruden än för att vara Tyget, speglas av att $\mathbf{E}$(Skruden) innehåller $\mathbf{E}$(Tyget) som en delmängd. Om egenskapen ”att vara identisk med Tyget” ingick i $\mathbf{E}$(Tyget) så skulle inte detta vara fallet, eftersom Skruden inte är identisk med Tyget. Notera att egenskapsmodellen $\mathbf{\Omega}_T$ är sluten. Men vad händer om vi lägger till egenskapen K, ”att vara identisk med Tyget”, i $\mathbf{X}_T$? Svar: I världen V har då Tyget båda egenskaperna J och K. Men eftersom ingenting i $\mathbf{\Omega}_T$ har båda dessa egenskaper essentiellt, så vore inte denna egenskapsmodell sluten uppåt. I denna mening hindrar slutenheten K från att ”vara med i modellen”.

Med hjälp av slutenhet kan vi nu bevisa en filosofiskt intressant sats. För att formulera satsen så introducerar Yablo begreppet *profil*: α:s profil, i en viss värld W, är mängden $\mathbf{P}_W(\alpha)$ av kumulativa egenskaper som tinget har i den världen.

4. Låt $\alpha \in \mathbf{D}$ vara ett ting. Slutenhet hindrar inte helt och hållet att $\mathbf{X}$ kan inkludera egenskapen A, ”att vara identisk med α”. Ett trivialt exempel: om α är det enda tinget i diskursdomänen och A är den enda egenskapen i $\mathbf{X}$, så är egenskapsmodellen ändå sluten. Så A är faktiskt kumulativ i vissa slutna egenskapsmodeller. Enligt systemet, så beror alltså en egenskaps kumulativitet på hur egenskapsmodellen ser ut i stort. En egenskap är inte kumulativ, eller icke-kumulativ, i sig.

Sats 9. *Anta att* $\mathbf{\Omega}$ *är sluten och att* $\beta \geq \alpha$. *Låt* $W \in \mathrm{W}$.

1. *Om* β *existerar i* W, *då existerar* α *i* W.

2. *Om* β *existerar i* W, *då är* $\mathbf{P}_W(\beta) = \mathbf{P}_W(\alpha)$.

Bevis. 1. Genom att sätta $\mathbf{Y} = \varnothing$ i definitionen av slutenhet nedåt, får vi att $\alpha \in \varnothing[W]$, ur vilket följer att α existerar i W.

2. Anta att β existerar i W; då existerar även α i W, enligt del 1. Eftersom $\mathbf{\Omega}$ är sluten uppåt, så existerar det $\alpha^\circ \geq \alpha$ och $\beta^\circ \geq \beta$ i W vars essenser inkluderar $\mathbf{P}_W(\alpha)$ respektive $\mathbf{P}_W(\beta)$: sätt till exempel $\mathbf{Y} = \mathbf{P}_W(\alpha)$ i definitionen av slutenhet uppåt, för att få fram α°. Transitiviteten av $\geq$ ger att $\beta^\circ \geq \alpha$, så eftersom $\mathbf{\Omega}$ är sluten nedåt, får vi att $\mathbf{E}(\beta) \subseteq \mathbf{P}_W(\beta) \subseteq \mathbf{E}(\beta^\circ) \subseteq \mathbf{P}_W(\alpha) \subseteq \mathbf{E}(\alpha^\circ)$. Det innebär att $\alpha^\circ \geq \beta$, så eftersom $\mathbf{\Omega}$ är sluten nedåt, får vi till och med att $\mathbf{E}(\beta) \subseteq \mathbf{P}_W(\beta) \subseteq \mathbf{E}(\beta^\circ) \subseteq \mathbf{P}_W(\alpha) \subseteq \mathbf{E}(\alpha^\circ) \subseteq \mathbf{P}_W(\beta)$. Detta innebär att $\mathbf{P}_W(\beta) = \mathbf{P}_W(\alpha)$, som önskat. QED.

Ett nyckelsteg i beviset av del 2 är att slutenhet uppåt ger att ett existerande ting α kan förädlas så pass långt att alla dess kumulativa egenskaper är essentiella egenskaper hos förädlingen α°. Detta tillämpas på både α och β. Slutenhet nedåt ger sedan genom ett par steg både att α:s profil inkluderar β:s och att β:s profil inkluderar α:s. Alltså har de samma kumulativa egenskaper.

α och β har samma essens om och endast om $\alpha \geq \beta$ och $\beta \geq \alpha$. Så av satsen följer att om α och β har samma essens, då existerar de i samma världar och har samma kumulativa egenskaper där de existerar. Systemet fångar på så vis den metafysiska ståndpunkten att ett tings essens karaktäriserar vad tinget är. Men satsen säger mer, nämligen att det räcker att β:s essens inkluderar α:s essens för att α ska existera och ha samma kumulativa egenskaper som β i alla världar där β existerar.

Yablo använder sig av distinktionen mellan hypotetiska och kategoriska egenskaper (1987, s. 305). Jag förstår honom som att egenskapen att förångas vid 100 °C är en egenskap vattnet i glaset har hypotetiskt, medan egenskapen att bestå av H_2O-molekyler är en egenskap det har kategoriskt. Den hypotetiska egenskapen är en benägenhet hos vattnet som skulle aktualiseras under vissa omständigheter, medan den kategoriska egenskapen är en egenskap vattnet har aktuellt. Yablo försöker ge en formell definition av begreppet *kategorisk* inom ramen för sitt system:[5]

5. Det är inte uppenbart för mig att denna definition överensstämmer med den gängse bilden av kategoriska egenskaper, och jag hänvisar läsaren till Yablos förklaring i dennes (1987, s. 303–10).

Definition 10. Anta att **Ω** är sluten. Låt P vara en egenskap. P är en *kategorisk* egenskap om för varje värld W och för varje α, β som existerar i W, $\alpha \leq \beta \rightarrow (\alpha \in P(W) \leftrightarrow \beta \in P(W))$.

Kategoriska egenskaper är alltså sådana egenskaper som inte kan skilja mellan två ting som existerar i samma värld och där det ena förädlar det andra. Det följer av Sats 9 del 2 att varje kumulativ egenskap är kategorisk. Men notera att det kan finnas kategoriska egenskaper även utanför **X**. Med hjälp av detta begrepp definierar Yablo kontingent identitet:

Definition 11. Anta att **Ω** är *sluten. Låt* $W \in \mathbf{W}$ och låt $\alpha, \beta \in \mathbf{O}(W)$. α och β *sammanfaller* eller är *kontingent identiska i* W, skrivet $\alpha \approx_W \beta$, om de har samma kategoriska egenskaper i W.[6]

Sammanfattningsvis kan Yablos system betraktas som en variant av den mängdteoretiska semantiken för modallogiken. Förutom en diskursdomän **D** av ting, en mängd världar **W**, och en funktion **O** som beskriver vilka ting som existerar i vilka världar, så består varje egenskapsmodell också av en mängd egenskaper **X**. Ett tings essens är mängden egenskaper som tinget har i alla världar där det existerar. Men Yablo anser att essens även ska fungera som ett mått på vad som krävs för att vara ett visst ting. Att det krävs mer för att vara Skruden i Turin än att vara Tyget i Turin förklaras av att Skrudens essens inkluderar Tygets essens som en delmängd. Yablo uttrycker detta som att Skruden förädlar Tyget. För att essens ska fungera som ett mått på det här sättet, inför Yablo kravet att egenskapsmodellen ska vara sluten. Om egenskapsmodellen är sluten, så kan egenskaperna i **X** beskrivas som kumulativa, och kontingent identitet kan definieras. En annan viktig konsekvens av slutenhet är Sats 9 ovan, som innebär att två existerande ting, sådana att det ena förädlar det andra, har samma kumulativa egenskaper. Yablo skiljer mellan kategoriska och hypotetiska egenskaper, och han använder även denna distinktion i sin definition av kontingent identitet. Men jag ska strax ge en betydligt enklare definition av kontingent identitet som tycks ha undgått Yablo.

4. KOLLAPS AV KONTINGENT IDENTITET

I detta avsnitt genomförs en kritisk granskning av Yablos formella sys-

6. Notera att $\approx_W$ är en reflexiv, symmetrisk och transitiv relation, det vill säga en ekvivalensrelation.

tem. Kritiken baseras på nya satser som jag härlett från Yablos definitioner. Kärnan i kritiken är Sats 15, vilken visar att kontingent identitet har en stark benägenhet att kollapsa. Vi ska börja med att härleda att Yablos definition av kontingent identitet är ekvivalent med en betydligt enklare definition.

Sats 12. *Anta att* $\mathbf{\Omega}$ *är sluten. Låt* $W \in \mathbf{W}$ *och låt* $\alpha, \beta \in \mathbf{O}(W)$. *α och β är kontingent identiska i W om och endast om de har samma kumulativa egenskaper.*

Bevis. Vi behöver visa att $\alpha \approx_W \beta$ om och endast om $\mathbf{P}_W(\alpha) = \mathbf{P}_W(\beta)$. Anta att $\alpha \approx_W \beta$. Eftersom kumulativa egenskaper är kategoriska, har vi att $\mathbf{P}_W(\alpha) = \mathbf{P}_W(\beta)$. Omvänt, anta att $\mathbf{P}_W(\alpha) = \mathbf{P}_W(\beta)$ och låt P vara en kategorisk egenskap. Tack vare slutenhet uppåt så existerar det γ i W som uppfyller $\mathbf{E}(\gamma) \supseteq \mathbf{P}_W(\alpha) = \mathbf{P}_W(\beta)$. Det vill säga, $\gamma \geq \alpha, \beta$. Så eftersom P är kategorisk, har vi att $\alpha \in P(W)$ om och endast om $\gamma \in P(W)$, och att $\beta \in P(W)$ om och endast om $\gamma \in P(W)$. Således är $\alpha \in P(W)$ om och endast om $\beta \in P(W)$. Eftersom P är en godtycklig kategorisk egenskap, så följer det att $\alpha \approx_W \beta$. QED.

För riktningen $\Leftarrow$ i detta bevis används en liknande teknik som i beviset av Sats 9. Från antagandet att α och β har samma kumulativa egenskaper härleds att de båda har en gemensam förädling γ, vilken har alla dessa kumulativa egenskaper essentiellt. Från Yablos definition av kategorisk egenskap följer därefter i ett par steg att α och β har samma kategoriska egenskaper, vilket i sin tur innebär att de sammanfaller. Vi kan alltså lika gärna definiera kontingent identitet som ”att ha samma kumulativa egenskaper”, vilket är enklare än Yablos Definition 11 ovan, och undviker behovet att involvera distinktionen mellan kategoriska och hypotetiska egenskaper.

Av Satserna 9 och 12 följer nu ett korollarium som belyser hur förädling förhåller sig till kontingent identitet:

Korollarium 13. *Anta att* $\mathbf{\Omega}$ *är sluten. Om β förädlar α, och β existerar i världen W, så existerar även α i W, och* $\alpha \approx_W \beta$.

Bevis. Enligt Sats 9, så existerar α i W och $\mathbf{P}_W(\alpha) = \mathbf{P}_W(\beta)$. Nu följer det av Sats 12 att $\alpha \approx_W \beta$. QED.

Låt oss nu återigen undersöka vilken filosofisk konsekvens som följer om α och β har samma essens. Detta är ekvivalent med att $\alpha \geq \beta$ och $\beta \geq \alpha$. Alltså: Om α och β har samma essens, så existerar de i samma världar och är kontingent identiska i de världarna. Essens bestämmer alltså tingen upp till kontingent identitet. Detta preciserar ytterligare ståndpunkten att ett tings essens karaktäriserar vad tinget är.

Yablo intresserar sig särskilt för en sorts slutna egenskapsmodeller som han kallar fulla (1987, s. 310–11):

Definition 14. Anta att $\mathbf{\Omega}$ är sluten. $\mathbf{\Omega}$ *är full* om för varje mängd världar $\mathbf{V} \subseteq \mathbf{W}$, och för varje funktion f: $\mathbf{V} \rightarrow \mathbf{D}$, det finns $\alpha \in \mathbf{D}$, så att $\alpha \in \mathbf{O}(W) \leftrightarrow W \in \mathbf{V}$ och $W \in \mathbf{V} \rightarrow \alpha \approx_W f(W)$.

Fullhet innebär informellt att om man godtyckligt ”plockar” ett ting från var och en av en uppsättning världar, så finns det ett ting α som existerar i precis dessa världar och som sammanfaller med precis de tingen i dessa världar. Så om det finns en fotboll i W och en gräsmatta i W', då finns det ett ting som sammanfaller med fotbollen i W, men med gräsmattan i W'. Som jag läser Yablo (1987, s. 307), så medger han att fullhet innebär existensen av många kontraintuitiva ting, men att det ändå är försvarbart utifrån att metafysik söker förstå verkligheten i sig, oberoende av observatörernas förhållande till koncepten. Men följande sats, som jag bevisat, innebär att fullhet har drastiska konsekvenser för kontingent identitet.

Sats 15. *Anta att* $\mathbf{\Omega}$ *är sluten och full. Anta vidare att något ting* α *existerar i två olika världar* $W' \neq W$ *vari det har samma kumulativa egenskaper. Då sammanfaller alla ting som existerar i* W (*och därmed även i* W' *av symmetri*).

Bevis. Anta att β existerar i W. Antagandet i satsen säger att det existerar α i W, sådant att α även existerar i en annan värld $W' \neq W$ och att $\mathbf{P}_W(\alpha) = \mathbf{P}_{W'}(\alpha)$. Eftersom β är godtycklig så räcker det att bevisa att $\alpha \approx_W \beta$. Det följer av fullhet, att det finns ett ting $\gamma \in \mathbf{D}$ som existerar i W och W', samt uppfyller att $\gamma \approx_W \beta$ och att $\gamma \approx_{W'} \alpha$. Eftersom α har samma kumulativa egenskaper i W som i W', så följer nu att $\mathbf{E}(\gamma) \subseteq \mathbf{P}_W(\beta) \cap \mathbf{P}_{W'}(\alpha) = \mathbf{P}_W(\beta) \cap \mathbf{P}_W(\alpha)$. Eftersom $\mathbf{\Omega}$ är sluten uppåt, så existerar det även α° och β° i W, sådana att $\mathbf{E}(\alpha^\circ) \supseteq \mathbf{P}_W(\alpha)$ och $\mathbf{E}(\beta^\circ) \supseteq \mathbf{P}_W(\beta)$. Alltså är $\mathbf{E}(\gamma)$ en delmängd både av $\mathbf{E}(\alpha^\circ)$ och $\mathbf{E}(\beta^\circ)$. Det vill säga, $\alpha \leq \alpha^\circ \geq \gamma \leq \beta^\circ \geq \beta$. Nu följer det av Korollarium 13 att $\alpha \approx_W \beta$. QED.

Den som vill använda Yablos modell i sin fullt utvecklade form (inklusive slutenhet och fullhet) för att modellera kontingent identitet, måste acceptera konsekvensen att antagandet i Sats 15 är falskt för varje icketrivial modell. Låt säga, som exempel, att vi vill konstruera en enkel modell av två ting α och β, med deras färg-egenskaper, sådan att det är möjligt att α och β har vilka färger som helst, oberoende av varandra, utan att α och β sammanfaller med varandra. Till exempel så är det möjligt att båda är blå, och det är även möjligt att α är blå och β är

vit. Men då uppfylls antagandet i Sats 15, med följden att α och β sammanfaller. Detta enkla förehavande kan alltså inte modelleras i Yablos system. Detsamma gäller naturligtvis många andra sammanhang av större komplexitet. Till exempel så har Denby i sin (2014) argumenterat för en essentialistisk position, i vilken endast intrinsikala egenskaper kan tillhöra ett tings essens. I Yablos modell, motsvaras denna filosofiska position av att alla egenskaper i **X** är intrinsikala. Då uppfylls kriteriet i Sats 15 i en stark mening: intrinsikala egenskaper är sådana som ett ting har oberoende av andra tings intrinsikala egenskaper, så om alla egenskaper i **X** är intrinsikala, och det finns två olika ting, då kan vilken kontingent **X**-egenskap som helst förändras för det ena tinget utan att påverka det andra tingets **X**-egenskaper. Därmed är Yablos system inkompatibelt med denna form av essentialism.

Låt oss undersöka mer informellt hur beviset av Sats 15 kan tillämpas i vår värld. Enligt fullhet så finns det ett ting γ som sammanfaller med en fotboll i vår aktuella värld W, men som sammanfaller med en gräsmatta i en annan värld W'. Denna andra värld W' kan vi välja fritt, så länge den inte är identisk med den aktuella världen. Låt säga att W' är mycket lik W, bara att någon sten på månen ligger en aning annorlunda, eller något sådant som orimligen påverkar Gräsmattans kumulativa egenskaper. Det innebär att varje essentiell egenskap hos γ är en kumulativ egenskap som både Fotbollen och Gräsmattan har i W. (Här används antagandet att gräsmattan har samma egenskaper i W' som i W.) Av slutenhet uppåt följer att det finns en Fotboll$^\circ$ och en Gräsmatta$^\circ$ som har alla sina kumulativa egenskaper essentiellt. Därför är γ:s essens en delmängd av var och en av Fotbollen$^\circ$:s och Gräsmattan$^\circ$:s essenser. Så båda dessa förädlar γ. Nu följer det av Korollarium 13, att alla dessa ting sammanfaller i vår aktuella värld W, men det är absurt att Fotbollen och Gräsmattan skulle vara kontingent identiska.

För att Yablos system ska kunna tillämpas, måste det finnas någon kumulativ egenskap som Gräsmattan får i W' av att stenen på månen ligger lite annorlunda. Vi kan förstås konstruera vissa artificiella kandidategenskaper som potentiellt skulle rädda systemet här, såsom ”att vara sådan att stenen på månen ligger så och så”. Men ska denna typ av (i relation till Gräsmattan) extremt extrinsikala egenskaper betraktas som kumulativa, värdiga att inkluderas i ett tings essens? Finns det ett ting, kontingent identiskt med gräsmattan, som har den essentiella egenskapen ”att vara sådan att stenen på månen ligger så och så”? I princip så kräver Yablos system detta! Jag betraktar det som en oattraktiv

aspekt av systemet, som pekar på ett behov av att modifiera Yablos fullhetskriterium.

5. FÖRSLAG TILL MODIFIKATION

Jag föreslår att Yablos fullhetsaxiom ersätts med ett nytt axiom, som presenteras i detta avsnitt. Det går utanför artikels omfång att närmare studera det modifierade systemet. Min avsikt är endast att argumentera för dess plausibilitet, som en startpunkt för vidare forskning.

Låt oss börja med en formell definition av begreppet *karaktärisering* inom ramen för slutna egenskapsmodeller.

Definition 16. Anta att $\mathbf{\Omega}$ är sluten. En delmängd $\mathbf{K} \subseteq \mathbf{X}$ är en *karaktärisering* om för alla ting α, β och för alla världar W, vi har att $\alpha, \beta \in \mathbf{K}[W] \rightarrow \alpha \approx_{w} \beta$.

Alltså, i en godtycklig värld, om två existerande ting båda har alla egenskaper i en viss karaktärisering, då sammanfaller de.

Följande definition, skulle mer informellt kunna uttryckas som att karaktäriseringarna och essenserna är desamma.

Definition 17. Anta att $\mathbf{\Omega}$ är sluten. $\mathbf{\Omega}$ är *välkaraktäriserande* om följande kriterium är uppfyllt. För varje $\mathbf{Y} \subseteq \mathbf{X}$, så är $\mathbf{Y}$ en karaktärisering om och endast om det finns $\gamma \in \mathrm{D}$ sådant att $\mathrm{E}(\gamma) = \mathbf{Y}$.

Mitt förslag är att ersätta Yablos krav på att egenskapsmodellen ska vara sluten och full, med kravet att egenskapsmodellen ska vara sluten och välkaraktäriserande. Notera först delkriteriet att det för varje karaktärisering måste finnas ett ting vars essens är den karaktäriseringen. Detta delkriterium är en svagare variant av Yablos fullhetskriterium, men som undviker fullhetskriteriets oönskade konsekvenser. Mer om detta strax. Det andra delkriteriet, att varje essens måste vara en karaktärisering, hjälper oss att bevisa följande sats, som stärker den filosofiska ståndpunkten att ett tings essens karaktäriserar vad det tinget är. Enligt denna sats, så är egenskaperna i ett tings essens tillräckliga för att bestämma tinget upp till kontingent identitet (i alla världar där tinget existerar).

Sats 18. *Anta att* $\mathbf{\Omega}$ *är sluten och välkaraktäriserande. Betänk ett ting* α *som existerar i W. Om* β *existerar i W och där har alla egenskaper i* $\mathbf{E}(\alpha)$, *då sammanfaller* β *med* α *i W.*

Bevis. Eftersom $\mathbf{E}(\alpha)$ är en karaktärisering, och vi har i W att både α och β existerar och har egenskaperna i $\mathbf{E}(\alpha)$, så sammanfaller α och β i W. QED.

Låt oss nu titta närmare på det första delkriteriet, att det för varje karaktärisering måste finnas ett ting vars essens är den karaktäriseringen. Mitt kriterium säger att för varje ”godtyckligt ihopplock” av kumulativa egenskaper, så finns det ett ting i diskursdomänen som har just detta ”ihopplock” som sin essens, förutsatt att ”ihopplocket” identifierar tinget i fråga upp till kontingent identitet. Mitt kriterium öppnar inte för den kollaps som fullhetskriteriet resulterar i. Det är exempelvis fullt möjligt att ha en sluten välkaraktäriserande egenskapsmodell där alla egenskaper är intrinsikala, utan att kollapsa kontingent identitet.

6. AVSLUTNING

Jag har i denna text gått igenom Yablos system för kontingent identitet, och funnit att det har problematiska konsekvenser. I synnerhet leder det till en känslighet för kollaps, som endast tycks kunna undvikas genom att inkludera extremt extrinsikala egenskaper i många tings essenser. Enklare modeller, där alla egenskaper är intrinsikala, är exempelvis inkompatibla med systemet. Jag föreslår därför att vi byter ut Yablos fullhetskriterium mot ett kriterium som jag kallar välkaraktärisering. Därmed uppnås en tillåtande ontologi, i enlighet med Yablos intention, samtidigt som kollaps undviks.

LITTERATUR

Denby, David. 2014. ”Essence and Intrinsicality”. I *Companion to Intrinsic Properties*, red. Robert M. Francescotti, s. 87–109. Boston: De Gruyter.

Kripke, Saul. 1971. ”Identity and Necessity”. I *Identity and Individuation*, red. Milton K. Munitz, s. 135–64. New York: New York University Press.

Yablo, Stephen. 1987. ”Identity, Essence, and Indiscernibility”. *The Journal of Philosophy* 84, nr 6, s. 293–314.

SOFIA JEPPSSON

FÖRTJÄNST OCH STRAFF

Hur kan det anses berättigat att låsa in brottslingar i fängelse? Genom årens lopp har flera olika svar getts på den frågan. Ett av de enklaste och mest klassiska är att brottslingar *förtjänar* att straffas för det de har gjort. Teorin att brottslingar bör få det straff de förtjänar kallas ”retributivism”. Retributivismen dras dock med flera problem, varav vissa är mer uppmärksammade än andra. I denna artikel kommer jag att argumentera för att retributivismen bör överges helt, eftersom vi inte kan veta vilket straff brottslingar förtjänar. Dels kan vi inte veta säkert att moraliskt ansvar existerar, och att människor alls kan förtjäna att straffas. Dels kan vi inte veta i vilken grad enskilda brottslingar var ansvariga för det de gjorde, även om moraliskt ansvar finns till att börja med. Till sist har vi ingen pålitlig metod för att matcha brott och straff med varandra. Att helt överge retributivismen är radikalt, eftersom de flesta straffsystem (inklusive det svenska) i alla fall delvis bygger på den. Det finns dock en del intressanta alternativ i den filosofiska litteraturen, som jag kort diskuterar mot slutet av artikeln.

1. RETRIBUTIVISM

Påståendet att brottslingar ska få det straff de förtjänar kan låta både brutalt och otidsenligt i mångas öron. Ändå är det en idé som spelar en viktig roll för många länders straffsystem – inklusive Sveriges, faktiskt (se till exempel von Hirsch och Aswhorth 2005). Visserligen heter det Kriminal*vården*, och vi har ett ideal som säger att brottslingar bör rehabiliteras när de väl sitter i fängelse. Brottsbalkens regler om utdömandet av skyddstillsyn är också ganska pragmatiska: ”Vid val av påföljd skall rätten som skäl för skyddstillsyn beakta om det finns anledning att anta att denna påföljd kan bidra till att den tilltalade avhåller sig från fortsatt brottslighet” (kapitel 30, paragraf 9). Grundregeln är dock att straffen ska vara *proportionerliga*.[1] I kapitel 29 stadgas att ett brotts straffvärde ska bestämmas av en kombination av skadan, kränkningen eller faran

1. Jag följer här den filosofiska traditionen att prata om ”straff”. I svensk lagstiftning brukar man dock använda ordet ”påföljd” i stället.

Filosofisk tidskrift 2017 nr 3, 58–68

som brottet orsakades, brottslingens insikter eller brist på insikt om dessa, hens motiv och avsikter. Försvårande respektive förmildrande omständigheter handlar också främst om avsikter och motiv hos brottslingen. Det är tveksamt om all denna vikt som läggs vid brottslingens sinnestillstånd kan motiveras utifrån vare sig vikten av avskräckning eller rehabilitering. Hur snabbt det går att rehabilitera en brottsling (i den mån vi över huvud taget vågar hoppas att fängelse kan ha en rehabiliterande effekt) torde bero på många andra faktorer än brottslingens motiv och avsikter vid brottstillfället. Det är också mycket möjligt att straffsystemet skulle vara mer avskräckande om det var mycket enkelt, och en och samma typ av brott alltid gav exakt samma straff. Det finns heller inga direkt pragmatiska skäl för regeln att straffvärdet delvis ska avgöras av nivån på skadan, kränkningen eller risken som brottet orsakade. Det finns ingenting som säger att brottslingar som orsakar mindre skador generellt är mer lätt-rehabiliterade än brottslingar som orsakar större skador, eller att det är svårare att avskräcka folk från att orsaka stora än små skador. Principen om proportionerliga straff, och ett straffvärde som bestäms av skada-insikt-avsikt, bygger snarare på retributivistiska idéer. En brottsling som orsakar stor skada *förtjänar* att straffas hårdare än en brottsling som orsakar liten skada. En brottsling som hade ett elakare motiv för sitt brott *förtjänar* mer straff än den brottsling som hade ett mindre elakt motiv.

Det klassiska alternativet till retributivism är förstås avskräckningsteorin, enligt vilken målet för straffsystemet bör vara att avskräcka människor från att begå brott. Detta är också en mycket inflytelserik teori, och de flesta människor instämmer nog i att detta är en viktig roll som straffsystemet har att spela. Dock verkar avskräckningsteorin ha svårt att stå på helt egna ben. Anta, till exempel, att vi skulle kunna minska mängden snatteri radikalt genom att införa fängelsestraff för snattare. Eller anta att det är väldigt viktigt, om vi ska avskräcka människor från att begå mord, att alla mord som får uppmärksamhet i media klaras upp, och att avskräckningsmålet därför gynnas rejält av att vi sätter dit oskyldiga människor för mord som annars skulle ha förblivit ouppklarade. Detta är förstås oacceptabelt. Men *varför* är det oacceptabelt? Ett svar som ligger nära till hands är att snattare inte *förtjänar* att sitta i fängelse, och den som är oskyldig inte *förtjänar* att straffas över huvud taget. Men då är vi tillbaka vid retributivismen igen. Tesen att ingen får straffas som inte förtjänar det, och att ingen får straffas hårdare än hen förtjänar, kallas ibland ”negativ retributivism” (medan ”positiv retributivism”, det som jag

bara kallar ”retributivism” i större delen av artikeln, också innefattar tesen att vi *ska* straffa de som förtjänar att straffas).

Retributivistiska idéer spelar alltså en stor roll i många straffsystem, liksom i ”vanligt folks” idéer om brott och straff. Men retributivismen dras också med stora problem.

2. KLASSISK SKEPTICISM OM ANSVAR

Idén att vi kan förtjäna straff för felaktiga handlingar bygger på antagandet att vi kan vara *moraliskt ansvariga* för det vi gör. Det är därför Brottsbalken lägger sådan vikt vid att brottslingen insåg eller borde ha insett vilken skada hen orsakade – om hen inte kunde inse att skada skulle följa på hens handling, så kan hen inte vara moraliskt ansvarig för skadan hen orsakade. Även de flesta av de uppräknade försvårande och förmildrande omständigheterna behandlar faktorer som anses öka eller minska moraliskt ansvar och klandervärdhet. En hel del filosofer argumenterar dock för att moraliskt ansvar bara är en myt. I själva verket är ingen någonsin moraliskt ansvarig för det hen gör – och därmed kan inte heller någon förtjäna att straffas när hen handlat fel (t.ex. Pereboom 2014; Waller 2004; Strawson 2002). I extrem korthet så går skeptikernas argumentation ungefär såhär. Vi är inte ansvariga för sådant som inträffar på grund av faktorer som vi saknar kontroll över. I slutändan så är dock alla våra handlingar resultatet av faktorer som vi saknar kontroll över. Vi saknar kontroll över vilka gener vi har, hur vår barndom tedde sig och över huvud taget vilka olika miljöfaktorer som vi har exponerats för under livets gång. Någon kanske vill hävda att allt detta visserligen är sant, men att jag ändå har ett *val* angående vad jag gör med min genuppsättning och mina olika influenser. Men, frågar sig skeptikern, vart kommer detta ”jag” ifrån? Måste inte ”jaget” självt vara resultatet av gener som kombinerats ihop med miljöfaktorer och resulterat i en viss personlighet som sedan gör vissa typer av val? *Allt*, säger skeptikern, inklusive mitt ”jag” och de val jag gör, har uppstått genom ett samspel mellan arv och miljö, och eftersom både arvet och min ursprungliga miljö är faktorer som jag saknar kontroll över så kan jag inte vara moraliskt ansvarig för någonting som jag gör.

Den här filosofiska läran står förstås inte oemotsagd. *Kompatibilister* argumenterar för att moraliskt ansvar inte är beroende av den sortens ultimata kontroll som skeptikerna påpekar att vi inte kan ha. Vad som krävs för moraliskt ansvar, säger många kompatibilister, är vissa ratio-

nella förmågor. Agenten måste till exempel ha förmågan att väga alternativ mot varandra och utföra den handling som hen har mest skäl att göra (se till exempel Fischer och Ravizza 1998 och Nelkin 2011). Det här är vanliga, icke mystiska förmågor som de allra flesta vuxna människor oftast har, och därför så är de allra flesta vuxna människor moraliskt ansvariga för det mesta som de gör.

Även om man anser att kompatibilisterna har starkare argument än skeptikerna, så kvarstår dock ett problem: Hur *säkra* kan vi vara på att kompatibilisterna har rätt? Vi diskuterar ju att *straffa* människor – i fallet fängelse att ta ifrån dem deras frihet, något som normalt sett är en rättighet. Ska vi göra något så drastiskt så bör vi ha rejält på fötterna. Det är inte rätt att sätta Kalle i fängelse för att han stal en bil om vi inte är *väldigt säkra* på att Kalle verkligen stal bilen. På samma sätt så borde vi väl behöva vara *väldigt säkra* på att Kalle också var moraliskt ansvarig för det han gjorde och verkligen förtjänar att sitta inlåst, innan vi straffar honom. Men kan vi verkligen vara *väldigt säkra* på att alla skeptiker har fel, när de säger att ingen, alltså inte heller Kalle, kan förtjäna att straffas? Trots allt så har en hel del intelligenta människor som tänkt djupt och länge på saken kommit fram till skeptiska slutsatser. Även den som själv är övertygad kompatibilist bör nog erkänna *möjligheten* att hen har fel trots allt och skeptikerna rätt (Kearns 2015; Vilhauer 2013, s. 162). Men då bör vi akta oss för att gå runt och dela ut de straff som vi tror att folk förtjänar.

3. DEN NYARE, EMPIRISKT MOTIVERADE SKEPTICISMEN

Den metafysiskt motiverade skepticism som jag beskrev ovan har väldigt gamla anor. Olika former av denna skepticism har existerat i tusentals år. Under senare decennier har dock en ny sorts skepticism växt fram. Anta för diskussionens skull att kompatibilisterna har rätt i att någon sorts ultimat kontroll som verkar omöjlig att uppnå inte är nödvändig för att vara moraliskt ansvarig för det man gör – det räcker att ha vissa rationella förmågor. Vi kan fortfarande fråga oss om de flesta av oss verkligen har de där rationella förmågorna i tillräckligt hög grad.

Dels har vi den s.k. situationistiska litteraturen – psykologiska experiment som verkar visa att vårt beteende i hög grad styrs av den situation vi hamnar i, snarare än av våra värderingar. Ett av de mest kända experimenten är Stanley Milgrams lydnadsexperiment (Milgram 1974/2010). Försökspersonerna som anlände till experimentet hade fått höra att det handlade om straffs påverkan på inlärning, och att de skulle lottas till

antingen en lärar- eller elevroll. Läraren förhörde eleven på en lista med ord som han skulle ha memorerat, och om han svarade fel, så skulle läraren ge honom en elchock. Först var chockerna små, men läraren instruerades av testledaren att gradvis vrida upp strömmen. Eleven började så småningom protestera, sedan skrika av smärta och till slut tystna som om han hade svimmat. De flesta lärare fortsatte dock ändå att vrida upp strömmen på testledarens instruktioner, och ge starkare och starkare chocker, ända tills de var uppe på apparatens maxnivå märkt med ”fara”. Nu handlade i själva verket inte experimentet om inlärning utan om auktoritetslydnad. Lottningen var riggad så att alla äkta försökspersoner blev lärare – ”eleven” var i maskopi med Milgram. Inga elchocker gavs heller, utan eleven fejkade. Experimentet chockade världen, eftersom ingen hade förväntat sig att helt vanliga, hyggliga människor, människor som i många fall mådde tydligt dåligt av att (som de trodde) skada en annan människa, skulle lyda order på det här sättet. Det verkar som om *situationen*, med en auktoritetsfigur som ger order, spelade betydligt större roll för de allra flesta försökspersoners beteende än vad deras värderingar gjorde. Sammantaget har en stor mängd experiment gjorts där man kunnat visa att irrelevanta faktorer påverkar människors beteende: Personer som nyss hittat ett mynt blir exempelvis betydligt mer benägna att hjälpa någon som tappat sina papper på gatan (Isen och Levin 1972), och närvaron av främmande människor kan minska vår benägenhet att hjälpa en människa i nöd (Latané och Rodin 1968). Allt det här verkar tyda på att vi gör som vi gör beroende på den situation vi befinner oss i, snarare än vilka värderingar vi har. Vi kanske inte är så rationella som kompatibilistiska teorier om moraliskt ansvar brukar utgå ifrån (Nelkin 2005)?

Ett annat psykologiskt fenomen som diskuterats av senare skeptiker är *implicit bias*, vilket kanske kan översättas till ”förutfattade meningar” på svenska. I ett experiment fick deltagarna välja vilken av två kandidater som var mest lämplig som polischef (Uhlmann och Cohen 2005). Den ena kandidaten hade en lång formell utbildning, men inte så mycket praktisk erfarenhet av polisarbete, medan den andra kandidaten hade det omvända cv:t. Hälften av försökspersonerna fick en kvinnlig sökande med formell utbildning och lite praktisk erfarenhet och en manlig sökande med lite utbildning och mycket praktisk erfarenhet, medan den andra hälften av försökspersonerna fick det omvända, alltså en praktiskt erfaren kvinna och en högutbildad man. Det intressanta var att båda grupperna övervägande valde mannen. Båda kunde ge ”rationella” skäl för sitt val: I den gruppen där mannen var högutbildad så underströk för-

sökspersonerna, när de ombads motivera sitt val, att det faktiskt är ett *chefsjobb*, vilket kräver utbildning mer än praktisk erfarenhet. I den andra gruppen menade försökspersonerna på att *polis*chef inte är som vilken chef som helst, utan praktisk erfarenhet måste väga tungt. Försökspersonerna trodde själva att de gjort ett rationellt val, men styrdes uppenbarligen av sexistiska fördomar som de själva inte ville kännas vid – ja som de själva kan ha varit helt omedvetna om. Neil Levy frågar sig om vi verkligen kan vara moraliskt ansvariga för de val vi gör i situationer när vi styrs av *omedvetna* fördomar, samtidigt som vi själva uppfattar de val vi gör som rationella och opartiska (Levy 2014, s. 92–95).

Det finns en bred konsensus kring tesen att vi måste ha en viss kapacitet för rationella val och rationella handlingar för att kunna vara moraliskt ansvariga för det vi gör. Jag använder ”rationell” här i en tunn bemärkelse – när jag gör någonting så är det ”rationellt” om jag inte kände till ett annat *mycket* bättre alternativ. I vissa situationer så verkar dock ”vanliga” människor, utan några grava psykiska funktionshinder eller problem, ha väldigt svårt för att välja och handla rationellt, även enligt denna tunna rationalitetsdefinition. De känner till att ett visst alternativ finns där, och det är mycket bättre enligt deras egna preferenser och värderingar än det som de faktiskt gör. Ändå har de svårt att se det som en riktig möjlighet. Så verkar det ha varit med i alla fall vissa av Milgrams försökspersoner, i försöket jag beskrev tidigare. De upplevde det som hemskt att (som de trodde) skada en oskyldig människa – att vägra elchocka ”studenten” hade varit *mycket* bättre för dem. De visste också att ingenting direkt hindrade dem från att vägra – de var med i ett frivilligt försök, testledaren hade ingen auktoritet att straffa dem om de vägrade lyda hans order och så vidare. Men ändå uppfattade de inte vägran som en riktig möjlighet (i alla fall inte till att börja med; vissa försökspersoner fick en plötslig uppenbarelse och insåg att de faktiskt kunde säga ”nej” en bit längre in i testet) (Milgram 1974/2010, s. 10, 52, 54–55, 87, 120). På ett liknande sätt så verkar en del kriminella, *trots* att de är missnöjda med sin kriminella livsstil och ser en laglydig livsstil som mycket mer attraktiv, ha svårt att uppfatta ett laglydigt liv som en verklig möjlighet. Det finns terapiformer för återfallsförbrytare som helt koncentrerar sig på att få klienten att just börja uppfatta ett laglydigt liv som möjligt att välja (Langlands et al. 2009; Garland och Fredrickson 2009). Detta kan ofta vara svårt och kräva en hel del terapi. Kort sagt så verkar det finnas empiriskt stöd för att även människor som inte lider av någon svår psykisk sjukdom eller intellektuellt funktionshinder ändå kan ha väldigt svårt att tänka rationellt ibland – faktorer så som stressande situationer,

omgivningens förväntningar, de kretsar man rör i sig med mera kan få det alternativ man egentligen skulle föredra att framstå som på något sätt overkligt eller helt glömmas bort. Om en kapacitet för rationella val och handlingar är nödvändig för moraliskt ansvar, så innebär rationella svårigheter rimligtvis ett minskat ansvar.

4. ATT MATCHA IHOP BROTT OCH STRAFF

Även om vi, för diskussionens skull, antar att vi är helt säkra på att moraliskt ansvar existerar och dessutom har fullt pålitliga metoder för att avgöra i hur hög grad en viss brottsling var moraliskt ansvarig för det hen gjorde, så kvarstår fortfarande ett problem för den som vill ge brottslingar det straff de förtjänar: Hur matchar vi ihop brott och straff (Braithwaite och Pettit 1990 s. 149–51)?

Det klassiska sättet att göra detta är förstås öga för öga, tand för tand. Om man tar denna princip helt bokstavligt så blir den dock svår att leva upp till. Om någon har stulit en bil, ska då straffet vara att ta hens bil ifrån hen? Om hen inte har någon bil, vad gör vi då? För att principen ska bli någorlunda rimlig får den nog tolkas som att brottslingen ska få uppleva lika mycket lidande som hen orsakat sitt offer, men även den principen blir snabbt problematisk.

Anta först att både Kalle och Pelle stjäl en bil. Kalle och Pelle har dock olika personlighet. Kalle blir lätt upprörd och stressad i pressade situationer, eller när han helt enkelt tvingas genomlida stora förändringar. Pelle däremot är väldigt anpassningsbar och har en förmåga att luta sig tillbaka och ta det lugnt i nästan vilken situation som helst. Denna skillnad i personlighet mellan Kalle och Pelle innebär att Pelle måste sitta tre gånger så länge i fängelse som Kalle för att uppnå samma nivå av lidande. Men det verkar knappast rättvisa att ge Pelle ett extra långt straff bara för att han råkar ha en viss personlighet.

Något liknande gäller om vi antar att det inte är Kalle och Pelle, utan deras offer, som skiljer sig kraftigt åt i personlighet. Anta att personen som Kalle stal bilen ifrån blir extremt upprörd över stölden, medan personen som Pelle stal bilen ifrån tar allt i livet med en klackspark, inklusive bilstölden. Om brottslingar ska åsamkas lika mycket lidande som sina offer blir slutsatsen att Pelle ska komma mycket lindrigt undan, medan Kalle ska straffas ganska hårt. Återigen så verkar detta orättvist – de har ju begått precis likadana brott!

Därmed inte sagt att det *alltid* skulle vara fel att ta hänsyn till sådana

faktorer som att en viss brottsling kanske är extra känslig och skulle lida extra mycket av att sitta i fängelse, eller att ett visst brottsoffer på grund av *sin* känslighet drabbats extra hårt av brottet. Det kan kanske finnas situationer där detta verkar rimligt. Men om vi ska ha som konsekvent princip att åsamka brottslingen lika mycket lidande som offret blev utsatt för får det konstiga konsekvenser. Dessutom återkommer rejäla epistemiska problem – hur ska vi kunna veta att exempelvis en brottsling som hävdar sig vara oerhört känslig och därför kommer att ha det extra jobbigt i fängelse verkligen är det?

I stället för att försöka matcha ihop brott och straff utifrån lidandegrader, så kan man helt enkelt rådfråga sina intuitioner. Vilken typ av straff känns intuitivt passande för exempelvis en bilstöld? I praktiken är det också denna metod som används i länder där målsättningen är att straffet ska vara ”förtjänt” eller ”proportionerligt mot brottets allvarsgrad” (inklusive Sverige). Men kan vi lita på våra intuitioner? En del filosofer som optimistiskt svarar ”ja” på den frågan pekar på det faktum att det råder en bred (inte total förstås, men bred) konsensus, även när man jämför människor från olika länder och kulturer, rörande det relativa allvaret hos olika brott (Robinson 2013, kap. 1). Alltså, presenterar man människor med en lista på olika typer av brott och ber dem ranka brotten från minst till mest allvarligt, så kommer olika människor upp med ungefär samma rankning. Vill man att straffet ska vara proportionerligt mot brottets allvar, så ska förstås ett brott ha högre straff ju allvarligare det är.

Problemet här är att en rankning inte räcker, om vi vill veta vilket brott som ska ha vilket straff. Anta att väpnat rån av en bil är dubbelt så allvarligt som stöld av en bil. Väpnat rån av en bil ska alltså straffas dubbelt så hårt som stölden. Jaha? Vad innebär det? Ska tjuven böta tusen kronor och rånaren två tusen? Ska tjuven sitta fem år i fängelse och rånaren tio? Något annat? Om alla brott ska ges det straff som brottslingen förtjänar, så räcker inte en rankning. Och när det kommer till *absoluta* nivåer av straff så varierar folks intuitioner rejält, i alla fall om man jämför olika samhällen med varandra.

Det finns faktiskt de som föreslår att vi helt enkelt ska acceptera relativism på det här området (Matravers 2014). Vad en brottsling förtjänar för straff för exempelvis en bilstöld beror på våra intuitioner om detta, och eftersom intuitionerna varierar mellan olika samhällen så varierar det också vad brottslingen förtjänar. Men detta är problematiskt. Anta att man i ett visst samhälle har en vitt spridd intuition som säger att en tjuv förtjänar att få handen avhuggen. Är då detta rättvist? Precis vad tjuven

förtjänar? Förespråkare för retributivism brukar också hävda att retributivismen ger ett skydd mot alltför hårda straff, just eftersom den säger att ingen får straffas hårdare än hen förtjänar (von Hirsch och Ashworth 2005). Men detta faller ju om människors förtjänst bestäms av intuitionerna i det samhälle de lever i, och detta samhälle har en väldigt brutal syn på brott och straff.

5. RETRIBUTIVISMEN ÄR OHÅLLBAR – MEN VILKA ÄR ALTERNATIVEN?

Retributivism ifrågasätts ofta av filosofer som själva är övertygade skeptiker, av de klassiska metafysiska skälen, om fri vilja och moraliskt ansvar (se till exempel Pereboom 2014 och Caruso 2016). Och det är förstås helt riktigt att vi inte kan rättfärdiga straffandet av människor på retributivistisk grund om moraliskt ansvar är omöjligt, och ingen förtjänar att straffas. Jag har argumenterat för att vi inte behöver utgå ifrån den klassiska skepticismen för att komma fram till att retributivismen har problem. Att det alls är *möjligt* att skeptikerna skulle ha rätt innebär att vi inte kan vara *säkra* på att brottslingar förtjänar att straffas, och vi bör vara säkra innan vi straffar någon. Även om moraliskt ansvar finns, har vi skäl att tro att många människor i många situationer saknar *fullt* ansvar för sina handlingar; psykologisk forskning visar att vi inte är så rationella eller har så mycket kontroll över våra handlingar som vi tenderar att tro. Till sist så saknar vi pålitliga metoder för att matcha ihop brott med straff. Men om inte retributivismen kan rättfärdigas – blir då slutsatsen att även de mest farliga brottslingarna ska få löpa vind för våg?

Inte nödvändigtvis. Det klassiska alternativet till retributivismen är avskräckningsteorin, enligt vilken straff kan motiveras utifrån sina avskräckande effekter. Som jag tidigare nämnt så har dock avskräckningsteorin vissa problem; den kan inte ensam förklara varför det vore fel att exempelvis sätta snattare i fängelse eller sätta dit en oskyldig för ett olöst mord. Ben Vilhauer (2013) har dock argumenterat för en typ av avskräckningsteori som inte har dessa problem. Han utgår ifrån John Rawls idé att det rättvisa samhället är det samhälle vi skulle välja att skapa om vi befann oss ”bakom okunnighetens slöja” – om vi skulle skapa ett samhälle utan att veta om vi själva skulle bli rika eller fattiga, män eller kvinnor, svarta eller vita, etc. (Rawls 1971). Vilhauer menar att detta tankeexperiment även kan appliceras på frågan om brott och straff (något som Rawls själv inte gjorde). Om vi inte visste ifall vi själva skulle bli brottslingar eller

laglydiga medborgare, vilket straffsystem skulle vi skapa? Hans svar är att vi skulle vilja ha någon form av straffsystem för avskräckningens skull, men om vi inte själva visste ifall vi skulle drabbas av det så skulle vi inte satsa på så hög avskräckning som möjligt, utan nöja oss med vad som kan uppnås på en relativt mild och human väg.

Derk Pereboom och Gregg Caruso har argumenterat för att vi har rätt att låsa in riktigt farliga brottslingar, även om de inte *förtjänar* inlåsning, på samma grunder som vi har rätt att sätta folk i karantän om de bär på en riktigt farlig, smittsam sjukdom (Pereboom 2001, kap. 6; 2014, kap. 7; Pereboom och Caruso kommande). Precis som samhället har en plikt att i första hand förebygga farliga epidemier och bara ta till karantän som en sista utväg, så har vi en plikt att i första hand förebygga brottslighet. När det gäller mindre farliga brottslingar öppnar de upp för att milda straff, som böter, kan få användas i avskräckande syfte. Min åsikt är att Vilhauers teori och Perebooms och Carusos teori kan ses som komplement till varandra snarare än konkurrenter. Pereboom och Caruso nämner också s.k. *restorative justice* som någonting som skulle kunna ha en plats i deras tänkta system. I en *restorative justice*-process får offer, brottsling och andra berörda mötas, ge sina versioner av händelsen och sedan försöka komma överens om lämpliga uppgifter som brottslingen ska utföra för att så långt som det är möjligt reparera den skada som hen har ställt till med. Även här tror jag att Vilhauers teori kan vara ett bra komplement; vi kan använda Rawls tankeexperiment med okunnighetens slöja för att få ett fräscht perspektiv på hur *restorative justice* bör användas. Jag kommer inte att gå in närmare på detta i denna korta artikel, men konstaterar i alla fall att det finns intressanta alternativ till retributivism i den filosofiska litteraturen.

För att sammanfatta: Den retributivistiska synen på brott och straff dras med stora problem. Den förutsätter nämligen att vi kan veta vad folk förtjänar, men det kan vi inte. Lyckligtvis finns det alternativa sätt att se på brott och straff, och andra sätt att rättfärdiga domstolsväsende och kriminalvård än att hänvisa till att brottslingar ska få vad de förtjänar. Själv tror jag att vi kommer att få se mer och mer av dessa icke förtjänst-baserade teorier i framtiden.

LITTERATUR

Braithwaite, John och Philip Pettit. 1990. *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford: Oxford University Press.

Caruso, Gregg. 2016. ”Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model”. *Southwest Philosophy Review* 32, nr 1, s. 25–48.

Caruso, Gregg och Derk Pereboom. Kommande. ”Hard Incompatibilism Existentialism: Neuroscience, Punishment and Meaning in Life”. I *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, red. Gregg Caruso och Owen Flanagan. New York: Oxford University Press.

Fischer, John och Mark Ravizza. 1998. *Responsibility and Control.* Cambridge: Cambridge University Press.

Garland, Eric och Barbara Fredrickson. 2009. ”Application of the Broaden-and-Build Theory of Positive Emotions to Intimate Partner Violence”. I *Strengths-Based Batterer Intervention: A New Paradigm in Ending Family Violence*, red. Peter Lehmann och Catherine Simmons, s. 189–215. New York: Springer

Isen, Alice och Paula Levin. 1972. ”Effect of Feeling Good on Helping: Cookies and Kindness”. *Journal of Personality and Social Psychology* 21, nr 3, s. 384–88.

Kearns, Stephen. 2015. ”Free Will Agnosticism” *Noûs* 49, s. 235–52.

Langlands, Robyn et al. 2009. ”Applying the Good Lives Model to Male Perpetrators of Domestic Violence”. I *Strengths-Based Batterer Intervention: A New Paradigm in Ending Family Violence*, red.Peter Lehmann och Catherine Simmons, s. 217–235. New York: Springer.

Latané, Bibb och Judith Rodin. 1969. ”A Lady in Distress: Inhibiting Effects of Friends and Strangers on Bystander Intervention”. *Journal of Experimental Social Psychology* 5, nr 2, s. 189–202.

Matravers, Matt. 2014. ”Proportionality Theory and Public Opinion”. I: Jesper Ryberg och Julian Roberts (utg.) *Popular Punishment: On the Normative Significance of Public Opinion.* Oxford University Press, s. 33–53.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View.* Nytryck 2010. New York: HarperCollins.

Nelkin, Dana. 2005. ”Freedom, Responsibility and the Challenge of Situationism”. *Midwest Studies in Philosophy* 29, nr 1, s. 181–206.

———. 2011. *Making Sense of Freedom and Responsibility.* Oxford: Oxford University Press.

Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.

———. 2014. *Free Will, Agency and Meaning in Life*. Oxford: Oxford University Press.

Robinson, Paul. 2013. *Intuitions of Justice and the Utility of Desert*. Oxford: Oxford University Press.

Strawson, Galen. 2002. ”The Bounds of Freedom”. I *The Oxford Handbook of Free Will*, red. Robert Kane, s. 441–60. Oxford: Oxford University Press.

Uhlmann, Eric och Geoffrey Cohen. 2005. ”Constructed Criteria: Redefining Merit to Justify Discrimination”. *Psychological Science* 16, nr 6, s. 474–80.

Vilhauer, Ben. 2013. ”Persons, Punishment and Free Will Skepticism”. *Philosophical Studies* 162, nr 2, s. 143–63.

Von Hirsch, Andrew och Andrew Ashworth. 2005. *Proportionate Sentencing. Exploring the Principles. Oxford:* Oxford University Press.

Waller, Bruce. 2004. ”Virtue Unrewarded: Morality Without Moral Responsibility”. *Philosophia* 31, nr 3–4, s. 427–47.

JAKOB LETH FINK

ARISTOTELES OM MORALISK BLINDHET

I sin diskussion av praktiskt förnuft (*phronēsis*) i *Den nikomachiska etiken* beskriver Aristoteles ett fenomen jag här kommer att kalla ”moralisk blindhet”. Han talar om en människa som har blivit så fullständigt förstörd av njutning och smärta att moraliska principer ”inte framträder” (*ou phainetai*) för denna person. I denna text ämnar jag beskriva de delar av Aristoteles’ representationsteori som förklarar en sådan moralisk blindhet (svenska ”representation” är min översättning av det grekiska ”*phantasia*”). Det är alltså den aristoteliska teorin om *phantasia* som vi här ska granska närmare särskilt där den är relevant för Aristoteles’ moralfilosofi. Jag ska först kort presentera den text där Aristoteles beskriver den moraliska blindheten (sektion 1) och sedan ta upp de fysiologisk-psykologiska (sektion 2 och 3) och de praktisk-filosofiska (sektion 4) elementen inom hans representationsteori. Till slut återvänder jag till den moraliskt blinde (sektion 5).

Jag kommer inte att ge en fullständig framställning av *phantasia* hos Aristoteles, utan endast ta upp de element som är avgörande för frågan om moralisk blindhet. Min diskussion bygger på relativt ny forskning, forskning som tenderar att betona den aristoteliska fysiologins och psykologins betydelse för den aristoteliska etiken. För ett mer detaljerat studium av *phantasia* och relaterade ämnen hänvisar jag till denna forskning (Corcilius 2008; Moss 2012).

1. MORALISK BLINDHET

Aristoteles beskriver moralisk blindhet på följande sätt i en passage från *Den nikomachiska etiken* bok 6, kapitel 5:

> Orsaken till att vi för självkontroll (*sōphrosynē*) använder denna benämning, är att den bevarar det praktiska förnuftet (*phronēsis*). Den bibehåller nämligen just denna typ av begrepp (*hypolēpsis*). Ty det angenäma och det plågsamma förstör eller förvränger inte alla begrepp, såsom t.ex. beträffande vinkelsumman i en triangel, huruvida den är lika med två räta vinklar eller inte, utan endast begrepp om handlingar (*peri to prakton*).

Filosofisk tidskrift 2017 nr 3, 69–81

> Principerna för handlingarna utgörs nämligen av den omständighet för vars skull de utförs, fast för honom som har blivit alldeles korrumperad av det angenäma eller det plågsamma framträder inte omedelbart principen (*euthys ou phainetai archē*), dvs. den sak för vars skull eller till följd av vilken han bör välja och göra allt vad han gör. Ty uselheten (*kakia*) förstör principen. (*EN* 6.5.1140b11–18. Översättning av Ringbom 1993, s. 164. Ändrad)

Vi kommer att återvända till denna text mer i detalj nedan (sektion 5). För ögonblicket räcker det att påpeka följande: Det praktiska förnuftet arbetar med en typ av begrepp som rör föremål som kan realiseras genom handling. Till skillnad från teoretiska begrepp, t.ex. om triangelns egenskaper, påverkas denna typ av praktiska begrepp i så hög grad av njutning (det angenäma) och smärta (det plågsamma) att de inte alls uppmärksammas av eller framträder för vissa personer. Dessa personer är moraliskt blinda på det sättet att principerna inte skiner igenom eller inte representeras för dem (*ou phainetai*). Men vilken betydelse har då detta ”att representeras” inom den moraliska kontexten? Svaret till det finns i Aristoteles’ fysiologisk-psykologiska reflektioner över representation (*phantasia*). För att kunna besvara frågan om hur vi skulle förklara den moraliska blindheten, måste vi därför först granska *phantasia*. Detta leder oss till de naturfilosofiska texterna *Om själen* (*De anima*) och *Om djurs rörelse* (*De motu animalium*).

2. *PHANTASIA*: DEN GRUNDLÄGGANDE VERSIONEN

Mycket har skrivits om *phantasia*, inte minst därför att det inte är klart om Aristoteles överhuvudtaget har en koherent teori om detta och i så fall vad denna är. De många problemen och delfrågorna, t.ex. om *phantasias* kritiska funktion, om den exakta relationen mellan perception och *phantasia* eller om *phantasia* som kreativ förmåga, ska jag inte här behandla utan nöjer mig med att hänvisa till den nyare litteraturen.[1]

Aristoteles skiljer mellan åtminstone två sorters *phantasia*: perceptiv *phantasia* och rationell *phantasia*, varav en typ kallas ”övervägande” (*DA* 3.10.433b29–30; 3.11.434a5–10). I analogi med det svenska ”perceptiv” (av medeltidslatin *perceptivus*) ska jag här använda ”deliberativ” (av latin *deliberativus*) om denna övervägande typ av *phantasia*, även om det kanske känns lite främmande. Perceptiv *phantasia* tycks vara den primära i

1. Scheiter 2012, s. 251–78, har en fin sammanfattning av litteraturen om *phantasia*. Jfr också Moss 2012, s. 51–57.

förhållande till den rationella och förefaller också utgöra basen för den deliberativa sorten. Vi skall därför granska den perceptiva först.

Phantasia är i sin mest grundläggande form en rörelse som beror på att ett percipierande väsen har haft en aktuell perception (*aisthēsis*). Denna perception bearbetas av *phantasia* och görs till en representation (*DA* 3.3.428b11–14). Det att *phantasia* förstås som en rörelse betyder att man ska tänka sig den som en förändring i det percipierande väsendets kropp. Detta är en fysiologisk process på samma sätt som en perception är det. Att *phantasia* beror på en aktuell perception betyder att *phantasia* basalt sett är receptiv. Man måste ha sett, hört, berört, smakat eller luktat någonting aktuellt för att kunna ha *phantasia* om detsamma. *Phantasia* sparar sådana perceptioner och är därför primärt receptiv eller retentiv och bara sekundärt aktiv. Resultatet av en sparad perception kallar Aristoteles ett *phantasma* (*DA* 3.3.428a1–2). Detta översätts ibland, felaktigt, med ”bild” (engelska: image) men behöver inte exklusivt beteckna en visuell representation. Ett *phantasma* kan spara alla sensoriska representationer även ursprungligt komplexa perceptioner såsom ”en vit kaffekopp” eller en situation såsom en fotbollsmatch mellan Sverige och Danmark (Scheiter 2012, 252).

Phantasia är en rörelse i kroppen och denna rörelse, säger Aristoteles, är nödvändigtvis lik den ursprungliga perceptionen (*DA* 3.3.428b14). Det betyder att en representation (*phantasia*) har en liknande kausal verkan i kroppen som den ursprungliga perceptionen. Äter jag en kanelbulle percipierar jag aktuell söthet och min mun tillförs mera saliv. Kanelbullens smak utgör här den verkande orsaken (*causa efficiens*) till dessa förändringar. Min *phantasia* av kanelbullens smak skulle nu ha en liknande kausal inverkan på min kropp som den ursprungliga perceptionen, och har det också: mina tänder löper i vatten (som man säger på danska) när jag har (eller får) en *phantasia* av kanelbullen, t.ex. om jag minns den eller föreställer mig den. Men här är mitt *phantasma*, alltså min representation, av smaken den verkande orsaken (*causa efficiens*) till förändringen (för jag äter ju inte just nu en kanelbulle).

Skillnaden mellan perception och *phantasia* består i att perceptionen endast aktualiseras när ett externt perceptionsobjekt föreligger, medan *phantasia* kan aktualiseras också utan att något sådant föreligger.[2] Jag kan föreställa mig kanelbullen och därvid reproduceras en fysiologisk reaktion som liknar den ursprungliga smakperceptionen. När en repre-

2. *DA* 3.3.429a4–8. Aristoteles säger här, att *phantasia* ”förblir” i perceptionsapparaten och att detta förklarar t.ex. drömsyner.

sentation en gång har bildas på basis av perception, behöver den inte längre ett externt objekt för att aktualiseras. Den kan aktualiseras internt, alltså i och av själen, även utan ett externt objekt.

Slutligen ser *phantasia* ut att vara avgörande för begreppsbildningen. Aristoteles säger, att *phantasia* varken är perception eller diskursivt tänkande och att *phantasia* inte kan finnas utan perception medan, å andra sidan, begreppet (*hypolēpsis*) inte kan finnas utan *phantasia* (*DA* 3.3.428a14–16). Vi borde alltså kanske tänka oss *phantasia* som en sorts bro mellan perception och tänkande. Det är dock inte helt klart vad Aristoteles menar att ett begrepp (*hypolēpsis*) är, men låt oss här anta att vi någorlunda vet vad detta betyder (ett mentalt påstående, eller en mental representation som subsumerar många likartade förhållanden under ett) (jfr *METAPH*. 1.1.981a5–7).

Phantasia är ett nödvändigt men inte ett tillräckligt villkor för begreppsbildning därför att ingen mänsklig intellektuell aktivitet äger rum utan ett innehåll och *phantasia* levererar detta innehåll åt människosjälen när den tänker, varvid ett eller flera *phantasmata* föreligger för själen och uppfyller samma funktion som perceptioner (*aisthēmata*) (*DA* 3.7.431a14–17). Utan perception, bevarad av *phantasia*, skulle begreppet (*hypolēpsis*) vara tomt och tänkandet blint, för att uttrycka sig i kantianska termer. Det betyder inte att *phantasia* ensam kan bilda begrepp. Kom ihåg att *phantasia* beror på perception och perceptionen ”tänker” inte. För att bilda begrepp behövs förutom *phantasia* både minne och tänkande i form av intellektet (*nous*) (*METAPH*. 1.1.980b25–981a7; *AN.POST*. 2.19.100a3–15).

Perceptioner är alltid partikulära och de *phantasmata* som beror på den basala, perceptiva *phantasia* måste därför också vara det. Likväl ingår *phantasia* i begreppsbildningsprocessen och spelar därvid en roll för generella, mentala påståenden eller representationer. Det är därför inte bara partikulära föremål som representeras av *phantasia* utan också generella och universella som t.ex. moraliska principer. För att kunna förklara bildandet av moralbegrepp, alltså den sorts begrepp som är relevanta för moralisk blindhet, måste vi emellertid först granska den sorts representation som ingår i moraliskt övervägande.

3. *PHANTASIA*: DEN DELIBERATIVA VERSIONEN

De anima, bok 3, kapitel 7–12, behandlar en rad problem som rör handling och relaterade ämnen, häribland särskilt strävan (*orexis*). Aristoteles introducerar också i denna förbindelse den rationella *phantasia*

(*DA* 3.10.433b29–30). Rationell (*logistikē*) *phantasia* karakteriseras av sin relation till den intellektuella själsförmågan. När den intellektuella själsförmågan bekräftar eller förnekar (*phēsēi ē apophēsēi*) någonting som gott eller uselt, initierar den en rörelse eller handling genom att få oss att eftersträva eller undvika någonting (*DA* 3.7.431a15–16). Detta kräver en speciell typ av *phantasia*, den deliberativa (*bouleutikē phantasia*), och denna måste då uppfattas som subsummerad under den rationella *phantasia*.

Den deliberativa *phantasia* introduceras i följande svårbegripliga passage:

> Den *phantasia* som anknyter sig till perceptionen (*aisthētikē phantasia*) finns alltså, som sagt var, även hos de övriga djuren, medan den *phantasia* som hör samman med deliberation (*bouleutikē*) finns hos de djur som är förnuftsbegåvade. Huruvida man skall göra det ena eller det andra kräver redan överläggning (*logismos*), och man måste nödvändigtvis väga valet med något som är enhetligt, ty man strävar alltid efter det större. Och det är alltså möjligt att göra ett *phantasma* av flera. Detta är också grunden till att [*sc*. andra djur] inte tycks ha förmågan att bilda åsikter (*doxa*), därför att de inte har *phantasia* utifrån resonemang (*ek syllogismou*). (*DA* 3.11.434a5–11. Övers. K. Järvinen 1998, 117. Ändrad)[3]

För att förstå passagen måste vi inse att Aristoteles här beskriver den psykologisk-fysiologiska sidan av det praktiska förnuftet (*phronēsis*). Vi vet från etiken att *phronēsis* primärt består i deliberation och hör till den del av förnuftssjälen som är åsiktsbildande (*EN* 6.5.1140a25–31; b25–26), vilket förklarar varför Aristoteles i *De anima*-passagen fokuserar på deliberation och åsikt.

Då *phantasia* per definition inte finns utan perception, och då deliberativ *phantasia* i *De anima*-passagen tycks utgå ifrån resonemang, måste den deliberativa *phantasia* vara ett resultat av både perception och resonemang. Om det är korrekt, får vi följande bild av den deliberativa *phantasia*: i sin mest grundläggande betydelse sparar och reproducerar *phantasia* perceptioner (se ovan sektion 2). Dessa *phantasmata*

3. Tolkningen av denna text är omstridd. Det är inte säkert att Aristoteles talar om ”andra djur” i den sista delen av texten och heller inte att han pratar om ”*phantasia* utifrån resonemang” här. Som Ross (och många andra) menar jag att Aristoteles nog ändå talar om *phantasia* utifrån resonemang (jfr Ross 1961, s. 319). Dessutom är själva texten osäker. Efter orden ”utifrån resonemang” läser handskrifterna *hautē de ekeinēn* som betyder ”den själv har denna” och vissa utgivare emenderar till *hautē de kinei* (”den själv rör”). Jag följer Bywater som atheterar, alltså stryker, orden *hautē de ekeinēn*. Polansky 2007, s. 531, diskuterar de olika tolkningsmöjligheterna.

bearbetas av tänkandet, varvid de samlas i ett *phantasma*. Det är denna process, skulle jag tro, som Aristoteles också kallar ”*phantasia* utifrån resonemang” (*ek syllogismou*), då *syllogismos* (resonemang) har grundbetydelsen ”att räkna samman”, vilket på ett sätt bara omformulerar uttrycket ”att göra ett av flera”. Den deliberativa *phantasia* måste alltså utgå från både perception och intellekt.[4]

Deliberativ *phantasia* bevarar därmed *phantasmata* som har bearbetats av intellektet. ”Att göra ett *phantasma* av flera” betyder kanske att intellektet lägger ihop flera enskilda *phantasmata* till ett generellt *phantasma*. Ett exempel: genom *phantasia* av många enskilda situationer där det sociala fenomenet erkännande spelar en roll lägger intellektet ihop en överordnad representation (*phantasma*) av detta sociala fenomen.

Under denna process tycks intellektet också tillskriva detta *phantasma* ett moraliskt värde. Vi såg ju ovan, att intellektet ibland bekräftar eller förnekar (*phēsēi ē apophēsēi*) någonting som gott eller uselt och därmed som någonting som bör eftersträvas eller undvikas (*DA* 3.7.431a14–16). Jag förstår det som så, att intellektet på detta sätt tillordnar ett moraliskt värde (gott eller uselt) till det *phantasma* som intellektet gör till *ett*. Intellektet bekräftar, t.ex., att i denna situation var det bra att få erkännande och likaså i denna andra situation osv. Med utgångspunkt i dessa enkla *phantasmata* producerar intellektet ett enda *phantasma*: ”att få erkännande är generellt sett gott”. Detta *phantasma* kan nu motivera till handling (att eftersträva det som medför erkännande och undvika det som inte gör det).

Men hur vet intellektet att det är gott eller uselt att få erkännande? Det gör det därför att alla handlingar och aktiviteter åtföljs av njutning eller smärta (*EN* 2.3.1104b34–1105a5; *EN* 10.5.1175b26–27). Handlingar som följs av njutning tenderar att uppfattas som ”goda” oavsett om de är goda eller inte (jfr *EN* 2.3.1105a2–5). Om alltså erkännande åtföljs av njutning, kommer detta att disponera intellektet till att anse erkännande som någonting gott. Denna position har nyligen beskrivits som motivationell hedonism.[5]

Den deliberativa *phantasia* skapar representationer (*phantasmata*) som enligt den aristoteliska representationsteorin måste ha en inverkan

4. Aristoteles är inte helt klar på denna punkt. Han säger ibland att man kan uppfatta *phantasia* ”som ett slags tänkande” (*hōs noēsin tina*) (*DA* 3.10.433a9–10) och andra gånger att *phantasia* antigen utgår från perception eller tänkande (*MA* 8.702a19). Om min interpretation är korrekt, kan disjunktionen i den senare texten inte vara exklusiv.

5. Corcilius 2008, s. 95–102. Kort sagt betyder det, att Aristoteles menar att njutning eller smärta i alla fall utgör den verkande orsaken (*causa efficiens*) till all handling. Det betyder inte att njutning utgör syftet (*causa finalis*) för all handling, vilket skulle vara en radikal psykologisk hedonism och inte motsvarar Aristoteles’ position, jfr *EN* 1.4.1095b14–22.

på kroppen lik den ursprungliga verkan i den situation från vilken representationen utgår; och de har precis denna verkan. Den deliberativa *phantasia* skapar *phantasmata* som orsakar njutning eller smärta. Det fungerar så att *phantasia* förändrar kroppen genom känslor av värme och kyla, och dessa känslor utgör den fysiologiska basen för njutning och smärta (*MA* 7.701b16–22).[6] Det gäller för dessa känslor att det smärtfyllda undviks medan det njutningsfulla eftersträvas (*MA* 8.701b35–702a2). Vårt *phantasma* ”att få erkännande är generellt sett gott” åtföljs alltså av en värmekänsla och därmed av njutning. Därför, alltså beroende på att vi känner njutning, motiverar detta *phantasma* till handling.

Deliberativ *phantasia* blir därmed en avgörande faktor för strävan (*orexis*) därigenom att den ”bearbetar” (*paraskeuazei*) strävan (*MA* 8.702a17–19). Det måste betyda att *phantasia* bidrar till motivation genom att leverera objekt att sträva efter, och därför uppfattar Aristoteles också *phantasia* som ett nödvändigt villkor för strävan (*DA* 3.10.433b27–29) och därmed för handling.[7]

Vi vet, slutligen, att *phantasia* nödvändigtvis ingår i begreppsbildningsprocessen. Det måste vara den deliberativa *phantasia* som ingår i bildandet av praktiska begrepp (däribland moraliska principer). Och detta är också den sorts begrepp rörande handling som Aristoteles talar om i förbindelse med moralisk blindhet (*EN* 6.5.1140b15–16). Det är denna typ av begrepp som påverkas av smärta och njutning, till skillnad från teoretiska begrepp, som är stabila och inte påverkas av dessa känslor.

4. SOCIALISERING OCH MORALISK REPRESENTATION

Den moraliskt blinde har blivit ”alldeles korrumperad av det angenäma eller det plågsamma” (*EN* 6.5.1140b17). Han, för Aristoteles pratar nästan alltid om män, har en helt igenom usel karaktär eller personlighet, och hans moraliska uselhet (*kakia*) har förstört hans representation av moraliska principer (*EN* 6.5.1140b19–20). Han har ingen *phantasia* av dessa, de kommer inte alls in på hans moraliska ”radar”.

6. Jag accepterar inte utgivaren Nussbaums strykning (athetering) av orden *thermou ē psychrou* (”av värme eller kyla”) i *MA* 7.701b20. Orden läses i alla handskrifter och det finns inga goda skäl att ta bort dem.

7. Handling beror såklart inte endast på strävan (*orexis*) utan också på beslut (*proairesis*), alltså ett rationellt element. Men beslutet är just antingen strävande förnuft (*orektikos nous*) eller förnuftig strävan (*orexis dianoētikē*) (*EN* 6.2.1139b4–5). I sig själv orsakar förnuftet inte rörelse (*EN* 6.2.1139a35–36). Det måste blandas med strävan. Den deliberativa *phantasia* förklarar hur denna blandning är möjlig, och beslutet förutsätter därmed deliberativ *phantasia*.

Att socialisering överhuvudtaget är relevant här ser vi i det att den moraliskt blinde har blivit korrumperad. Han var alltså inte alltid usel, men har blivit så genom sin uppväxt och livsföring. Aristoteles' socialiseringsteori fokuserar på vana (*ethos*) och på karaktär eller personlighet (*ēthos*). Att socialiseras betyder för Aristoteles att bilda en personlighet eller karaktär i interaktion med andra människor, och denna personlighetsbildning äger rum baserat just på vana. Detta är välkänt och har behandlats många gånger i den nyare litteraturen (Burnyeat 1980). Det är mindre känt att socialiseringsprocessen också implicerar representation (*phantasia*) av moraliska principer.

Men detta framgår av två passager som jag här kort ska kommentera. Den första passagen finner vi i inledningen till *Den nikomachiska etiken*, när Aristoteles talar om vad som måste förutsättas hos den som ska lyssna på hans föreläsning, om denne skall ha någon nytta av den:

> Detta är för övrigt anledningen till att den, som skall ha utbyte av att lyssna till föreläsningar om det sköna och det rätta och samlevnadsproblem överhuvudtaget, måste ha blivit väl uppfostrad genom sina vanor. Principen (*archē*) är nämligen ett *att* och om detta framträder tillräckligt klart (*phainoito*), behöver man inte ett *därför att*. En sådan person är antingen redan i besittning av principerna eller så kan han lätt tillägna sig dem. (*EN* 1.4.1095b4–8. Övers. M. Ringbom 1993, s. 24–25. Ändrad)

Den ideale åhöraren måste ha vanor av en speciell sort, nämligen goda eller fina vanor. Vanorna utgör som vi vet förstadiet till karaktärsbildningen (*EN* 2.1.1103a17–26) och om åhöraren har vant sig vid det fina eller ädla, har han det som krävs för att få ut något av etikföreläsningen. Annars har han det inte. Tillvänjningen till det fina implicerar nämligen uppenbart i vissa fall att åhöraren redan har en representation av en korrekt moralprincip. Med ”korrekt moralprincip” menar jag en princip för handling som inte endast förefaller vara god för den som handlar utan faktiskt är god.

Aristoteles pratar om den relevanta utgångspunkten som ett *att*. Han menar härmed en sorts vetande eller erfarenhet om handling och situationer som rör handling. Detta är ett vetande om ”fakta”, man vet (eller menar sig veta) att någonting är fallet, men man vet inte varför det är fallet. Denna sorts vetande utgör den naturliga utgångspunkten för inlärning, därför att alla människor redan har den och känner sig hemma i den. Den rör, säger Aristoteles, föremål som är kända för oss (*EN* 1.4.1095a30–b4). Den blivande åhöraren vet alltså redan t.ex. *att* erkännande är gott och *att* just denna konkreta situation kräver av honom att

ge eller få erkännande. Han vet inte nödvändigtvis *varför* det förhåller sig så, men behöver heller inte göra det. Om principen *att* representeras tillräckligt klart för honom, skulle man lätt genom undervisning kunna bibringa honom förklaringen, alltså ett *därför att*.

Vi måste emellertid notera att vanor inte nödvändigtvis ger en stabil representation av moralprinciper. Aristoteles påpekar ju att om ett *att* framträder tillräckligt klart, då behövs inte ett *därför att*. Men det måste betyda, att vanor inte nödvändigtvis leder till en tillräckligt klar representation. Den stabila representationen finns först när vanorna omvandlats till riktiga karaktärsdrag.

Detta lär vi oss bland annat i den andra texten. Vi befinner oss denna gång i *Den nikomachiska etiken*, bok 3, kapitel 5. Aristoteles behandlar här frivillighet, ofrivillighet och ansvar och kommer i samband med detta in på ämnet moralisk representation:

> Ifall någon nu säger, att alla strävar efter det skenbart goda (*to phainomenon agathon*) utan att ha inflytande på det skenbara, och att syftet för var och en tycks vara (*to telos phainetai*) av samma karaktär som han själv .... (*EN* 3.5.1114a31–b1. Övers. M. Ringbom 1993, s. 82. Ändrad)

Han beskriver här någon som försöker frånskriva sig ansvar för sin egen uppfattning av vad som är gott. Det goda eller det som förefaller gott är i slutändan (alltså som *causa finalis*) det som motiverar varje handling; vi handlar för att uppnå någonting gott (*EN* 1.1.1094a2–3) antingen omedelbart eller förmedlat. Det är därför av vikt att kunna fastslå vem som bär ansvaret för någons uppfattning om det goda. Är det vi själva eller är det inte det? Aristoteles berättar här om någon som förnekar att det skulle vara vi själva som bär detta ansvar. Om dennes argument är giltigt, skulle det innebära att ingen är ansvarig för sina egna handlingar, därför att ingen har ansvar för vad som förefaller honom eller henne gott. Argumentet beror på tre premisser varav den sista kommer att vara viktig för oss:

(i) Alla människor begär det goda eller det som förefaller dem vara det goda. (*EN* 3.4.1113a15–16; *EN* 1.2.1094a18–22)
(ii) Ingen är herre över sin egen representation (*phantasia*) av det goda.
(iii) Hur syftet (alltså det goda eller det som förefaller gott) representeras (*phainesthai*) av en person beror på den sorts person han eller hon är.

Aristoteles accepterar (i) och (iii) men avvisar (ii).[8] Det är (iii) vi nu helt kort skall granska närmre.

Även om Aristoteles här inte direkt använder termen karaktär (*ēthos*), är det vad han menar här. Karaktären grundas i vanorna, men i motsats till dessa är karaktären fast och stabil (*EN* 2.4.1105a32–33). Vanor kan ändras; det kan karaktären endast under stort besvär. Med en fullt utvecklad personlighet följer alltså också en stabil representation av moralprinciper. Om man har blivit riktigt socialiserad får man också en riktig uppfattning om moralprinciperna. Om man inte har blivit socialiserad riktigt, får man en moraliskt sett felaktig uppfattning om moralprinciperna. En god socialisering är alltså helt avgörande (*EN* 2.1.1103b212–5). Inte bara därför att den utvecklar goda egenskaper, dygder, utan också därför att dygderna för med sig en riktig uppfattning (*phantasia*) om moralprinciperna. Vi såg tidigare att den moraliskt blindes uselhet förstör hans uppfattning om principen (*EN* 6.5.1140b19–20) och att självkontrollen, som är en viktig karaktärsdygd, bevarar moralbegrepp (*EN* 6.5.1140b11–13). Samtidigt påpekar Aristoteles att dygden hos den moraliskt sett goda karaktären gör målet för handling korrekt (*EN* 6.12.1144a7–9). Karaktären stabiliserar, bevarar och korrigerar alltså moraliska principer.[9] Den moraliskt goda karaktären bevarar korrekta principer medan den moraliskt usla karaktären förstör och förvanskar korrekta principer.

5. MORALISK BLINDHET

Moralisk blindhet består i en brist på representation (*phantasia*) av korrekta principer för moralisk handling och moraliskt tänkande. Vi såg i passagen där moralisk blindhet presenterades, att detta fenomen är knutet till dygden självkontroll (*sōphrosynē*) eller bristen på densamma. Självkontrollen bevarar moralbegrepp t.ex. om moraliska principer (*EN* 6.5.1140b11–13). Men den moraliskt blinde saknar just självkontroll. Denna karaktärsdygd består i en korrekt (välavvägd) relation till njutning och smärta (*EN* 3.10.1117b23–26), men just dessa känslor har förstört den moraliskt blinde. Det är precis kontroll över dem som han saknar. Det är det som gör honom usel: Han gör fula saker på grund av njutning och gör på

8. Han menar alltså att varje person är ansvarig för sin uppfattning av det goda (*EN* 3.5.1114b1–25).

9. Det betyder också att uppfattningen av det goda i högre grad beror på vår karaktär än på vårt tänkande, jfr Moss 2012, s. 153–99.

grund av smärta inte det ädla (*EN* 2.3.1104b8–11). Han eftersträvar alltså njutning och undviker smärta i alla situationer (absolut hedonism).

Självkontrollen håller just njutningen i schack och skapar därmed utrymme för bildandet av representationer av det goda som inte sätter njutningen högst. Men den moraliskt blinde har genom ett helt liv vant sig vid njutning och har därför endast representationer av moraliska principer som sätter njutningen högst. Njutning är för den moraliskt blinde det viktigaste och ledande goda, medan Aristoteles skulle kvalificera detta och säga, att njutningen *förefaller* honom vara det goda men inte verkligen är det.

Alla moralprinciper som på något sätt strider emot njutningen kommer den moraliskt blinde inte att uppmärksamma. Njutning och smärta förstör eller förvränger hos den moraliskt blinde andra moralbegrepp (*EN* 6.5.1140b13–14). Aristoteles beskriver på andra ställen, där han behandlar en annan sorts *phantasmata* (nämligen drömmar), vad han menar med detta. Han använder en metafor: Om man tar ett fat med vatten för att använda det som en spegel och skakar det våldsamt, ser man antingen ingenting i vattnet (då är representationen ”förstörd”) eller så känner man inte igen den bild man ser, men antar att det är någonting annat än det verkligen är (då är representationen ”förvrängd”) (*DI* 3.461a14–17). Den moraliskt blinde är antingen så korrumperad av njutning och smärta att korrekta moralprinciper inte alls registreras av honom (på samma sätt som ingenting ses i det vatten som är i våldsam rörelse) eller så korrumperad att han registrerar en korrekt moralprincip men tolkar den fel på ett sätt som svarar mot hans egen representation av det goda (som när man ser någonting i vatten men tror det är någonting annat, som det inte egentligen är). Detta sista svarar mot en förvrängning av moralprincipen.

Men vilken sorts moralprincip är det som den moraliskt blinde inte ”ser” eller, i den mån han ”ser” den, förvränger? Är han blind för den överordnade principen (syftet) för handling eller är han blind för den konkreta situation han måste handla i? Det framgår av beskrivningen i *Den nikomachiska etiken* 6.5 att den moraliskt blinde inte ”ser” syftet (*EN* 6.5.1140b18–19). Men låt oss stanna upp ett ögonblick och tänka över vad Aristoteles menar att moraliskt övervägande (deliberation) är. Detta omfattar nämligen inte bara en uppfattning om syftet för handling, som vi ser av följande omstridda passage från *De anima*:

> Vi har dels ett på det allmänna riktat begrepp (*hypolēpsis*) och överläggning, dels ett på det särskilda riktat begrepp. Det förra säger att den som är så beskaffad bör göra det så beskaffade, medan det senare säger att det och det är så beskaffat och att jag är så beskaffad …. (*DA* 3.11.434a16–19. Övers. K. Järvinen 1998, s. 118. Ändrad)

Aristoteles pratar här om att kunna applicera en generell moralisk princip (t.ex. ”den tappre bör hålla stånd i farliga situationer”) på en konkret situation (”detta är en farlig situation och jag är en tapper person”).[10] Exemplet gör klart att deliberation är en ganska komplicerad process, som innebär uppfattningen av (a) en universell princip, (b) en partikulär situation och (c) vem den handlande är (alltså självförståelse). Vi vet redan att den moraliskt blinde saknar (a). Men vi vet ju också att han inte ”ser” (a) därför att han är en speciell karaktär, nämligen en usling. Detta har att göra med (c), alltså med hans personlighet och självuppfattning. Det är alltså inte bara syftet (a) för handling som den moraliskt blinda inte ”ser”, han ”ser” heller inte sig själv (c) som en person som måste handla i överenstämmelse med syftet (a). Kanske ”ser” han inte heller att situationen (b) kräver handling, då principen (a) och hans självuppfattning (c) inte framträder för honom; och situationen därmed inte skiljer sig från andra tillfälliga situationer som inte heller kräver att han agerar.[11]

Den moraliskt blinde saknar (a) därför att han genom sin socialisering inte har vant sig vid att handla efter (a) och därför inte har skapat och bevarat en representation av (a). Han kan gott och väl, i teoretisk mening, förstå vad (a) innebär, om någon skulle informera honom om (a), men det skulle vara irrelevant, då deliberation endast ger önskat resultat om han också praktiskt ”förstår” (a). Men det är ju just (a) som praktisk princip som hans uselhet har förstört (*EN* 6.5.1140b19–20). Detta innebär att (a) inte alls motiverar honom till handling. Vi vet att representationer från den deliberativa *phantasia* ledsagas av njutning eller smärta och därigenom levererar objekten för strävan (*orexis*). Men den moraliskt blinde har ingen sådan *phantasia* och därför heller inte det motsvarande moraliska begrepp (*hypolēpsis*) som skulle kunna motivera honom att handla. Att förklara vad (a) betyder rent teoretiskt skulle lika lite få honom att handla i överenstämmelse med (a) som om vi förklarade

10. Jfr Shields 2016, s. 368–69 för utförligare diskussion av texten.

11. Medeltidens latinska aristoteliker utarbetade detta i stor detalj, jfr Iacopo Costa, ‘Φρόνησις, pleasure and the perception of the goal: Medieval Latin tradition on *Nicomachean Ethics* 6.5.1140b11–21’, i J. L. Fink (red.), *Phantasia in Aristotle’s Ethics and Its Reception in the Arabic, Greek, Hebrew and Latin Traditions* (Bloomsbury, under utgivning).

Pythagoras' teorem för honom. Den moraliskt blinde är inte en idiot utan bara precis detta: moraliskt blind.[12]

LITTERATUR OCH FÖRKORTNINGAR

Aristoteles

AN. De anima, *Om själen*: David Ross (red.), *Aristotle: De anima*, edited with Introduction and Commentary. Clarendon, 1961.

AN.POST. Analytica Posteriora, Den annan analytik: David Ross (red.), *Aristotle's Prior and Posterior Analytics*, edited with Introduction and Commentary. Clarendon, 1949.

DI. De insomniis, *Om drömmar*: David Ross (red.), *Aristotle: Parva Naturalia*, A Revised Text with Introduction and Commentary. Clarendon, 1955.

EN. Ethica nicomachea, *Den nikomachiska etiken*: Ingram Bywater (red.), *Aristotelis Ethica Nicomachea*, recognovit brevique adnotatione critica instruxit. Clarendon, 1962 [1894].

MA. De motu animalium, Om djurs rörelse: Martha C. Nussbaum (red.), *Aristotle's De Motu Animalium*, Text with Translation, Commentary and Interpretive Essays. Princeton University Press, 1978.

METAPH. Metaphysica, *Metafysik*: David Ross (red.), *Aristotle: Metaphysics*, edited with Introduction and Commentary, 2 vol. Clarendon, 1924.

Burnyeat, Myles. "Aristotle on Learning to Be Good". I *Essays on Aristotle's Ethics*, red. Amelie O. Rorty, s. 69–92. Berkeley: University of California Press, 1980.

Corcilius, Klaus. *Streben und Bewegen: Aristoteles' Theorie der animalischen Ortsbewegung*. Berlin: Walter de Gruyter, 2008.

Costa, Iacopo. "Φρόνησις, Pleasure and the Perception of the Goal: The Medieval Latin Tradition on *Nicomachean Ethics* 6.5.1140b11–21." *Phantasia in Aristotle's Ethics and Its Reception in the Arabic, Greek, Hebrew and Latin Traditions,* edited by Jakob L. Fink. London: Bloomsbury, kommande.

Järvinen, Kimmo. *Aristoteles: Tre böcker om själen*, översättning K. Järvinen. Göteborg: Daidalos, 1998.

Moss, Jessica. *Aristotle on the Apparent Good: Perception, Phantasia, Thought, and Desire*. Oxford: Oxford University Press, 2012.

Polansky, Roland. *Aristotle's De anima*. Cambridge: Cambridge University Press, 2007.

Ringbom, Mårten. *Aristoteles: Den nikomachiska etiken*, översättning och kommentar M. Ringbom. Göteborg: Daidalos, 1993.

Scheiter, Krisanna. "Images, Appearences, and Phantasia in Aristotle." *Phronesis* 57 (2012), 251–78.

Shields, Christopher. *Aristotle: De anima*, translated with an Introduction and Commentary. Oxford: Oxford University Press, 2016.

12. Denna artikel är skriven som en del av projektet *Representation and Reality* som finansieras av Riksbankens Jubileumsfond och Göteborgs universitet. Tack till Börje Bydén och Anna-Sofia Maurin för språkgranskning av min svenska.

ANNA SOFIA MAURIN

METAFYSIK OCH (ANNAN) VETENSKAP

1. INLEDNING

I vissa sammanhang är det att säga om något att det är ”metafysik” att betrakta som en förolämpning. Hur kommer det sig? Det finns såklart lika många svar på denna fråga som det finns sätt att förstå termen ”metafysik” (så väldigt många!). Anta därför att vi med ”metafysik” menar studiet av en av oss oberoende verklighet. Anta mer specifikt att det en metafysiker gör är att formulera teorier om vilken *sorts* entiteter denna verklighet är uppbyggd av. Anta att hon söker svar på frågor rörande vad som är *fundamentalt* och vad som inte är det; vad som snarare existerar och är som det är *på grund av* existensen av och naturen hos det mer fundamentala. Anta vidare att metafysikern resonerar sig fram till sina slutsatser med hjälp av i huvudsak *a priori* metoder, utifrån en minimal – och okontroversiell – mängd observationella fakta, och med en idé om att dessa slutsatser bör vara ”vetenskapligt adekvata” (det vill säga: att de (helst) inte ska stå i direkt motsägelse till vedertagna vetenskapliga resultat). Anta med andra ord att vi med ”metafysik” avser ungefär det som moderna analytiska filosofer ägnar sig åt när de påstår att de ägnar sig åt metafysik. Vad beror anklagelser om meningslöshet, oklarhet, ovetbarhet och allmän dålighet på i fallet metafysik i just denna mening?

Återigen finns det många förklaringar. Ibland beror kritiken troligtvis på missförstånd, okunskap eller fördom. Ibland baserar sig dock omdömet att metafysik är något djupt problematiskt på kritik som är väl värd att tas på allvar både av vän och ovän till metafysiken. Carnaps (1950) konstaterande att metafysiska påståenden är meningslösa (alltså inte bara falska eller ovetbara) tillhör den senare sortens kritik. Enligt Carnap bör vi skilja mellan interna och externa frågor. Interna frågor är frågor vi ställer om världen runt omkring oss *givet* ett visst sätt att förstå, begreppsliggöra och tala om densamma. Externa frågor är frågor vi ställer om samma värld, så att säga ”oberoende” av hur vi begreppsliggör eller talar om den. Metafysikerns frågor, menar Carnap, är visserligen meningsfulla om de uppfattas såsom interna. Men det är knappast så metafysikern tänker sig att de ska förstås. Givet vårt normala sätt att begreppsliggöra världen runt omkring oss skulle nämligen ett påstående

Filosofisk tidskrift 2017 nr 3, 82–95

som <Det finns materiella objekt> i så fall trivialt följa från en mängd andra påståenden vi accepterar som sanna, påståenden som t.ex. att <Det finns ett bord> och <Det finns en stol>. Påståenden som i sin tur enkelt kan låta sig verifieras genom att vi tar en titt runt vårt arbetsrum, kök eller det lokala caféet. Men så enkelt tänker sig knappast metafysikern att det är att besvara metafysiska frågor. Problemet, menar Carnap, är att alternativet – att betrakta dessa frågor som externa – är oacceptabelt. Externa frågor ställs så att säga ”utifrån” varje sätt vi har att begreppsliggöra verkligheten på. De frågar vilket av dessa som är det sanna eller mest korrekta sättet att representera en av oss oberoende verklighet. Sålunda betraktade, menar Carnap, kan inte längre vare sig fråga eller svar verifieras. Orsaken till detta är att verifikation, och därmed också ett påståendes sanning eller falskhet, är något vi endast kan nå fram till från insidan av ett begreppssystem. Det vill säga ett system med syntaktiska och semantiska regler som just talar om för oss hur enskilda påståenden – vare sig de är analytiska eller syntetiska – verifieras.

2. ONÖDIGHETSINVÄNDNINGEN

Carnaps idéer har kanske inte samma ställning nu som för några år sedan.[1] De flesta är överens om att de verifikationistiska principer de vilar på är ohållbara och försök att omformulera det Carnapianska problemet i mer epistemologiska (kanske rent av Kantianska) termer kan kritiseras för att ”övergenerera” skepticism. Risken är med andra ord stor att ovetbarhetsanklagelser även slår hårt inom områden – som till exempel den grundläggande teoretiska fysiken – många anser producera både meningsfulla, vetbara och kanske inte minst oundgängliga teorier av stor vikt för de mer tillämpade vetenskaperna. Hur det än förhåller sig med detta kan den sorts kritik av det metafysiska projektet som ska diskuteras här varken likställas med den klassiskt Kantianska eller den klassiskt Carnapianska.

Anta därför (möjligtvis kontrafaktiskt) att varken Kant eller Carnap – eller för den delen någon Kantian eller Carnapian – lyckats få metafysiken på fall. Ändå är metafysiken inte utan sina kritiker. En sorts kritik som man kan rikta mot det metafysiska projektet ifrågasätter poängen med metafysik. Denna ”onödighetsinvändning” pekar helt enkelt på det faktum att en utredning av samma slag som den metafysiken ägnar sig

1. Se dock t.ex. Chalmers, Manley och Wasserman (2009) för flera exempel på texter som utvecklar vad man kan kalla den (neo)Carnapianska idén.

åt vad gäller studieobjekt (verklighetens grundläggande konstitution och natur) redan görs inom de olika specialvetenskaperna, bara bättre. Varför hålla på med metafysik då? Detta är en sorts kritik man ofta finner framförd av filosofer med särskilt starka (natur)vetenskapliga intressen, men det är också en sorts kritik man ibland ser riktas mot metafysik (och filosofi i allmänhet, men här kommer jag framförallt koncentrera mig på just metafysiken) av vetenskaparna själva. Skillnaden mellan metafysik och annan vetenskap, menar kritikerna, ligger i att när man inom vetenskapen formulerar sina teorier om verklighetens grundläggande natur gör man detta med hjälp av väl utarbetade (empiriska) metoder, metoder som erbjuder möjlighet att på ett oberoende sätt testa och falsifiera hypoteser, upprepa experiment, och så vidare. Det är därför vetenskapen, inte metafysiken, som kan och därmed också bör säga oss något om verklighetens grundläggande konstitution.

Denna invändning har uttryckts på olika sätt av olika människor. I en av sina senaste böcker skriver t.ex. Stephen Hawking i introduktionen:

> Vi existerar endast en mycket kort tid och under denna tid utforskar vi endast en mycket liten del av hela universum. Samtidigt är vi nyfikna varelser. Vi frågar, och söker svar. Levandes i denna vidsträckta värld, en värld som är omväxlande vänlig och grym, blickandes upp mot den enorma rymden ovanför oss, frågar vi: Hur förstår man bäst den värld vi nu befinner oss i? Hur beter sig universum? Vilken är verklighetens natur? Vad är alltings ursprung? Behöver universum en skapare? De flesta av oss spenderar ingen stor del av vår tid med att oroa oss över dessa frågor, men vi oroar oss alla för dem ibland. Traditionellt betraktas dessa frågor som filosofiska frågor, *men filosofin är död.* Filosofin har inte hängt med i utvecklingen av den moderna vetenskapen, i synnerhet fysiken. Det är därför nu vetenskaparna som bär upptäckternas fackla i vår jakt på kunskap. (Hawking och Mlodinow, 2010, min övers.) [2]

Liknande kritik har också framförts filosofer emellan. I boken *Every Thing Must Go* (2007) fokuserar t.ex. författarna James Ladyman och Don Ross på den, som de menar, höggradigt problematiska metod med vars hjälp metafysikern resonerar sig fram till sina slutsatser. För trots att metafysikens studieobjekt är detsamma som fysikens, tror meta-

2. I ett mycket uppmärksammat tal vid en *Google Zeitgeist* sponsrad konferens i Hartfordshire (2011) upprepar han i princip samma sak. Detta uttalande, liksom påståendena i hans senaste bok, gav upphov till en del diskussion. Ett litet relativt populärvetenskapligt urval inkluderar: Norris (2011) och Reisz (2015). Den här animerade diskussionen är också värd att ta en titt på: https://iai.tv/video/hawking-vs-philosophy

fysikern att hon kan förlita sig på rena ”länsstolsintuitioner” när hon formulerar sina teorier. Detta är problematiskt av åtminstone två skäl:

> För det första kräver det att vi ignorerar det faktum att vetenskapen, och i synnerhet fysiken, har lärt oss att universum är väldigt mycket annorlunda än hur vi tenderar att tänka oss att det är. För det andra kräver det att vi ignorerar några centrala implikationer av evolutionsläran samt av våra bästa kognitiva och behavioristiska teorier om medvetandets natur. (ibid., min översättning)

Allra värst tycker Ladyman och Ross att de filosofer/metafysiker som explicit annonserat att de är *naturalister*, och att de därför lägger stor vikt vid vetenskapliga resultat, är. För trots sin djupt kända tilltro till den vetenskapliga metoden utmärks det de i praktiken gör bland annat av att de:

1. ignorerar vetenskapliga resultat även när dessa är höggradigt relevanta, samt att de
2. använder omodern eller ”domesticerad” vetenskap, snarare än modern dito, när vetenskapliga resultat väl uppmärksammas.

Det finns flera saker att säga om den sorts kritik av metafysik (och i vissa fall av filosofi i allmänhet) som framförts av Hawking, Ladyman och Ross med flera. Här kommer jag, med utgångspunkt i Paul (2012) koncentrera mig på att sätta denna sorts kritik i välbehövligt perspektiv. Även om det finns flera sätt att förstå kritiken på verkar det nämligen som om den genomgående utgår ifrån att åtminstone följande två antaganden är sanna:

Samma studieobjekt: Vetenskap och metafysik ställer och försöker besvara samma (sorts) frågor om samma (sorts) studieobjekt.

Olika metodologi: Vetenskap och metafysik gör ovanstående med hjälp av radikalt skilda metodologier.

I det följande ämnar jag försöka övertyga läsaren om att dessa båda antaganden är långt ifrån uppenbart sanna och att metafysik och vetenskap därför inte nödvändigtvis är att betrakta som varandras rivaler. Tvärtom verkar det som om metafysik och vetenskap står i någon form av beroendeförhållande till varandra. Att vetenskapen i en viss mening utgår från (metafysiska) antaganden, samt att metafysiken i normalfallet begränsar sig till att hävda sådant som går att betrakta som vetenskapligt adekvat.

3. SAMMA STUDIEOBJEKT?

Vid en första anblick kan det tyckas som om vetenskap och metafysik mycket riktigt ägnar sig åt att studera samma sak. Precis som inom metafysiken, är man inom vetenskapen intresserad av att identifiera verklighetens mest grundläggande sorters entiteter. Och precis som inom metafysiken, vill man inom vetenskapen förstå hur det mer grundläggande förhåller sig till det mindre grundläggande (även om man inom vetenskapen ofta snarare talar om förhållandet mellan entiteter/processer på mikro- och makronivå). Enligt förespråkare för den s.k. *onödighetsinvändningen* är de två projekten därmed rivaler, något som inte talar till metafysikens fördel.

Mot detta kan man dock invända. För det första är det oklart om metafysik och vetenskap verkligen studerar precis samma portion av verkligheten. Och även om det visar sig att de faktiskt gör det, är det likväl oklart om det följer av detta att de också studerar ”samma sak”. En förespråkare för *onödighetsinvändningen* vill såklart inte råka hävda att, bara för att till exempel kemi och fysik studerar fenomen som faller inom samma domän (fysiska ting och processer), så studerar de därmed samma sak. Samma fenomen kan uppenbarligen studeras ur mer än ett perspektiv, där valet av perspektiv i sin tur påverkar vilken kunskap vi får om det. När vi ska värdera *onödighetsinvändningen* måste vi därför först försöka besvara åtminstone följande två frågor:

1. Studerar metafysik och vetenskap fenomen som faller inom samma domän?
2. Studerar metafysik och vetenskap fenomenen som faller inom denna domän på samma sätt?

Det finns skäl att tro att svaret på båda dessa frågor är *nej*. Hur man besvarar den första frågan beror på en rad olika faktorer. En sådan faktor är *vad* det är vi jämför metafysiken med. Så här långt har jag talat om ”vetenskapen”, men direkta jämförelser kommer alltid att behöva göras mellan metafysiken och enskilda (special)vetenskaper. Allra vanligast när den här sortens jämförelser är på tapeten är att man ställer metafysiken mot (grundläggande) fysik, men det är långt ifrån uppenbart att detta är den enda eller ens den mest intressanta jämförelsen i sammanhanget (även om det råkar vara den jämförelse även jag kommer att fokusera på i det som följer). Betraktade som en enhet kan man kanske säga att vetenskap*erna* studerar fenomen som faller inom samma do-

män som den som studeras inom ramen för en naturalistisk metafysik (en metafysik enligt vilken endast det som existerar i rumtiden verkligen existerar). Inom metafysiken är man dock ofta intresserad av att studera inte bara vad som finns i den fysiska rumtiden, utan också vad som finns utanför rumtiden (det abstrakta), liksom vad som endast möjligtvis finns (i vissa fall även vad som endast omöjligtvis finns). Metafysik i denna ”bredare” mening kommer att ha ett studieobjekt som skiljer sig rejält från det vetenskapliga.

Och även om det skulle visa sig att vetenskap och metafysik i någon mening studerar fenomen som faller inom *samma domän*, har vi inga goda skäl att anta att man därmed studerar vadhelst faller inom denna domän på *samma sätt*. Vi har med andra ord inga goda skäl att tro att de kategorier som identifieras inom respektive disciplin – de distinktioner som görs – är i någon intressant mening likvärdiga. Likvärdiga, vill säga, så till vida att om vi har tillgång till resultatet av den ena utredningen har vi inte längre något behov av resultatet av den andra. Orsaken till detta är att de (kategoriska) distinktioner som gör inom metafysiken inte bara är *mer generella* utan också *mer grundläggande* än de som görs inom (special)vetenskaperna.

För att se detta, anta, inte helt grundlöst, att man inom fysiken betraktar entiteter som faller inom kategorin *fält* som fundamentala (detta exempel presenteras i Paul 2012 s. 5–6). *Vilken sorts entitet är ett fält?* Anta att fysikern svarar att *fält är objekt med en viss natur* (följt av en lång och komplex utläggning om vilka egenskaper fält har, hur dessa är distribuerade, hur de relaterar till varandra, osv.). Fysiken kan ta lång tid på sig att komma fram till den slutgiltiga beskrivningen av fältens natur (om de någonsin kommer fram till en sådan). Det intressanta är dock att oavsett hur komplett denna beskrivning i slutändan blir, menar metafysikern att en rad frågor rörande fältens natur återstår att besvara. Exempel på frågor av detta slag är:

i. Vad är ett objekt?
ii. Vad är en egenskap?
iii. Vad innebär det för ett objekt att ha/äga/instantiera en egenskap?

Här är några exempel på hur en metafysiker hade kunnat resonera kring fältens natur, och därmed hur hon hade (och har) besvarat ovanstående frågor:

i. Fält är substrat i vilka (universella eller partikulära) egenskaper är instantierade.[3]
ii. Fält är ”knippen” (bundles) av (universella eller partikulära) egenskaper.[4]
iii . Fält är *sui generis* entiteter som saknar ytterligare ontologisk struktur; distinkta fält har samma egenskaper genom att t.ex. ingå i samma likhetsklass.[5]

Beroende på vilken teori om fältens natur man omfattar kommer bland annat sådant som vad man betraktar som fundamentalt att variera. En substratteoretiker kommer till exempel betrakta åtminstone två, möjligtvis tre, kategorier som fundamentala (egenskaper, substrat, och (eventuellt) sakförhållanden), medan en bundleteoretiker kommer nöja sig med en betydligt mer ekonomisk teori (en teori som endast postulerar existensen av egenskaper). Olika teorier öppnar också upp för olika idéer om hur det mer fundamentala bygger upp det mindre fundamentala. Av skäl som vi inte har utrymme att gå in på här, menar till exempel de flesta substratteoretiker att sakförhållanden byggs upp av sina ”delar” – substrat och egenskaper – medelst någon sorts icke-mereologisk kombinationsprincip (se t.ex. Armstrong 1997). Åtminstone vissa förespråkare för bundleteorin (Paul 2002 är en av dessa) menar däremot att mereologi är allt som behövs för att ”bygga” objekt av en bas av egenskaper. Och om du är en renodlad klassnominalist verkar det som om mängdlära är allt du behöver.

Det tycks till och med som om vetenskapen räknar med och utgår ifrån produkten av det metafysiska studiet av världen; att vetenskapen i åtminstone denna mening är *beroende av* resultatet av den sorts utredning metafysiken ägnar sig åt. Inom vetenskapen tycks man nämligen ofta utgå ifrån att vi har en förteoretisk förståelse av naturen hos de kategorier i vilka vi placerar de entiteter vi ämnar studera. Vi har redan sett ett exempel på detta: kategoriseringen av fält som objekt. Men även sådant som vad en naturlag, kausalitet, tid, persistens, och egenskaper är tas i någon mening för givet inom vetenskaperna, samtidigt som naturen hos dessa fenomen utgör kärnan i en rent metafysisk utredning. Metafysikens sätt att studera vadhelst som faller inom dess domän – även i det fall metafysikens och vetenskapens domän sammanfaller – *föregår* i denna mening vetenskapens. Med Pauls ord:

3. Ett svar som omfattas av t.ex. Armstrong (1978) och Lowe (2006).
4. Ett svar som omfattas av t.ex. Campbell (1990), Ehring (2011) och mig (2002).
5. Ett svar som omfattas av t.ex. Lewis (1983) (klassnominalist) och Rodriguez-Pereyra (2002) (likhetsnominalist).

> Faktum att metafysikens ämne i ontologisk mening föregår vetenskapens ämne reflekteras i det faktum att många av metafysikens begrepp föregår vetenskapens begrepp. För att kunna informera oss om medlemmarna i olika kategorier förutsätter vetenskapen alltså någon förteoretisk kännedom om dessa (metafysiska) kategoriers natur. Det finns inget sätt att förstå de centrala begreppen i klassisk fältlära och kvantkromodynamik utan en föregående tillgång till begreppet om en egenskap. Det går inte heller att greppa idén om en mekanism så som denna används inom organisk kemi utan tillgång till begreppet kausalitet. Och det går inte att göra reda för de begrepp vi använder i en biologisk representation av citronsyra utan tillgång till begreppet persistens. I fall som dessa börjar vi med det metafysiska begreppet och betraktar det som ett villkor under vilket de vetenskapliga begreppen sedan ska förstås. (2012, s. 6, min övers.)

Även om vetenskap och metafysik studerar fenomen som helt och hållet faller inom samma domän (något som vi har sett inte heller verkar särskilt troligt), så studerar de med andra ord sannolikt inte dessa fenomen på samma sätt.

4. OLIKA METOD?

Samtidigt som *vad* metafysik och vetenskap studerar i ovanstående mening kan säga skilja sig åt, kan man hävda att de två disciplinernas respektive metodologier är väldigt lika. Båda disciplinerna – och i det som följer är jag främst intresserad av jämförelsen metafysik–fysik – är intresserade av att upptäcka sanningar om ofta oobserverbara fenomen i världen, och båda disciplinerna använder sig *både* av a priori och a posteriori resonerande. Det är alltså inte uppenbart så som kritikern påstår, att metafysiken till skillnad från t.ex. fysiken *endast* ägnar sig åt länsstolsspekulationer. Inte heller är det så att den vetenskapliga metoden är rent empirisk. Skillnaden mellan vetenskap och metafysik har snarare att göra med var man tenderar att lägga tyngdpunkten.

Naturligtvis är ovanstående påståenden sanningar med modifikation. Metafysik kan bedrivas på mer än ett sätt, och det kan även vetenskaperna. Här utgår vi ifrån en förståelse av metafysik liksom av modern vetenskap som i hög grad ”modellerande” discipliner. Discipliner, vill säga, som försöker förstå världen runt omkring genom att konstruera modeller av det man är intresserade av att undersöka. Mer exakt (men inte särskilt exakt alls) utgår modellerande discipliner i normal-

fallet från något (empiriskt observerbart eller på annat sätt åtkomligt) fenomen man önskar förstå bättre. Man abstraherar sedan bort allt som man menar är att betrakta som irrelevant. Man idealiserar/förenklar för att därmed kunna producera så generella resultat som möjligt. Resultatet är en modell. En *teori* är sedan en uppsättning (sinsemellan konsistenta) modeller. Teorin betraktas som sann(olik) när dess modeller är isomorfa med det de representerar. Mer än en modell kan i princip alltid konstrueras för varje fenomen man önskar förklara, och mer än en teori om samma sak kan därmed också alltid sammanställas. Teori- och modellval sker med hänvisning till de s.k. teoretiska dygderna. Om två (eller fler) teorier är empiriskt ekvivalenta (eller, vilket är vanligare inom vetenskaperna, om de är nära nog empiriskt ekvivalenta) så bör du föredra den teori som uppvisar ett större förklaringsvärde, är mer elegant, är enklare, och som kan samexistera på ett harmoniskt sätt med redan accepterade teorier, intuitioner, och antaganden. Och så vidare. Allt detta (eller det mesta av det) är *a priori* överväganden, dvs. överväganden som ingen erfarenhet kan tala vare sig för eller emot.

Detta sätt att vetenskapa och filosofera är naturligtvis inte utan sina egna (filosofiska) problem. Det tycks vila på antagandet att man relativt enkelt kan ta sig från en (förenklad) beskrivning av någon abstrakt (ofta matematisk) struktur till ”motsvarande” struktur i världen. Tanken är ju till och med att dessa som det verkar ytterligt olika sorters strukturer kan vara isomorfa. Detta antagande ger upphov till en rad frågor. En fråga handlar om meningsfullheten i att ägna tid åt att bygga upp dessa modeller. Varför, givet att motsvarande struktur står att finna i den yttervärld vi i första hand är här för att studera, gå via modellerande? Varför inte studera världen direkt? Ett vanligt svar på den senare frågan är att man, om man går via förenklade och idealiserade modeller av verkligheten, har möjlighet att fokusera på och (förhoppningsvis uttömmande) beskriva vissa aspekter av verkligheten, som hade förblivit så att säga osynliga för oss om vi studerat den ytterligt komplexa och därmed också svåröverskådliga verkligheten direkt. Det här innebär såklart och som sagt att modellen och den (del av) verklighet(en) som modelleras inte är identiska. Och det är detta som i sin tur ger upphov till frågor rörande påståenden om ”isomorfa” strukturer. Varför tro att vissa (men inte andra) strukturer/kategoriseringar i verkligheten på detta sätt kan avspeglas i modellen? Detta är en filosofisk fråga, och en fråga som gett upphov till en stor och livaktig diskussion i metafysiken (närmare bestämt i *meta*metafysik). Bara valet

av modellerande som en gångbar metod introducerar med andra ord filosofi (och (meta)metafysik) i vetenskapen!

Modellerande är som sagt inte bara något man ägnar sig åt inom vetenskapen. Även om det inte alltid refereras till som ”modellerande” förekommer detta sätt att representera världen även inom t.ex. metafysiken. Metafysikens modeller, liksom vetenskapens, utgår från en uppsättning empiriska data, och de involverar också idealisering, abstraktion och förenkling. Precis som inom vetenskapen jämförs sedan modeller med avseende på elegans, enkelhet, och förklaringsvärde, och teorival görs medelst s.k. *inference to the best explanation* (IBE). Att modellerande är något man ägnar sig inom *både* vetenskap och metafysik understryks av Paul:

> Allt vetenskapligt teoretiserande, alltså även sådant som sker inom de empiriska vetenskaperna, baserar sig bland annat på a priori resonemang rörande enkelhet, elegans och förklaringsstyrka; värden som alltså spelar en viktig roll, både för utvecklandet av framgångsrika vetenskapliga teorier och för dito metafysiska teorier. Inom vare sig vetenskap eller metafysik används a priori resonemang för att undersöka direkt observerbara eller testbara egenskaper hos världen. […] Användandet av a priori resonemang, både inom vetenskapen och inom metafysiken, rättfärdigas snarare med hänvisning till den roll sådana resonemang spelar i ”härledningar till den bästa förklaringen” [IBE], en metod som baserar sig på idén att teorier som maximerar enkelhet, styrka elegans och de andra teoretiska dygderna är mer sannolika. (2012, s. 19, min övers.)

Detta innebär såklart inte att det *inte* finns några skillnader mellan den vetenskapliga och den metafysiska metoden. En uppenbar skillnad rör de empiriska data som fungerar som input för teoretiserande och modellerande. När metafysiska modeller konstrueras tenderar *vardaglig* erfarenhet att ges en priviligierad roll, åtminstone i den mening att denna betraktas som baslinje eller utgångspunkt för vidare teoretiserande. Det antas med andra ord att någon del av denna erfarenhet är självklar och given (om än möjlig att i slutändan bortförklara), och modeller/teorier konstrueras i akt och mening att underbygga detta antagande. Vetenskapen utgår däremot i normalfallet från erfarenhet av ett betydligt mer raffinerat slag. Data samlas in på ett systematiskt och välorganiserat sätt, experiment designas enligt förutbestämda regler, och upprepas sedan tillräckligt många gånger för att utfallet ska förtjäna sin roll i modellen.

Hur kommer det sig att det finns denna skillnad? Och hur kan metafysiken försvara sin klart mindre avancerade utgångspunkt? En möjlighet är att göra detta med hänvisning till metafysikens ämne. Metafysiken är som sagt intresserad av att komma fram till fullt ut *generella* sanningar om verkligheten. Sanningar som man tänker sig ska gälla oavsett vilken sorts erfarenhet vi utgår ifrån. Det är därför som våra vardagliga erfarenheter fungerar som en god utgångspunkt för vidare teoretiserande. Detta sätt att resonera vilar såklart på ett viktigt antagande: att entiteter på mikro- respektive makronivå är underkastade samma (metafysiska) regulariteter. Ibland har detta antagande ifrågasatts (bland annat i ljuset av resultat från modern vetenskap). Mer vetenskapligt intresserade metafysiker har då valt att koncentrera sig på vetenskapliga erfarenheter. Att använda vardagliga erfarenheter som utgångspunkt är därför inte nödvändigt – även om det är vanligt – ens för en metafysiker.

Omvänt kan man fråga om inte också vetenskapen delvis utgår från ”vardagliga” erfarenheter. Vetenskapen laborerar visserligen med en enorm mängd empiriskt innehåll, innehåll man skaffat sig på ett avancerat och vetenskapligt sätt. Som vi redan har konstaterat kan man dock hävda att en delförklaring till hur vetenskapen får tillgång till denna stora mängd data är den långa rad förteoretiska – *metafysiska* – antaganden med vars hjälp den, åtminstone initialt, kan tolka sin observationella evidens. Det är först när vi vet vad kausalitet, egenskaper, händelser och objekt är som vi kan identifiera kausalitet, egenskaper, händelser, och objekt i världen. Och det är först när vi vet det som vi kan formulera teorier om fält, elektroner, organismer och energier. Men kausalitet, egenskaper, händelser och objekt är något vi känner igen på grund av våra *vardagliga* erfarenheter av dem. Vi slår foten i stolen. Vi spelar biljard. Vi får en stöt av hårtorken. När vi frågar oss vad för sorts sak ett fält är, utgår vi förmodligen från just denna sorts erfarenheter. Vi vet vad det innebär att fråga om ett fält är ett objekt, eftersom vi vet vad ett objekt är. Och vi vet vad ett objekt är därför att vi vet vad en stol, ett bord, och en hårtork är. Bland annat. I åtminstone denna mening har vetenskap och metafysik samma erfarenhetsmässiga utgångspunkt. Detta innebär dock såklart inte att metafysiken helt kan bortse från vad vetenskapen har att säga när den ska utvärdera sina teser. Det *kan* vara så att våra vardagliga erfarenheter står i konflikt med de vetenskapliga – och i så fall bör vanligtvis antaganden som vilar på vardaglig erfarenhet vara de som förkastas.

En annan viktig skillnad mellan den metafysiska och den vetenskapliga metoden är att man inom vetenskapen kan göra finare distinktioner

mellan olika, och mer eller mindre väletablerade, uppsättningar data. Vetenskapen, men inte metafysiken, kan med andra ord manipulera och testa sina teser, och mängden rimliga modeller kan därmed krympas betydligt. Detsamma tycks inte vara sant om metafysiken. Anta t.ex. att vi utgår från följande empiriska data: distinkta ting kan vara ”samma”/ha samma natur (ett datum baserat i våra vardagliga erfarenheter av bord med samma färg, hus med samma form, osv.). Inom metafysiken har denna vardagliga och höggradigt generella observation kallats observationen om ”enhet i mångfald”, och den har gett upphov till en uppsjö av förklaringar. En sådan är nominalismen, enligt vilken denna observation bör (bort)förklaras med hänvisning till de involverade objektens primitiva natur, eller till den likhetsklass de ingår i, eller till det predikat de råkar falla under. En annan, och radikalt annorlunda teori är universalierealismen. Enligt denna teori är distinkta objekt ”samma” på grund av att det finns något – en universalie – som karakteriserar båda. Och så vidare. Inget sätt att manipulera och testa våra erfarenheter kan omkullkasta någon av dessa eller liknande teorier. Inte ens om det skulle visa sig att vår erfarenhet av enhet i mångfald är en illusion har vi därmed bevisat att t.ex. universalierealismen är falsk. En konsekvens av detta är att de erfarenheter metafysikern utgår ifrån i normalfallet är konsistenta med en mycket stor mängd rivaliserande modeller, många fler än vad som är fallet inom vetenskaperna. I denna mening spelar visserligen (vardagliga) erfarenheter en roll inom metafysiken, men de spelar lång ifrån den centrala, kanske avgörande, roll de spelar inom vetenskaperna.

Huvudskillnaden mellan modellerande inom vetenskap och metafysik ligger alltså framförallt i *hur mycket* av de respektive teorierna som involverar det oobserverbara och endast indirekt konfirmerbara, och därmed *hur många* olika rivaliserande modeller som kan maximera de teoretiska dygderna samtidigt som de gör ett adekvat jobb med att underbygga de erfarenheter de är satta att förklara/representera. Vetenskapen ställer generellt sett högre krav på sina modeller i meningen att det är svårare att formulera empiriskt adekvata modeller. Metafysikern kan relativt enkelt producera alternativa, och likvärdigt empiriskt adekvata modeller för samma fenomen, och har därför fler teorier att välja mellan. Notera dock att det, mängden empiriskt adekvata teorier till trots, är långt ifrån så att varje empiriskt adekvat teori är i någon mening ”gångbar”. Orsaken till detta är att metafysiska teorier inte bara konkurrerar med varandra med avseende på sin empiriska adekvans utan också, kanske framförallt, vad gäller sina rent teoretiska kvaliteter.

Liksom inom vetenskapen används med andra ord klassiska teoretiska desiderata som guider till sanning. Men då den empiriska evidensen spelar mindre roll för att skilja acceptabla från icke-acceptabla teorier inom metafysiken, spelar dessa desiderata eventuellt en större roll här. Detta får ytterligare konsekvenser. Precis som den empiriska evidens metafysikern vilar sig mot inte tillåter henne att utesluta särskilt många teorier kommer inte de teoretiska dygder hon sedan är hänvisad till lyckas mycket bättre. En viktig anledning till detta är att det finns mer än ett sätt att väga dygderna mot varandra.

5. AVSLUTNING

Även om vetenskap och metafysik skulle visa sig studera *samma* (del av) verklighet(en), så finns det goda skäl att tro att de studerar densamma från olika perspektiv och med olika stora generalitetsanspråk. Vetenskap och metafysik använder sig dessutom av i hög grad liknande metoder – metoder som involverar både empiri och a priori resonemang, om än som vi sett på lite olika sätt och med olika styrka. Vetenskapen laborerar dessutom med begrepp som metafysiken i sin tur förklarar och preciserar, *inklusive* begrepp som elegans, enkelhet och förklaringsvärde. Samtidigt sätter vetenskapen upp gränser för metafysiken: att en metafysisk teori som motsäger etablerad vetenskap bör förkastas är de flesta överens om. Snarare än ömsesidigt uteslutande är därför vetenskap och metafysik förmodligen odelbart sammantvinnade med varandra.

LITTERATUR

Armstrong, D. M. 1978. *Universals and Scientific Realism*, I–II. Cambridge: Cambridge University Press.

———. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Campbell, Keith. 1990. *Abstract Particulars*. Oxford: Basil Blackwell.

Carnap, Rudolf. 1950. ”Empiricism, Semantics, Ontology.” *Revue Internationale de Philosophie* 4, s. 20–40.

Chalmers, David, David Manley och Ryan Wasserman. 2009. *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Clarendon Press.

Ehring, Douglas. 2011. *Tropes: Properties, Objects, and Mental Causation*. Oxford: Oxford University Press.

Hawking, Stephen och Leonard Mlodinow. 2010. *The Grand Design: New Answers to the Ultimate Questions of Life*. London: Bantam Press.

Ladyman, James och Don Ross. 2010. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Lewis, David. 2010. ”New Work for a Theory of Universals.” *Australasian Journal of Philosophy* 61, s. 343–77.
Lowe, Jonathan. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Oxford University Press.
Maurin, Anna-Sofia. 2002. *If Tropes*. Dordrecht: Kluwer.
Norris, Christopher. 2011. ”Hawking contra Philosophy.” *Philosophy Now*, nr 82 (januari/februari), s. 21–24.
Paul, Laurie. 2002. ”Logical Parts.” *Noûs* 36, nr 4, s. 578–96.
———. 2012. ”Metaphysics as Modeling: The Handmaiden’s Tale.” *Philosophical Studies* 160, s. 1–29.
Reisz, Matthew. 2015. ”Is Philosophy Dead?” *Times Higher Education*, https://www.timeshighereducation.com/news/is-philosophy-dead/2018686.article
Rodriguez-Pereyra, Gonzalo. 2002. *Resemblance Nominalism: A Solution to the Problem of Universals*. Oxford: Oxford University Press.

INSTRUKTIONER TILL SKRIBENTER

Filosofisk tidskrift har som syfte att bidra till en allsidig och fruktbar diskussion av filosofiska problem, samt att på ett lättfattligt sätt informera om aktuell filosofisk forskning. Den vänder sig inte enbart till fackfilosofer, utan vill framför allt nå en bredare läsekrets av filosofiskt intresserade personer.

Tidskriften står öppen för olika filosofiska inriktningar, men den vill undvika bidrag som man inte kan tillgodogöra sig utan speciella förkunskaper eller tekniska färdigheter. Utöver längre artiklar på omkring 10–15 sidor, tar tidskriften gärna emot även kortare bidrag och inlägg av notiskaraktär.

MANUSKRIPT TILL FILOSOFISK TIDSKRIFT:

- skickas med e-post till lars.bergstrom@philosophy.su.se
- skall utformas i överensstämmelse med den typografi som är normal i tidskriften, utan onödig formatering
- skall vara försedda med namn och adress
- skall som regel vara skrivna på svenska, och citat från andra språk bör översättas till svenska (ej nödvändigt i fotnoter)
- särskild litteraturförteckning upprättas vid behov i alfabetisk ordning och placeras sist i manus
- för icke beställt material ansvaras ej
- korrektur läses i regel endast av redaktören
- införda bidrag honoreras inte
- i stället för särtryck erhåller författaren, gratis, 10 exemplar – för recensioner 5 exemplar – av det nummer av tidskriften i vilket bidraget varit infört