

Filosofisk tidskrift

Årgång 39 Nr 3 September 2018

- 3 Det lätta och det rätta
ERIK ANGNER
- 11 Jagträd och välfärd
KRISTER BYKVIST
- 24 No-Miracles-argumentet utan vetenskaplig realism
RICHARD DAWID
- 36 Växelsemantik (Switcher Semantics)
KATHRIN GLÜER OCH PETER PAGIN
- 52 Logik och normativitet
ANANDI HATTIANGADI
- 63 Förståelse ur ett kunskapsteoretiskt perspektiv
MIKAEL JANVID
- 72 Humes sentimentalism – inte non-kognitivism
JONAS OLSON
- 84 NOTISER

Filosofisk tidskrift

Utgges 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

Copyright © Respektive författare 2018

Grafisk form och produktion: Ulf Jacobsen

Satt med Arnhem Pro Blond

Tryck: Carlshamn Tryck & Media AB, Karlshamn 2018

ISSN 0348-7482

DET LÄTTA OCH DET RÄTTA

1. INLEDNING

För tio år sedan lanserades det s.k. nudge-programmet – ett icke politiskt program som stater och andra aktörer ska kunna använda för att hjälpa människor leva bättre liv. Programmet utvecklades av de amerikanska forskarna Richard H. Thaler och Cass Sunstein i en bok med titeln *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Thaler och Sunstein 2008). Författarna inspirerades av flera decenniers forskning inom beteendekonomin, den gren av nationalekonomin som vill förbättra ekonomisk teori genom att infoga insikter från modern psykologi. Beteendekonomin framställs ofta som ett alternativ till den neoklassiska ekonomin, mest känd för sina modeller för rationella val och förväntningar. Thalers och Sunsteins uttryckliga förhoppning var att beteendekonomin mer "realistiska" bild av mänskligt beteende skulle göra det möjligt att på vetenskapliga grunder förbättra människors välfärd, och detta utan att samtidigt inskränka deras handlingsfrihet eller autonomi.

Boken fick stor uppmärksamhet internationellt och gick inte obemärkt förbi i Sverige, vare sig inom akademien eller inom den privata sektorn. På universiteten har nudge-programmet uppmärksamats av forskare inom national- och företagsekonomi, psykologi, marknadsföring och filosofi, för att bara nämna några. Enstaka kritiska röster till trots har mottagandet överlag varit mycket positivt. Idéerna spreds sedan till den privata sektorn, där företag, stiftelser och mer löst sammanhållna nätverk börjat tillämpa dem för både kommersiella och välgörande ändamål.

Nu föreslår en regeringsrapport att nudge-programmet ska tillämpas även inom svensk offentlig administration. Rapporten som heter *När det lätta blir det rätta – en ESO-rapport om "nudging"*, är utgiven av Expertgruppen för studier i offentlig ekonomi (ESO) och författad av Francisca Ramsberg (2016).¹ Rapporten föreslår att Regeringskansliet bör tillsätta en ny funktion med uppdrag att utveckla, stödja och samordna

1. Om inte annat anges är alla sidhänvisningar till ESO-rapporten.

vad författaren på svenska kallar "beteendeinsatser" inom offentlig sektor (s. 68). Rapporten presenterades på ett seminarium i Rosenbad, där Finansmarknads- och konsumentminister Per Bolund gav förslaget om ett kansli för beteendeinsatser sitt entusiastiska bifall. Thalers Ekonomipris 2017 gav förslaget extra skjuts.

I den här artikeln avser jag granska rapporten från ett mer filosofiskt perspektiv. Min huvudsakliga tes är att de insatser som ESO-rapporten förordar är väsensskilda från dem som Thaler och Sunstein försvarar i sin bok. Medan Thaler och Sunstein presenterar nudging som ett medel för att befrämja individens egen välfärd, betraktar Ramsberg beteendeinsatser som en metod för att nå politiska mål och/eller agera i det allmännas intresse. Skillnaderna är inte uppenbara från texten i rapporten, men från både ett filosofiskt och ett praktiskt perspektiv är de avgörande. Rapporten gör sig alltså skyldig till en ansenlig och opportunistisk betydelseglidning, som orsakar onödig förvirring och försvårar en informerad analys av dess slutsatser.

2. LIBERTARIANSK PATERNALISM OCH NUDGE-AGENDAN

Thaler och Sunstein utgår från en doktrin de kallar *libertariansk paternalism* (Thaler och Sunstein 2003).² Doktrinen säger att det är godtagbart t.ex. för en stat att befrämja individens välfärd om det är möjligt att göra det utan att inskränka på hennes valfrihet eller autonomi. Ordet "paternalism" i sammanhanget speglar syftet att förbättra individens egen välfärd. Ordet "libertariansk" signalerar ambitionen att uppnå syftet utan att inskränka på individens valfrihet eller autonomi.

När mer traditionella nationalekonomer diskuterar styrmedel fokuserar de ofta på två typer. Ett styrmedel är förbud och mandat. T.ex. är det förbjudet att mörda andra människor och obligatoriskt att deklarerera. Ett annat styrmedel är incitament. T.ex. måste den som väljer att köpa whisky betala alkoholskatt, medan den som väljer att köpa elcykel i stället får en premie. Ekonomer påpekar ofta att mandat och förbud bara är en annan sorts incitament, i form av t.ex. fängelsestraff eller böter. Juridiskt är det givetvis fortfarande skillnad.

Thaler och Sunstein (2008) har inget principiellt argument emot de traditionella styrmedlen under rätt omständigheter, men de menar att det finns ett tredje alternativ – *nudging* – som ibland kan vara påkallat.

2. Det här avsnittet bygger på Angner (2016, kap. 12).

En nudge i deras bemärkelse är en insats som uppfyller fyra kriterier:

- (1) Den hjälper människor att fatta bättre beslut på egen hand, i stället för att fatta besluten åt dem.
- (2) Den är inte förenad med kostnader för individer som utsätts för den.
- (3) Den har inga (eller endast triviala) välfärdseffekter på individer som redan är rationella och välinformerade.
- (4) Den kan dock ha (icke-triviala) positiva välfärdseffekter på individer som inte redan är rationella och välinformerade.

Det första kriteriet visar hur en nudge skiljer sig från mandat och förbud: en nudge är inte tvingande på samma sätt. Det andra kriteriet visar hur en nudge skiljer sig från ett negativt incitament: det kostar ingenting att utsättas för en nudge. De tredje och fjärde kriterierna pekar på en intressant asymmetri, i det att en nudge inte har någon effekt alls på individer som redan är rationella och välinformerade, men potentiellt stora positiva effekter på dem som inte är det. P.g.a. asymmetrin kallas libertariansk paternalism ibland också för *asymmetrisk paternalism* (Camerer m.fl. 2003).

Skillnaden mellan de tre typerna av styrmedel kan illustreras med vägmärken.³ *Förbud mot infart med fordon* – ett gult minustecken mot röd bakgrund – anger att det är förbjudet att köra förbi skylten. Vägmärket *Betalväg* – stiliserade mynt i en handflata mot en bakgrund av en svart väg – anger att en viss avgift eller särskild skatt måste betalas av den som kör förbi. Vägmärket *Varning för farlig kurva* – en svängd pil mot gul bakgrund inuti en röd triangel – förvarnar om en farlig kurva samt kurvans riktning.

Det gemensamma för de tre vägmärkena är de har samma mål: att reducera antalet trafikanter som kör rakt fram förbi skylten. Skillnaden är att de når sitt mål på olika sätt. Det första säger att det är olagligt att köra rakt fram, vilket betyder att den som kör vidare kan få böter. Det rör sig alltså om ett förbud. Det andra säger inte att det är förbjudet att köra vidare, utan att den som väljer att göra det måste betala en avgift. Det är ett incitament. Det tredje säger varken att det är förbjudet eller avgiftsbelagt att köra rakt fram. Varningsmärket har ingen effekt på trafikanter som redan är rationella och välinformerade, vare sig de har avsikten att hålla sig på vägen eller att köra rakt fram över stupet. Det kan dock

3. Informationen i det här stycket kommer från Transportstyrelsens hemsida <https://www.transportstyrelsen.se/sv/vagtrafik/Vagmarken/>

ha stora positiva effekter på trafikanter som är mindre rationella eller välinformerade, och som kanske saktar in lite för att kunna hålla sig på vägen. Det är en nudge.

Det finns mycket att säga om nudging och libertariansk paternalism. En poäng som är uppenbar från exemplet ovan är att nudging egentligen inte är någon nydaning, även om ordet är nytt i sammanhanget. En annan är att termen "libertariansk paternalism" är långt från idealisk. Programmet har bara ytliga likheter med libertariansk politisk filosofi, vilken ofta helt förkastar tanken att staten ska lägga sig i människors val. Programmet har också bara ytliga likheter med paternalism i ordets vanliga bemärkelse, då det brukar definieras som doktrinen att det är godtagbart att inskränka på människors valfrihet och autonomi för att befrämja deras välfärd. En hypotes är att Thaler och Sunstein valde termen för att de ville appellera till unionen av mängden libertarianer och mängden paternalister, men att de i stället appellerade till snittet mellan mängderna – vilket är en väldigt mycket mindre skara.

Här har jag givetvis inte ambitionen att diskutera alla de frågor som väcks av den libertarianska paternalismen och nudge-programmet. I stället vill jag fokusera på en av de centrala idéerna i Thalers och Sunsteins verk: förhoppningen om att förbättra människors välfärd.

3. INDIVIDUELL VÄLFÄRD ELLER POLITISKA MÅL?

Ambitionen att förbättra individens välfärd är en av grundbultarna i Thaler och Sunsteins projekt, och det yttersta syftet med deras nudge-program. Även om de emellanåt svävar på målet i sina texter, är det genomgående temat klart. Som vi såg ovan speglar begreppet libertariansk paternalism förhoppningen att det går att förbättra människors välfärd utan att inskränka deras valfrihet eller autonomi. Och för att en insats ska vara en nudge i deras bemärkelse måste den ha (icke-triviala) positiva välfärdseffekter på en individ som inte redan är rationell och välinformerad. Notera att det som avses i båda fallen är välfärdsnivån hos den individ som utsätts för beteendeinsatsen.

Ramsberg understryker genomgående släktskapet mellan hennes projekt och Thaler och Sunsteins. Även om hon ofta använder den mer svensk klingande termen "beteendeinsatser", så framhäver redan rapportens titel att det handlar om just nudging. Ramsberg citerar andra länders framgångar med att tillämpa nudge-programmet genom s.k. *nudge units* som stöd för förslaget att tillsätta ett kansli för beteendein-

satser. I rapportens förord av ESOs ordförande Hans Lindblad står det också att nudging (inom citationstecken) "är det internationella begreppet" för beteendeinsatser (s. 3). Även när det gäller de syften Ramsberg vill uppnå använder hon emellanåt formuleringar som ligger nära Thaler och Sunsteins. Hon börjar t.ex. med att citera Regeringsformen 1 kap. 2 §, där det står att "den enskildes personliga, ekonomiska och kulturella välfärd ska vara grundläggande för den offentliga verksamheten" (citerad på s. 7). Här står den individuella välfärden i fokus, precis som hos Thaler och Sunstein.

Därför är det förvånande att notera att Ramsbergs beteendeinsatser har ett helt annat syfte än Thaler och Sunsteins. Hennes definition av "beteendeinsatser" lyder: "Med beteendeinsatser avses utformningen av beslutsarkitektur inom offentlig sektor i linje med redan beslutade mål för politiken" (s. 21). (Beslutsarkitektur har att göra med hur de tillgängliga alternativen presenteras för beslutsfattaren.) Omedelbart därefter tillägger hon: "Beteendeinsatser ska vidare vara i det allmännas intresse och inte inskränka individens valfrihet". Senare i rapporten står det att "beteendeinsatser skulle kunna leda till samhällsekonomiska vinster genom en högre måluppfyllelse för politiken, att skattemedel används mer effektivt och, i vissa fall, även till besparingar" (s. 65). I Lindblads förord läser man dessutom att "kännedom om hur individer beter sig i olika valsituationer är viktig kunskap för att politiken ska nå sina uppsatta mål" (s. 3).

Ramsberg verkar alltså dela Thalers och Sunsteins vilja att respektera individens valfrihet och autonomi. Däremot får man intrycket att beteendeinsatser i Ramsbergs mening har två syften: (1) att uppnå politiskt beslutade mål och (2) att befrämja det allmännas intresse.

Det är värt att notera att de två syften som Ramsberg vill uppnå inte är identiska. Att uppfylla politiskt beslutade mål är inte detsamma som att befrämja det allmännas intresse. Politiskt beslutade mål kan komma att avvika från det allmännas intresse t.ex. om politikerna fattar beslut på andra grunder, eller om de försöker men misslyckas med att fatta beslut i det allmännas intresse. Ramsbergs användning av ordet "vidare" i ett av citaten ovan antyder att hon är väl medveten om skillnaden. Men det är oklart från de formuleringar som förekommer i rapporten om en beteendeinsats måste uppfylla båda syftena samtidigt, eller om det räcker med att ett är uppfyllt.

Det är också värt att notera hur radikalt Ramsbergs syften skiljer sig från Thalers och Sunsteins: att använda beteendeinsatser för att uppnå

politiska mål och/eller för att verka i det allmännas intresse är inte det samma som att befrämja individuell välfärd. De två syftena kan sammanfalla, men de kan också vara motstridiga; t.ex. kan det som är i det allmännas intresse vara direkt skadligt för individen. Även de exempel Ramsberg anför skiljer sig radikalt från Thalers och Sunsteins. När de två diskuterar om hur man kan ordna en buffé för att "nudga" individen att äta nyttigare mat för hans eller hennes egna bästa pratar Ramsberg om hur man kan använda beteendeinsatser t.ex. för att minska människors klimatbelastning. Att minska klimateffekterna är ett politiskt mål och kanske dessutom i det allmännas intresse. Däremot handlar det inte nödvändigtvis alls om att befrämja välfärdsnivån hos den individ som utsätts för beteendeinsatsen.

De styrmedel som ESO-rapporten rekommenderar är alltså väsensskilda från dem som Thaler och Sunstein förordade år 2008. Rapporten är ansvarig för en anseelig betydelseglidning, från en insats som avser att öka individens egen välfärd till en som avser att uppfylla politiska mål och/eller verka i det allmännas intresse.

Nu kan man invända att betydelseglidningen är oproblematiske, eftersom de politiska målen, det allmännas intresse, och välfärden hos de individer beteendeinsatserna riktar sig mot sammanfaller tillräckligt ofta för att man i praktiken ska kunna ignorera skillnaderna. Invändningen bygger på en ogrundad empirisk hypotes. När det gäller insatser som ska lösa klimatkrisen, t.ex., måste man ofta väga framtida generationers intressen mot de nu levande individernas välfärd. Beteendeinsatser av det slag Ramsberg föreslår kan mycket väl komma att sänka välfärdsnivån idag hos individer som utsätts för dessa. Dessutom, även om de olika syftena skulle sammanfalla i praktiken, är det fortfarande sant att ESO-rapporten förordar ett helt annat program än Thaler och Sunsteins. ESO-rapporten uppvisar ingen principiell ambition att höja välfärdsnivån hos de som beteendeinsatserna riktar sig mot, även om de olika syftena kan sammanfalla.

Det ska nämnas att Ramsbergs förslag har en klar fördel framför Thalers och Sunsteins. Den vanligaste invändningen mot det ursprungliga nudge-programmet har att göra med den inbyggda paternalismen. Vissa kritiker menar att alla former av paternalism bör förkastas av princip – oavsett välfärdskonsekvenserna. Andra menar att det kanske inte är något fel på den libertarianska paternalismen i sig, men att den oundvikligen placerar oss på ett sluttande plan ner mot mer aggressiva former av paternalism och vidare mot koncentrationslägren och gula-

gera – där välfärdsnivåerna är mycket låga. När Ramsberg klipper banden mellan nudge-programmet och den individuella välfärden har hon också frikopplat förslaget från alla former av paternalism. Till skillnad från Thaler och Sunstein kan hon inte anklagas för att vara paternalist av något slag.

Ramsbergs förslag väcker i stället andra frågor. Att försöka styra människors beteende för att befrämja deras egen välfärd är en sak; att styra deras beteende för att uppnå politiskt beslutade mål och/eller verka för det allmännas intresse är en helt annan. Det senare är inte heller oproblematiskt, om än av andra skäl. En individ som utsätts för en beteendeinsats i Ramsbergs mening behöver inte dela de politiskt beslutade målen, eller ens vara medveten om att de finns.

Det går att invända att Sverige skiljer sig från USA i den bemärkelsen att människors skepsis mot statliga insatser är lägre här än där, och att det i realiteten gör beteendeinsatser av den typ Ramsberg föreslår oproblematiska. Nu är det osannolikt att beteendeinsatser av det slag ESO-rapporten förordar vore okontroversiella i Sverige. Men även om det vore ett faktum att samtliga svenskar tolererar beteendeinsatser av det slag Ramsberg förordar är detta strängt taget inte ett försvar för att genomföra sådana insatser.

Detta leder till den viktigaste invändningen mot ESO-rapporten. Den betydelseglidning som rapporten är skyldig till fördunklar de föreslagna beteendeinsatsernas verkliga natur och orsakar därmed onödig förvirring. Det i sin tur försvårar utvärderingen av ESO-rapportens slutsatser. För att kunna utvärdera ett förslag om nya styrmedel är det förstås oundgängligt att veta vad styrmedlet har för syfte. Den läsare som redan är djupt bevandrad i litteraturen om libertariansk paternalism och nudge-programmet kanske omedelbart uppfattar skillnaden mellan Ramsbergs förslag och det ursprungliga. Den mer oerfarna läsaren kan däremot lätt få intrycket att de två är identiska när de inte är det.

Invändningen illustrerar ett problem som tidigare uppmärksammats i litteraturen om nudge-programmet: problemet om ulvar i fårakläder (Angner 2016, s. 264). Ju mer accepterat och populärt det ursprungliga nudge-programmet blir, desto mer tilltalande kommer det att vara att på ett opportunistiskt vis försöka sälja andra program med förevändningen att de också handlar om nudging när det i själva verket rör sig om något annat. Här finns det anledning att vara vaksam för skiftande filosofiska grundvalar.

LITTERATUR

- Angner, Erik. 2016. *A course in behavioral economics*. Andra upplagan. London: Palgrave Macmillan.
- Camerer, Colin F., Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, och Matthew Rabin. 2003. "Regulation for conservatives: Behavioral economics and the case for 'asymmetric paternalism'". *University of Pennsylvania Law Review* 151 (3): 1211–54. <https://doi.org/10.2307/3312889>.
- Ramsberg, Francisca. 2016. *När det lätta blir det rätta – en ESO-rapport om "nudging"*. Rapport till Expertgruppen för studier i offentlig ekonomi, 2016:7. Stockholm: Finansdepartementet. <https://eso.expertgrupp.se/rapporter/forutsattningar-nudging/>.
- Thaler, Richard H., och Cass R. Sunstein. 2003. "Libertarian paternalism". *The American Economic Review* 93 (2): 175–79. <https://doi.org/10.1257/000282803321947001>.
- Thaler, Richard H, och Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

JAGTRÄD OCH VÄLFÄRD

Med vinden ryckande i den uppfällda kragen möblerar jag på lek om i släktträdet. I stället för grandonklar och sysslingskusiner befolkar jag den med tystlåtna varianter av mig själv. Ett jagträd. Min barndom utgör stammen. Grenarna hårbärgerar variationer av den jag varit och den jag kunnat bli. Grenen längst ner grenslas av en lillgammal pianist som helst flyr noterna och låter fingrarna löpa fritt. Vid sidan om honom, med vipande fötter, en ängslig fotbollsvakt som fruktar bollen. Högre upp en blundande kompositör med fingrarna löpande längs barken i nya skalor: c-ciss-e-fiss-g-aiss; c-d-ess-fiss-a. Snett ovanför, i en klyka, sitter en mentalskötare och röker, vid sidan om honom en kranförare med frilagda reglage. Därövanför en gymnasielärare, tillika trebarnsfar, och en man som krampaktigt balanserar längst ut på grenen och tittar ned. Och någonstans i mitten sitter jag, som jag blev: en kritiker och filosofidocent med äran halvvägs i behåll.¹

1. INTRODUKTION

Det är en smärtsam insikt att vi väljer vissa typer av liv, vissa grenar i vårt jagträd, och därmed väljer bort andra. Det är speciellt svårt att göra ett rimligt val om det som är bäst för en beror på ens attityder och preferenser. Det som skapar detta problem är att de olika varianterna av en person (jag-versionerna) ofta föredrar olika livsgrenar.² Jag kommer att argumentera för en lösning på detta problem, som mycket kortfattat kan sägas gå ut på att låta varje jag-versions attityder bestämma värdet av sitt eget liv och således inte tillåta att en jag-versions attityder bestämmer värdet av en annan jag-versions liv.³

1. Anderberg (2014), s. 18.

2. Talet om jag-versioner ska inte förstås som ett ställningstagande i frågan om huruvida en person kan bokstavligen sägas vara utspridd över möjliga världar och därför ha olika delar i olika världar. Satsen "min jag-variant i ett visst liv i en viss värld har en viss attityd eller preferens" är bara en smidig förkortning för "om jag skulle leva detta liv i denna värld, så skulle jag ha denna attityd eller preferens".

3. Bricker (1980) är en tidig uppsats som gav inspiration till mitt eget reflekterande kring dessa frågor.

Några viktiga antaganden bör nämnas. Jag kommer att fokusera på välfärd, vad som är bättre för en person, och anta att rimligheten i våra val beror enbart på hur våra val påverkar vår välfärd. Det finns förstås andra former av rimlighet, men det är den "prudentiella" rimligheten som är fokus i denna uppsats, dvs. rimligheten från ett egennyttigt perspektiv. Mer exakt så kommer jag att anta att vi bör (prudentiellt) maximera vår välfärd, och det är (prudentiellt) orätt att inte göra det.

Jag kommer också att anta att vad som är bättre för en person beror på dennes attityder och preferenser. Visst kan andra saker spela roll för hur bra ett liv är för en person men det verkar rimligt att åtminstone säga att, allt annat lika, så bestämmer dina attityder och preferenser vad som är ett bättre liv för dig. Jag kommer därför att anta att allt annat är lika i de principer och exempel som jag diskuterar nedan (förutom exemplet med Helen Keller).

De välfärdsgrundande attityderna och preferenserna antas vara finala (för attitydobjektets egen skull) och ovillkorliga, så de är varken instrumentella, som t.ex. min preferens för att ta en huvudvärkstablett, eller villkorade på olika tillstånd, som t.ex. min finala preferens att leva hälsosamt i framtiden givet att jag kommer att vilja detta då.

Slutligen kommer jag att göra ett förenklande (och mycket orealistiskt) antagande: varje jag-version har endast en uppsättning preferenser och attityder. Detta betyder att jag inte behöver beakta problemet med oenighet mellan jag-versioner som sitter på olika delar av samma livsgren (existerar vid olika tidpunkter inom samma liv).⁴

2. NÅGRA PROBLEMFALL

Här är två exempel som visar de svårigheter som uppstår då våra jag-versioner från olika livsgrenar är oeniga.

Karriärsvalet: filosof eller pianist?

- Om du skulle välja att bli en filosof, så skulle du komma att föredra detta liv framför ett liv som en pianist.
- Om du skulle välja att bli en pianist, så skulle du komma att föredra detta liv framför ett liv som filosof.

Vilket liv är bättre för dig? Tittar man på filosof-jagets preferenser, så verkar filosof-livet bättre för dig. Men tittar man på pianist-jagets prefe-

4. Jag behandlar detta problem i Bykvist (2003) och (2007a).

renser, så verkar pianist-livet bättre för dig. Men ett och samma liv kan inte vara både bättre och sämre för dig. Så vilket liv är bättre för dig?

Kierkegaards dilemma: gift eller ogift?

- Om du skulle välja att gifta dig, så skulle du komma att föredra att leva som ogift.
- Om du skulle välja att inte gifta dig, så skulle du komma att föredra att leva som gift.

Tittar man på ditt gifta jags preferenser, så verkar det bättre för dig att vara ogift. Men tittar man på ditt ogifta jags preferenser, så verkar det bättre för dig att vara gift. Men, återigen, ett och samma liv kan inte både vara bättre och sämre för dig. Så vilket liv är bättre för dig?

3. PROBLEMATISKA LÖSNINGAR

En mycket enkel lösning är att bara ta i beaktande de preferenser som man skulle ha om man valde att leva ett visst liv. Så om du väljer en viss livsgren, så är det de preferenser du har vid denna gren som bestämmer välfärdsvärdet av alla livsgrenar. Tillämpat på exemplen ovan så säger denna lösning att

- om du blir filosof och föredrar detta liv, så är filosoflivet bättre för dig än pianistlivet, och
- om du gifter dig och föredrar att vara ogift, så är det bättre för dig att leva som ogift än att leva som gift.

Nu kan man fråga sig varför de preferenser du skulle ha givet ett visst val ska bestämma värdet av alla livsgrenar. Vad gör dem så speciella? De behöver ju inte vara mer välgrundade, eller baserade på bättre information, eller ha någon annan egenskap som gör dem mer kvalificerade för att bestämma välfärden hos en individ.

Ett annat problem med detta synsätt är att det inte finns någon stabil standard för välfärd: vilket liv som är bättre för dig beror på vilket liv som förverkligas. Men det verkar rimligt att förvänta sig en stabil standard som anger vilket liv som är bättre för en person utan att vi vet vilket liv som faktiskt kommer att förverkligas.

Vi får också problem med att uppfylla den normativa princip som baseras på detta instabila välfärdsvärde. Vilket liv bör jag välja om jag bara fokuserar på min egen välfärd? Det liv som är bäst för mig. Vilket liv är bäst för mig? Enligt den princip vi nu diskuterar så beror det på

vad jag väljer. Men detta betyder att vi ställs inför följande svårartade normativa dilemma:

- Om jag skulle välja att gifta mig, så skulle jag handla jag orätt, eftersom det är då vore bättre för mig att leva som gift.
- Om jag skulle välja att inte gifta mig, så skulle jag handla orätt, eftersom det då vore bättre för mig att leva ett liv som gift.

Denna situation kan fångas mer schematiskt på följande subjunktivfria sätt:

- Gifter mig: föredrar att vara ogift, alltså bättre för mig att vara ogift, alltså orätt att gifta mig
- Gifter mig inte: föredrar att vara gift, alltså bättre för mig att vara gift, alltså orätt att inte gifta mig

Alltså gäller det att oavsett vad jag skulle göra så skulle jag handla orätt. Det är klart att oavsett vad jag skulle göra, så skulle det ju finnas en rätt handling – gifter jag mig så är det rätt att inte gifta sig, gifter jag mig inte så är det rätt att gifta sig – men den rätta handlingen är enbart gäckande. Hade den utförts hade den varit orätt. Detta betyder att jag inte kan uppfylla den normativa principen i detta fall, för det finns ingen handling som är sådan att om jag utförde den, så skulle den vara rätt enligt den normativa principen.⁵

En annan lösning är att endast beakta *ideala* preferenser och säga att ett liv är bättre för dig än ett annat om och endast om du skulle föredra det, om du hade all relevant empirisk information och resonerade på ett rationellt sätt.

Tanken är sen att även om icke-ideala preferenser kan variera från en livsgren till en annan så gäller detta inte för ideala preferenser. Men detta är ett mycket tveksamt antagande. Vad jag skulle föredra om jag hade all relevant empirisk information och resonerade rationellt beror till hög grad på mina grundläggande psykiska egenskaper, t.ex. mina karaktärsdrag och benägenheter. Men olika jag-versioner från olika livsgrenar kan skilja sig avsevärt vad gäller grundläggande psykiska egenskaper, vilket betyder att även ideala preferenser kan komma i konflikt.⁶

5. Detta betyder också att principen bryter mot vad som Wlodek Rabinowicz, och senare Erik Carlson, har kallat *normativ invariants*, principen som säger att den normativa statusen hos en handling inte kan bero på huruvida handlingen utförs eller inte. En mer utförlig diskussion av normativ invariants återfinns i Carlson (1995), kap. 6, sektion 6.3, och Bykvist (2007b).

6. Varför inte ge all vikt åt de ideala preferenser välfärdssubjektet har vid hennes *val-*

4. BÄTTRE LÖSNING: UPPMÄRKSAMMA ATTITYDSPOLARITET OCH ANAMMA DIAGONALPRINCIPEN

Hittills har vi bara undersökt teorier som bara beaktar preferenser för en livsgren framför en annan. Ingen vikt ges åt de *positiva*, *negativa*, eller *neutrala* attityder man kan ha till en livsgren. Jag ska nu visa att vi kan hitta en bättre teori om vi släpper in attityder som uppvisar polaritet, positiv, negativ, eller neutral.

Först måste vi klargöra vad vi menar med attitydspolaritet. Attitydspolaritet kan förklaras i termer av motivation, lust/olust, och värdeupplevelser.⁷

- Att inta en *positiv* attityd till ett liv =_{df} att vara motiverad att leva detta liv, känna lust när man tänker på det, och att uppleva det som något bra.
- Att inta en *negativ* attityd till ett liv =_{df} att vara motiverad att undvika leva detta liv, att känna olust när man tänker på det, och att uppleva det som något dåligt.
- Att inta en *neutral* attityd till ett liv =_{df} att vara varken motiverad att leva det eller motiverad att inte leva det, att varken känna lust eller olust när man tänker på det, och att uppleva det som något neutralt.

Naturligtvis så finns det kopplingar mellan preferenser och polaritetsattityder.⁸ För personer med en viss ordning på attitydlivet har vi följande koppling: A föredras framför B om och endast om en av attitydkombinationerna (1–5) nedan är fallet:

tillfälle? Motfrågan är då, "vilket valtillfälle?" Något av de förflutna, det nuvarande, eller något av de framtida valtillfällena? Om man väljer det nuvarande valtillfället, så kan man fråga sig varför bara nuvarande ideala preferenserna ska räknas när det gäller att bestämma den framtida välfärden. Dessutom finns det inget nuvarande valtillfälle för vissa subjekt, t.ex. ett ofött barns vars möjliga framtider vi vill bedöma eller en nu död person vars förflutna vi vill bedöma.

7. Dessa formuleringar gäller endast för kognitivt avancerade individer och ignoreras därför mycket av attitydspolaritetens komplexitet. Mindre avancerade individer, som spädbarn och vissa djur, kan endast uppvisa polaritet i motivation, sedd som enklare handlingsdispositioner.

8. Det kanske till och med finns *definitionsgrundade* kopplingar mellan preferenser och polaritetsattityder. Roderick Chisholm har föreslagit följande definitioner: att vara positivt inställd till x är att föredra x framför något y som man är indifferent inför; att vara negativt inställd till x är att föredra något y, som man är indifferent inför, framför x; att vara neutralt inställd till x är att vara indifferent mellan x och något y, som man är indifferent inför; att vara indifferent inför x är att vara indifferent mellan x och icke-x (eller avsaknaden av x). För mina syften i denna uppsats behöver jag inte ta ställning till denna definition av polaritetsattityder i termer av komparativ indifferens och preferens.

	A	B
1.	<i>positiv</i>	<i>neutral</i>
2.	<i>positiv</i>	<i>negativ</i>
3.	<i>positiv</i>	<i>svagare positiv</i>
4.	<i>neutral</i>	<i>negativ</i>
5.	<i>negativ</i>	<i>starkare negativ</i>

För att lätt få en överblick över informationen om både preferenser och attitydpolaritet så kommer jag att använda mig av *attitydmatriser*, som exempelvis kan se ut så här:

	A	B
A-jag	5	8
B-jag	-2	5

Talen i en *rad* representerar en viss jag-versions attityder till olika möjliga livsgrenar. Positiva tal representerar positiva attityder, negativa tal representerar negativa attityder, och nollan representerar en neutral attityd. Ett större positivt tal representerar en starkare positiv attityd, ett större negativt tal representerar en starkare negativ attityd, och samma tal representerar samma styrka. Talen i en *kolumn* representerar olika jag-versioners attityder till ett visst liv (en viss livsgren). Notera att för mina syften i denna uppsats så behöver jag inte anta att vi har ett mått på välfärdsskillnader eller välfärdskvoter som kan jämföras mellan olika jag-versioner.

Så från denna matris kan vi tex läsa av att personens A-jag (den jag-version som existerar i A) har en positiv attityd till sitt A, och en starkare positiv attityd till B. Så A-jaget föredrar B framför A. Vi kan också se att personens B-jag har en negativ attityd till A och en positiv attityd till B, så B-jaget föredrar B framför A.

Vi kan nu beskriva karriärsvalet med följande attitydmatris:

	filosof-liv	pianist-liv
filosof-jag	a	b
pianist-jag	c	d

Hittills har vi bara tittat på preferensinformationen, och då vet vi bara att $a > b$, eftersom filosof-jaget föredrar filosof-livet framför pianist-livet, och att $c < d$, eftersom pianist-jaget föredrar pianist-livet framför filosof-livet. Men det finns mer information att beakta. Vad har filosof-jaget och pianist-jaget för attityder (positiva, negativa, neutrala) till de alternativa livet? Här är en möjlighet.

	filosof-liv	pianist-liv
filosof-jag	10	5
pianist-jag	-15	-10

I denna version har mitt filosof-jag en *positiv* attityd till båda livet men en *starkare positiv* attityd till filosof-livet, och mitt pianist-jag har en *negativ* attityd till båda livet, men en *svagare negativ* attityd till pianist-livet. I denna version av karriärsvalet står alltså valet mellan ett liv som man kommer att *gilla* (ha en positiv attityd till) och ett liv som man kommer att *ogilla* (ha en negativ attityd till). Det verkar då rimligt att säga att det liv jag kommer att gilla är bättre för mig. Alltså är filosof-livet bättre för mig.

Så här ser Kierkegaards dilemma ut i matrisform:

	liv som gift	liv som ogift
mitt gifta jag	a	b
mitt ogifta jag	c	d

Om vi bara har information om preferenser, så vet vi endast att $a < b$, eftersom mitt gifta jag föredrar livet som ogift framför livet som gift, och att $c > d$, eftersom mitt ogifta jag föredrar livet som gift framför livet som ogift. Men återigen så finns det mer information att ta hänsyn till. Vad har mitt gifta jag och mitt ogifta jag för attityder (positiva, negativa, neutrala) till de alternativa livet? Här är en möjlighet.

	liv som gift	liv som ogift
mitt gifta jag	10	15
mitt ogifta jag	-5	-10

I denna version så har mitt gifta jag en *positiv* attityd till båda livet men en *starkare positiv* attityd till livet som ogift, och mitt ogifta jag har en *negativ* attityd till båda livet, men en *svagare negativ* attityd till livet som gift. Så valet står mellan ett liv som man kommer att *gilla* (ha en positiv attityd till) och ett liv som man kommer att *ogilla* (ha en negativ attityd till). Det verkar då rimligt att säga att det liv jag kommer att gilla är bättre för mig. Alltså är livet som gift bättre för mig.

Den generella idé som antyds ovan är denna:

- A är *bra* för dig om och endast om ditt A-jag har en positiv attityd till A (och ju starkare positiv attityd du har till A desto större positivt värde har A för dig).
- A är *dåligt* för dig om och endast om ditt A-jag har en negativ attityd till A (och ju starkare negativ attityd du har till A desto större negativt värde har A för dig).
- A är *neutralt* för dig om och endast om ditt A-jag har en neutral attityd till A.
- A är *bättre* för dig än B om och endast om värdet av A för ditt A-jag är större än värdet av B för ditt B-jag.⁹

Med andra ord ska man ska *tänka diagonalt*: för att utröna vilket liv som är bättre för en person ska man endast titta på värdena i attitydmatri-sens huvuddiagonal (följden av tal från matrisens övre vänstra till dess nedre högra hörn). Så vi kan kalla denna idé för ”diagonalprincipen”.

Notera att diagonalprincipen säger att ingen av jag-versionerna be-stämmer välfärdsvärdet av *alla* livsalternativ. Varje jag-version bestämmer endast välfärdsvärdet av *sitt eget* liv genom de polaritetsattityder jag-versionen har till sitt liv.

Det finns förstås en del invändningar mot diagonalprincipen som måste besvaras.

5. INVÄNDNING (1): EXEMPLET MED HELEN KELLER

(DÖVSTUM AMERIKANSK FÖRFATTARE OCH AKTIVIST, 1880–1968)

Anta att den handikappade Helen föredrar ett liv med handikapp fram-för ett utan, eftersom ett liv utan handikapp innehåller så lite av de upp-levelser och prestationer som hon faktiskt bryr sig om. Anta vidare att

9. Dessa principer måste förstås revideras om vi släpper antagande att varje jag-variant har endast en uppsättning attityder och preferenser. Då måste vi ta ställning till hur man ska väga samman olika attityder i ett liv för att fastställa det totala värdet av livet för en person.

hade hon fötts utan handikapp så hade hon föredragit att vara utan handikapp, och hennes positiva attityd till det liv som hon då hade levt skulle ha varit något starkare. I matrisform:

	liv utan handikapp	liv med handikapp
Helen utan handikapp	20	-10
Helen med handikapp	5	15

Diagonalprincipen säger att livet utan handikapp är bättre för Helen än livet med handikapp. Det verkar också följa att Helen *bör önska sig* att hon hade fötts utan handikapp, eftersom detta liv är bättre för henne. Men är det rimligt?

Som svar så är det viktigt att skilja på välfärd och vad man bör önska sig från ett icke uteslutande egennyttigt perspektiv. Vad man bör önska sig i denna mening har inte bara att göra med ens välfärd: det har också att göra med *moraliska*, *politiska* och *konstnärliga* prestationer. Man kanske ibland bör önska sig något mindre välfärd men fler viktiga moraliska, politiska, och konstnärliga prestationer. Helen kan ses som ett bra exempel på detta.

Även om vi i stället tittar på vad man bör önska sig från ett uteslutande egennyttigt perspektiv, så är det tveksamt om *graden* av ens önskan måste sammanfalla med *graden* av välfärd. Hur starkt ens önskan bör vara för ett liv som är bra för en beror på flera faktorer, t.ex. hur svårt det är att realisera detta liv. Man bör ha en starkare önskan för ett bra liv som är realiserbart än för ett liv som är något bättre men inte realiserbart. Så, den handikappade Helen bör önska sitt eget liv mer än livet utan handikapp, som hon ju inte kan förverkliga.

Men måste inte diagonalprincipen fortfarande säga att livet utan handikapp är bättre för Helen? Jo, men endast om *allt annat är lika*. Kom ihåg att jag antog att, allt annat lika, så beror ens välfärd på ens attityder och preferenser. Om man anser att intellektuella och konstnärliga prestationer utgör en viktig del av ens välfärd, så är inte allt annat lika i detta fall, och diagonalprincipen utesluter inte då att Helens faktiska liv med handikapp är bättre för henne i kraft av de imponerande prestationer som hon uppvisade i detta liv.

6. INVÄNDNING (2): BORTSER FRÅN ENIGA PREFERENSER

En annan invändning är att diagonalprincipen krockar med en dominansprincip som kan verka rimlig. Anta att vi har följande fall:

	filosof-liv	pianist-liv
filosof-jag	10	15
pianist-jag	-15	-10

I detta fall så föredrar *både* mitt filosof-jag och mitt pianist-jag pianist-livet framför filosof-livet. Men min diagonalprincip säger att pianist-livet är sämre för mig, eftersom jag kommer att ogilla det men gilla filosof-livet. Diagonalprincipen kommer alltså i konflikt med:

Dominansprincipen

- Om både ditt A-jag och ditt B-jag föredrar A framför B, så är A bättre för dig än B.

I valet mellan diagonalprincipen och dominansprincipen så ger jag upp dominansprincipen. Den gäller inte generellt. Men den gäller förstås när dina jag-versioner har samma positiva, negativa, och neutrala attityder. Mer generellt så gäller den i alla situationer där varje jag-version uppvisar en av attitydkombinationerna (1–5) i tabellen ovan.

Det finns ett oberoende starkt skäl att ge upp dominansprincipen: den genererar *cykliska* välfärdsordningar, som följande schematiska exempel visar:

- A1-jag: A3 föredras framför A1, som föredras framför A2
- A2-jag: A1 föredras framför A2, som föredras framför A3
- A3-jag: A2 föredras framför A3, som föredras framför A1

I detta fall säger dominansprincipen att A1 är bättre för dig än A2, eftersom både ditt A1-jag och ditt A2-jag föredrar A1 framför A2; A2 är bättre för dig än A3, eftersom både ditt A2-jag och ditt A3-jag föredrar A2 framför A3; men A3 är bättre A1, eftersom både ditt A3-jag och ditt A1-jag föredrar A3 framför A1.

7. INVÄNDNING (3): KRÄVER INTE KOMPARATIVT STÖD

Diagonalprincipen står även i konflikt med följande tänkbara nödvändiga villkor för "bättre för":

Principen om komparativt stöd

- A är bättre för dig än B endast om A föredras framför B av ditt A-jag eller ditt B-jag.

I exemplet ovan så föredrar varken ditt A-jag eller ditt B-jag A-livet framför B-livet, men diagonalprincipen säger ändå att ditt A-liv är bättre för dig än ditt B-liv. Detta kan verka bestickande, men principen om komparativt stöd är inte en rimlig princip. Exempelvis så råkar den i trubbel med följande fall:

	A	B
A-jag	-10	-10
B-jag	10	10

Här intar ditt A-jag en negativ attityd till båda liven men A-jaget är indifferent mellan A och B. Ditt B-jag är också indifferent mellan liven, men intar en positiv attityd till dem. Det verkar då självklart att B är bättre för dig än A, men principen om komparativt stöd måste förneka detta.

8. INVÄNDNING (4): IGNORERAR ÅNGER

En annan, mer besvärande, invändning är att diagonalprincipen inte verkar kunna ta ånger i beaktande. Anta att vi har följande schematiska fall att ta ställning till:

	A	B
A-jag	10	5
B-jag	15	10

Diagonalprincipen säger här att A är lika bra för dig som B, men man kan ju tycka att det finns en viktig skillnad mellan dessa liv. Ditt A-jag skulle föredra sitt liv framför alternativet B medan ditt B-jag skulle före-

dra det alternativa livet A. Ska inte denna skillnad ses som tungan på vågen som gör A bättre för dig än B?

Nej, för om vi ser parvis ånger som utslagsgivande, så får vi *cykliska* välfärdsordningar, som följande schematiska exempel visar:

	A	B	C
A-jag	10	5	15
B-jag	15	10	5
C-jag	5	15	10

A är bättre för dig än B, eftersom i relation till A och B så finns det ingen ånger i A men ånger i B. B är bättre för dig än C, eftersom i relation till B och C så finns det ingen ånger i B men ånger i C. C är bättre än A, eftersom i relation till A och C så finns det ingen ånger i C men ånger i A.¹⁰

Finns det inget cyklicitetsfritt sätt att ta i beaktande ånger? Jo, en viktig aspekt av ånger kan fångas, om attityderna antas vara känsliga för den *olust* som ofta tillkommer när man tänker på att man inte fick det man nu föredrar. Attitydvärdena i matrisen kommer då redan att återspegla den ånger-relaterade olusten.

9. SLUTORD

Det finns förstås mycket mer att säga om hur attityder ska bestämma vår välfärd och rimligheten i våra val.¹¹ Framförallt behöver vi ta i beaktande attitydkonflikter över tid i en livsgren. Men jag hoppas att diagonalprincipen kan ses som en första byggsten i en komplett teori – det hoppas i alla fall min nuvarande jag-version.¹²

10. Om vi släpper kravet på parvis jämförelse och i stället definierar ånger relativt det maximala elementet i hela alternativmängden, så får vi ingen skillnad alls, eftersom alla liv då kommer att innehålla samma storlek av ånger (15–10).

11. Jag säger lite mer i Bykvist (2006a), (2006b), (2010).

12. Detta är en utökad version av det föredrag jag höll på Filosofiska föreningen, Filosofiska institutionen, Umeå universitet, den 24 september 2014. Jag är mycket tacksam för de kommentarer jag fick från publiken. Jag är också tacksam för de många hjälpsamma kommentarer som Nils Säfström gav på en tidigare version av denna uppsats.

LITTERATUR

- Anderberg, Thomas. 2014. *Det mesta blir aldrig av*. Stockholm: Norstedt.
- Bricker, Phillip. 1980. "Prudence". *Journal of Philosophy* 77, nr 7, s. 381–401.
- Bykvist, Krister. 2003. "The Moral Relevance of Past Preferences". I *Time and Ethics: Essays at the Intersection*, red. H. Dyke, s. 115–36. Dordrecht: Kluwer Academic.
- . 2006a. "Prudence for Changing Selves". *Utilitas* 18, nr 3, s. 264–83.
- . 2006b. "What Are Desires Good For? Towards a Coherent Endorsement Theory". *Ratio* 19, nr 3, s. 286–304.
- . 2007a. "Comments on Dennis McKerlie's 'Rational Choice, Changes in Values over Time, and Well-Being'". *Utilitas* 19, nr 1, s. 73–77.
- . 2007b. "Violations of Normative Invariance: Some Thoughts on Shifty Oughts". *Theoria* 73, nr 2, s. 98–120.
- . 2010. "Can Unstable Preferences Provide a Stable Standard of Well-Being?". *Economics and Philosophy* 26, nr 1, s. 1–26.
- Carlson, Erik. 1995. *Consequentialism Reconsidered*. Dordrecht: Kluwer Academic.
- Chisholm, Roderick. 1964. "The Descriptive Element in the Concept of Action". *The Journal of Philosophy* 61, s. 613–25.

NO-MIRACLES-ARGUMENTET UTAN VETENSKAPLIG REALISM

Enligt No-Miracles-argumentet är vetenskaplig realism den enda tillfredsställande förklaringen till vetenskapens många lyckade förutsägelser. I den här artikeln presenteras en annan förklaringsstrategi, grundad i en postulerad, stark begränsning av hur underbestämd vetenskapens teorikonstruktioner är i ljuset av tillgängliga empiriska data, och ett argument för att denna strategi ger en bättre förståelse av vetenskapens framgångar.

1. INTRODUKTION

Efter att länge ha överskuggats av andra former av underbestämning, har frågan om vetenskapliga teories underbestämning av tillgängliga empiriska data på senare tid fått mer uppmärksamhet. Stanford (2001, 2006) framför att den typ av underbestämning som han kallar "transient underdetermination" erbjuder det starkaste argumentet mot vetenskaplig realism. I ett mer specifikt vetenskapligt sammanhang argumenterar jag för att vissa indikationer om hur starkt begränsad just denna typ av underbestämning – där kallad "vetenskaplig underbestämning" – är, kan få fysiker att tro på en teori trots brist på empirisk konfirmering (Dawid 2006, 2013).

Denna artikel understryker vikten av vetenskaplig underbestämning för en debatt som hittills förts främst i andra termer. (Härefter även "teoretisk underbestämning".) Begränsningar i underbestämning kan förklara vetenskapens lyckade förutsägelser, och därför försvaras abduktivt. Argumentet detta utmynnar i påminner om det klassiska No-Miracles-argumentet (NMA) (Putnam 1975), som abduktivt härleder vetenskaplig realism från vetenskapens prediktiva framgång, och som ofta anses utgöra det bästa kända argumentet för sådan realism. Vi ska här visa att ett antagande om begränsningar i underbestämning utgör ett väsentligt mer tillfredsställande ramverk för att förklara de många lyckade förutsägelseerna inom vetenskapen, än en fullt utvecklad realism.

Efter en kort analys i avsnitt 2 av några avgörande invändningar mot det klassiska NMA-argumentet, ägnas avsnitt 3 och 4 åt en diskussion om vikten i detta sammanhang av frågan om underbestämning. I avsnitt 5 presenteras sedan den allmänna formen hos ett argument vars struktur liknar NMA, men som är grundat i underbestämningsförhållanden.¹

2. VARFÖR DET KLASSISKA NMA-ARGUMENTET ÄR DÖMT ATT MISSLYCKAS

NMA är ett argument i tre steg. Där hävdas först att vetenskapens många lyckade förutsägelser framstår som ett mirakel om vi inte är realister med avseende på dess teorier. Därefter hävdas att denna realistiska ståndpunkt – d.v.s. antagandet att väletablerade vetenskapliga teories påståendesatser ofta är näraliggande sanningen under en bokstavlig tolkning av dem – verkligen kan ge en tillfredsställande förklaring av vetenskapens framgångar i detta avseende. En abduktiv slutledning ger så i det sista ledet att vetenskaplig realism är sannolik. Vissa filosofer (se t.ex. Musgrave 1985) har i argumentets första led understrukit att det endast är lyckade förutsägelser av genuint *nya* fenomen som kräver en realistisk förklaring; förutsägelser som endast utgör en extrapolering av redan observerade regelbundenheter kan uttryckas i termer av induktionsprincipens giltighet, utan ytterligare antaganden. Denna mer specifika form av NMA kommer antas i det följande.

NMA har ifrågasatts på många sätt – här vill jag fokusera på ett fundamentalt problem, som kan uttryckas som ett dilemma med avseende på precis hur idén om ”vetenskapens många lyckade förutsägelser” förstås. Idén kan tolkas på två olika sätt. En tolkning förstår dessa framgångar som rörande enskilda vetenskapliga teorier: filosofen undrar då varför vissa specifika teorier gör framgångsrika förutsägelser. I enlighet med (Dawid and Hartmann 2017) kallar jag denna tolkning ”enskild-teori-baserad” förståelse av prediktiv framgång. Den version av NMA som är grundad i denna förståelse kallar jag ”enskild-teori-baserad NMA”. Alternativet är att förstå de många lyckade förutsägelserna inom vetenskapens som ett uttryck för vetenskapens framsteg. Under denna förståelse är fenomenet som NMA avser förklara inte de prediktiva framgångarna hos enskilda teorier, utan snarare det faktum att den vetenskapliga

1. Artikelns har översatts från engelska av Henning Strandin, doktorand i teoretisk filosofi i Stockholm.

processen i sig frambringar teorier som gör lyckade förutsägelser med så hög frekvens. Frågan flyttas då till en epistemisk nivå: varför lyckas forskare hitta prediktivt framgångsrika teorier så ofta? Denna tolkning av prediktiv framgång kommer här kallas den "frekvensbaserade" tolkningen, och den motsvarande versionen av NMA kallas "frekvensbaserad NMA". Både den enskild-teori-baserade och den frekvensbaserade tolkningen av NMA förekommer i litteraturen² och ofta är skillnaden inte tydligt uttryckt. När denna skillnad gjorts explicit är det dock inte svårt att visa att ingen av tolkningarna fullt ut kan grunda NMA: enskild-teori-baserad NMA visar sig inte kunna grunda det första ledet i det argument som beskrevs i föregående stycke; den epistemiska tolkningen kan inte grunda det andra ledet.

Om vi ser till enskild-teori-baserad förståelse av vetenskapliga framgångar, kan vi notera att denna är fokuserad på teorierna själva, och inte historiska fakta om forskningsprocessen, vilket gör denna förståelse okänslig för distinktionen mellan *ackommodering* och *konfirmering på basis av nya fenomen*. En teori kan *ackommodera* empiriska data som tagits i beaktning då teorin formulerades, men den kan även *konfirmeras* av lyckade förutsägelser av *nya fenomen*, som inte varit iblandade vid teorins konstruktion. Om man fokuserar på den enskilda teorin och dess egenskaper, snarare än den vetenskapliga processen, så reduceras en teoris vetenskapliga framgångar till att den "bekräftat fenomenen" ("saves the phenomena"). Lyckade förutsägelser av nya fenomen tycks därmed inte ha större konfirmationsvärde än ackommodering. Men, så som påpekats av van Fraassen (1980), så blir förklaringen till prediktiv framgång i detta fall en alldaglig historia. En teoris framgång kan då förstås som en omedelbar konsekvens av två primitiva förhållanden: å ena sidan teorins matematiska struktur (tillsammans med dennas fysiska tolkning), å andra sidan empiriska data. En teoris prediktiva framgångar är då fullgott förklarade om vi pekar på att, *qua* sin matematiska struktur och det sätt denna passats in på befintliga data, så är teorin empiriskt adekvat med avseende på någon datamängd insamlad efter att teorin formulerats. Realisten tycks sakna övertygande skäl för att insistera på att någon ytterligare förklaring till vetenskapliga framgångar behövs.

Realisten kan som ett försök till svar på detta hävda att ett rent analytiskt argument inte kan förklara två kontingenta förhållanden observe-

2. Ett exempel på en epistemisk tolkning av NMA är (Musgrave 1985), ett senare exempel på en analytisk tolkning är (Lyons 2003).

rade i samband med en viss teoris prediktiva framgångar: (i) det faktum att just denna teori snarare än någon annan visat sig vara framgångsrik; och (ii) det faktum att vi har en prediktivt mycket framgångsrik teori inom något område alls.

Bas van Fraassen har framgångsrikt neutraliserat den första invändningen med sitt darwinistiska argument. Att vi i förlängningen har en prediktivt framgångsrik teori snarare än en som inte är framgångsrik på denna punkt är helt enkelt en konsekvens av att teorival sker på basis av prediktiva framgångar. Den andra invändningen, som vilar på observationen att vi har någon prediktivt framgångsrik teori alls, faller på att denna observation inte kan visas vara i behov av någon förklaring inom enskild-teori-baserad förståelse av NMA. Det är förvisso sant att en specifik teori som valts slumpmässigt ur en mängd teorier, som i sin tur inte utmärks för att vara prediktivt framgångsrika, sannolikt inte kommer vara särskilt prediktivt framgångsrik, men detta är inte nog för att bestämma sannolikheten för att en prediktivt framgångsrik teori har formulerats. Den ytterligare information som skulle krävas för att bestämma denna sannolikhet är antalet möjliga teorier, samt antalet teorier forskare har formulerat. Vi kan jämföra situationen med ett lotteri där vinstchansen för en enskild spelare är försvinnande liten, men sannolikheten för att någon alls vinner är stor, helt enkelt för att antalet spelare är stort. På samma sätt kan sannolikheten för att vi har en prediktivt framgångsrik teori göras stor av det rent matematiska förhållandet att tillräckligt många teorier har formulerats. Eftersom enskild-teori-baserad NMA varken förhåller sig till antalet möjliga vetenskapliga teorier eller antalet teorier forskare faktiskt formulerat kan de som anammar denna förståelse av NMA inte ge oss ett scenario under vilket prediktivt framgångsrika teories existens tarvar en förklaring.

Detta problem relaterar direkt till invändningen, framförd ursprungligen av Howson (2000), att NMA innehåller ett base-rate-misstag ("base rate fallacy"). Enligt Howson tror förespråkaren av en enskild-teori-baserad NMA felaktigt att, om det är osannolikt att en falsk teori är prediktivt framgångsrik, och sannolikt att en sann teori är det, så följer det att en teori som granskats och visats vara prediktivt framgångsrik sannolikt är näraliggande sanningen. Som Howson påpekar följer ingen sådan slutsats när inte den initiala sannolikheten ("prior probability", även "bases rate" i detta sammanhang) för att teorin är näraliggande sanningen tagits med i beräkningen. Menke (2014), Henderson (2017), Dawid och Hartmann (2017) har alla argumenterat för att Howsons kritik

specifikt gäller den enskild-teori-baserade förståelsen av NMA, vilken inte erbjuder någon grund för att beräkna den initiala sannolikheten. I enlighet med denna analys kan man ge problemet följande formulering: så länge den relevanta initiala sannolikheten inte har angivits så har en teoris prediktiva framgångar inte visats vara ett problem som kräver en förklaring över huvudtaget.

Än så länge har vi inte lyckats framställa det första av de tre leden som krävs i ett framgångsrikt abduktivt argument från vetenskaplig framgång till vetenskaplig realism: vi har inte ens identifierat ett problem som realism kunde innebära lösningen på. Den frekvensbaserade tolkningen av vetenskapens prediktiva framgång levererar på denna punkt. Frekvensbaserad NMA faller sålunda inte till föga för de argument mot enskild-teori-baserad NMA som vi presenterade ovan. Den höga frekvensen av lyckade förutsägelser av nya fenomen inom vetenskapen är en faktisk aspekt av forskningsprocessen, som inte kan förklaras analytiskt genom att jämföra data med en teoris förutsägelser. Inte heller lider denna formulering av ett base-rate-misstag. Att påvisa en frekvens av prediktiva framgångar innebär – så som visats av Dawid och Hartmann (2017) – att påvisa den initiala sannolikheten som efterfrågas i Howsons analys, och det är därför tydligt vilket förhållande det är som kräver en förklaring. Vi har måhända inte någon grund för förvåning baserad på att någon specifik framgångsrik teori existerar, så länge vi inte känner till antalet misslyckade teorier som formulerats, men om en signifikant andel av formulerade teorier är framgångsrika så kan detta vara förvånande, och tarva en förklaring. Det darwinistiska resonemanget fallerar också i det frekvensbaserade fallet. Eftersom vi nu ser till den faktiska historiska utvecklingen, snarare än en enskild teori och dess egenskaper, så kan vi formulera frågan om vetenskapens framgångar i termer av lyckade förutsägelser av nya fenomen: varför väljer forskare så ofta "rätt" teorier med avseende på empiriska data av en typ som inte fanns i beaktande då teorin formulerades? Ett darwinistiskt svar, som pekar på teoriurvalsprocessen, innebär då inget svar alls.

Antirealisten kan fortfarande förneka att någon förklaring behövs, genom att förneka att naturvetenskapen uppvisar någon signifikant tendens mot lyckade förutsägelser av nya fenomen (se t.ex. Fine 1986). Men detta påstående framstår inte som trovärdigt. Även om det kan verka svårt att framställa en algoritm genom vilken teorier kan räknas, och som tillåter ett kvantitativt omdöme om kvoten av teorier inom ett forskningsfält som gör lyckade förutsägelser av nya fenomen, så är det

uppenbart att den kvoten är högre inom de exakta vetenskaperna än inom andra vetenskapliga områden (så som humaniora) eller andra intellektuella fält (så som science fiction). Någon slags förklaring av denna skillnad kan tyckas motiverad.

Epistemiskt orienterad NMA är därför avsevärt mer framgångsrik, jämfört med den analytiska versionen, för att fastslå att vår förståelse av vetenskapen erbjuder ett allvarligt problem här. Den lyckas med detta genom att utvidga kraven på realismen, från att förklara en enskild teoris prediktiva framgångar, till att förklara ett forskningsfältets prediktiva framgångar med avseende på nya fenomen.

Tyvärr höjer den därmed också tröskeln till en nivå som realismen inte kan ta sig över. Laudan (1981) och Fine (1986) var bland de första att påpeka att, även om realisten har rätt i att en teoris approximativa sanning kan härledas från att den varit prediktivt framgångsrik i det förflutna, så ger tillgänglig evidens oss inga skäl att tro att denna sanning omfattar också de teoretiska delar som är väsentligt involverade i förutsägelse av fenomen vi ännu ej undersökt. Problemets källa är att inte ens den mest övertygade realisten kan påstå att framgångsrika teorier är sanna i precis varje avseende. Vetenskapshistorien vittnar om att även de mest framgångsrika teorierna kan förväntas visa sig empiriskt inadekvata förr eller senare. Sådana misslyckanden kommer då leda till en formulering av nya teorier, som enligt realisten ligger än närmare sanningen. Detta synsätt möjliggörs hos den vetenskapliga realisten genom att framgångsrika teorier tillskrivs approximativ, snarare än fullständig, sanning. Syftet med detta är alltså att ge rum för *misslyckade* framtida förutsägelser utifrån en approximativt sann teori. Det tycks mycket svårt att försvara att precis samma egenskap hos teorier även förklarar deras prediktiva framgångar. Sålunda tycks den vetenskapliga realisten inte vara bättre positionerad än antirealisten, för att förklara forskares förmåga att framgångsrikt förutsäga nya fenomen.

3. ARGUMENTET FÖR VETENSKAPLIG UNDERBESTÄMNING

Analysen så här långt indikerar att NMA misslyckas på grund av ett ofrånkomligt och fatalt dilemma: om argumentet endast förhåller sig till de prediktiva framgångarna hos enskilda vetenskapliga teorier så finns där inte något "mirakulöst" att förklara; om det vänder sig snarare till benägenheten för prediktiva framgångar inom ett forskningsfält, så kan ett intressant problem identifieras, men den vetenskapliga realis-

men kan inte erbjuda någon lösning på det. I den här artikeln presenteras ett alternativ till vetenskaplig realism, som kan ge ett tillfredsställande svar på det senare, frekvensbaserade, problemet.

Vi har redan introducerat frågan som är avgörande för att nå vårt mål. Anledningen till att enskild-teori-baserad NMA inte kan avancera ligger i det faktum att inget sägs om antalet möjliga vetenskapliga teorier som kan passas in på existerande empiriska data. Argumentet som ska presenteras här implicerar att ett underliggande antagande om att antalet möjliga alternativa teorier är begränsat (härefter: "antagandet om en begränsning av vetenskaplig underbestämning"), är avgörande för NMA:s intuitiva lockelse, och även för de avancerade försöken att rädda argumentet från de invändningar som presenterades i det föregående avsnittet. Men det ska visa sig att detta antagande erbjuder en förklaring av forskningsprocessens prediktiva framgångar som är oberoende av vetenskaplig realism.

4. INTUITIONSBASERAD VETENSKAPLIG REALISM, OCH DESS EFTERFÖLJARE

Denna analys börjar med en fråga som tycks sakna omedelbar anknytning till vetenskaplig underbestämning. Givet den någorlunda rättframma invändning som presenterades i avsnitt 2, hur kommer det sig att NMA ändå framstår som intuitivt mycket plausibelt?

För att närma oss ett svar, låt oss föreställa oss en värld som är enklare än den vi faktiskt bebor. I denna enklare värld kan alla fysiska fenomen beskrivas genom en tillämpning av de klassiska fysiska lagar som styr observerbara processer även på ej observerbara ting. En vetenskaplig realist i denna "klassiska värld" kan omfamna "klassikalitetsvillkoret": alla verkliga objekt beter sig som de observerbara objekten. I denna värld kan de objekt en teoretiker kan införa, om hon vill undvika falska teorier, endast skilja sig i *storlek* från de observerbara, och teoretikern måste reproducera alla nya mikrofysiska data i enlighet med detta. Realism grundad i det strikta klassikalitetsvillkoret introducerar sålunda mycket starka begränsningar på vetenskaplig teorikonstruktion, genom att exkludera alla typer av teorier som passar in på existerande empiriska data men inte har en strikt klassisk tolkning. En strikt klassisk realist kan förklara prediktiva framgångar med hänvisning till de begränsningar i teorikonstruktion som gäller under denna strikta realism. Det kan fortfarande hända att nya typer av oförutsagda fenomen

observeras, som en konsekvens av att teorin inte omfattar alla relevanta mikroskopiska objekt (d.v.s. teorin är inte exakt sann, i bemärkelsen fullständig), men inga nya teoretiska principer introduceras, och teorin kan hävdas vara approximativt sann. Ett ytterligare antagande om sällsyntheten hos nya partiklar kan sedan grunda en förklaring av prediktiv framgång i detta fall. NMA fungerar i detta scenario därför att strikt klassisk realism introducerar en stark begränsning i hur existerande data underbestämmer teorikonstruktion.

Under lång tid syftade vetenskaplig teorikonstruktion till att bibehålla en svagare version av klassisk-värld realism, såtillvida att observerbara tings karaktär överfördes på ej observerbara objekt närhelst detta var möjligt.³ De framgångar som uppnåddes på detta sätt⁴ kunde då ges en – mer eller mindre övertygande – realistisk förklaring, i enlighet med den strategi som skissades i det klassiska-värld-exemplet ovan. 1900-talets fysik försvårade dock tillvaron avsevärt för den vetenskapliga realisten. Moderna fysiska teorier, så som den särskilda och den allmänna relativitetsteorin, kvantmekaniken och gaugeteori, överger så mycket av våra intuitiva föreställningar om fysiska objekt, rum och tid att det som återstår kan förväntas försvinna i framtida teorirevisioner. Klassiska intuitioner om fysiska objekts beteende spelar därmed inte en betydande roll i konstruktionen av vår tids fundamentala fysik. En konsekvens av detta är att även den vetenskapliga realismen övergett sina intuitiva rötter. Moderna formuleringar av NMA hänvisar inte på ett direkt sätt till intuitiva föreställningar om fysiska objekt⁵, utan fokuserar på en abstrakt relation mellan ett påståendes empiriska adekvans och dess sanning. Vi har dock redan redogjort för baksidan av detta: om vetenskaplig realism helt retirerar till denna kärndefinition, så förlorar den sin normativa auktoritet med avseende på vetenskaplig teoriformulering. Och i avsaknad av detta verktyg för att epistemologiskt förklara framgångsrika teorival, så faller realismen till föga för den invändning som skisserades i avsnitt 2.

En realist som vill rädda NMA måste därmed söka andra teoretiska

3. Tidiga påståenden om en entydig relation mellan observerbara fenomen och teorin som var kompatibel med dem, så som i Newtons "härledning från fenomenen", kan rimligen sägas vara till en hög grad grundad i detta tillvägagångssätt.

4. Kinetiska gasteorin och tidiga atommodeller utgör framstående exempel.

5. Realismen vid 1900-talets början var tvärtom fortfarande grundad i intuitiva föreställningar om fysiska objekt. När Duhem (1906) invände mot en realistisk tolkning av vetenskapliga teorier, riktade han dessa invändningar i första hand mot att klassiska idéer om fysiska objekt överfördes på mikroförhållanden.

egenskaper än överensstämmelse med en naivt klassisk ontologi, som kan leda teoriformulering mot sanna teorier genom att begränsa vetenskaplig underbestämning. De populäraste kandidaterna (se t.ex. Boyd 1984) är enkelhet, avsaknad av *ad hoc*-karaktär, universalitet och förutsägelsekraft,⁶ som alla tycks spela en betydande roll inom konstruktion och val av vetenskapliga teorier. Dessa kan utnyttjas av en realist på följande sätt. Man ser att (i) framgångsrika teorier brukar uppfylla de ovan nämnda villkoren och (ii) teorier som konstruerats med hänsyn till dessa villkor tenderar att vara framgångsrika. Båda dessa observationer förklaras bäst under antagandet att sanna beskrivningar av verkligheten uppfyller dessa, eller väsentligt likartade, villkor. Om vi då söker efter teorier som uppfyller de angivna villkoren så ökar våra chanser för att hitta sådana som är näraliggande sanningen.

En noggrann betraktelse av detta argument ger dock vid handen att dess framgång inte är betingad av antagandet om realism. Prediktiva framgångar förklaras här av allmänna villkor för vetenskaplighet, så som universalitet, vilka styr forskningsprocessen och därmed begränsar underbestämningen vid teorikonstruktion. Argumentet vilar enbart på antagandet att våra nuvarande teorier på ett eller annat sätt återger universaliteten, eller liknande egenskaper, hos den sanna teorin. Det förblir därmed giltigt oavsett om teoriernas substantiella innehåll i övrigt ligger nära sanningen eller inte.

Därtill är det inte alls självklart att införandet av villkor som universalitet är tillräckligt för att styra forskare mot prediktivt framgångsrika teorier. Universalitet är inte en tillförlitlig indikator för prediktiv framgång om inte mängden teorier som är kompatibla med tillgängliga data och därtill åtminstone lika universella som den teori som övervägs inte är signifikant begränsad – d.v.s. om den teoretiska underbestämningen inte begränsas i tillräcklig utsträckning av de postulerade egenskaperna tillsammans med data. Dessa antaganden om signifikanta begränsningar i teoretisk underbestämning utgör alltså den faktiska kärnan hos varje förklaring av vetenskapliga framgångar, som är baserad på att forskares teorival styrs av vissa förteoretiska postulater. Den vetenskapliga realistens hänvisning till ytterligare villkor för teorival är alltså baserat på en felaktig föreställning om var fokus i dennes argument ligger. Förklaringskraften hos dessa villkor är kort sagt oberoende av

6. En teoris förutsägelsekraft är omfattningen av dess förutsägelser av framtida empiriska data. Detta måste tydligt särskiljas från dess prediktiva framgångar, vilket rör graden till vilken dess förutsägelser bekräftas vid observation.

huruvida de valda teorierna är approximativt sanna eller ej. Den är i stället grundad helt i den begränsning av teoretisk underbestämning som villkoren implicerar.

5. ETT ANNAT SÄTT ATT FÖRKLARA VETENSKAPLIG FRAMGÅNG

En bild framträder, med ett nytt perspektiv på frågan om vetenskapens prediktiva framgångar. Medan realismen förlorar sin förklaringskraft i de mer abstrakta och icke-intuitiva vetenskapliga sammanhangen, så förblir frågan om underbestämning central för vår förståelse av vetenskaplig framgång. Sökandet efter en förklaring till dessa framgångar bör därför fokusera på en analys av begränsningarna av vetenskaplig underbestämning, utan att insistera på en genomgående vetenskaplig realism.

Jag föreslår följande tillvägagångssätt. Graden i vilken vetenskaplig framgång är förväntad antas kunna uttryckas i termer av en forskares möjligheter till att formulera en teori i ljuset av (i) tillgängliga empiriska data och (ii) vissa godtagbara förteoretiska postulat om vetenskapliga teoriers karaktär. I de fall där teoriformulering är avsevärt underbestämd av dessa faktorer så är chansen att en vald teori framgångsrikt förutsäger nya fenomen minimal. Ju mer begränsade forskarens alternativ är, d.v.s. ju mindre den teoretiska underbestämningen är i ljuset av dessa antaganden, desto större är chansen att en vald teori gör lyckade förutsägelser av nya fenomen.

I denna bild förklaras vetenskaplig framgång på basis av två avgörande antaganden: att vissa allmänna förteoretiska postulat inom vetenskapen är godtagbara; och att dessa postulat begränsar vetenskaplig underbestämning. Låt oss formulera dessa antaganden något mer noggrant. Antagandet som gäller begränsningar av vetenskaplig underbestämning avgränsar mängden av teorier som är kompatibla med tillgängliga data. Detta antagande är bara meningsfullt givet ett ramverk som specificerar vilken mängd teorier vi talar om. Vi kan förvänta oss att ett sådant ramverk tillhandahålls av förteoretiska postulat, som styr teorikonstruktion. Eftersom antagandet om begränsningar i teoretisk underbestämning ska förklara vetenskaplig framgång måste det vila på en robust, objektiv grund. De förteoretiska postulat som tillhandahåller det nödvändiga ramverket kan därför inte utgöras bara av pragmatiskt motiverade tumregler. Dessa postulat måste i stället relatera till världen på ett sådant sätt att vi kan spåra vetenskapens prediktiva framgångar tillbaka till några egenskaper hos världen själv. Vi antar därför att en

sann – eller mer försiktigt en empiriskt adekvat – teori existerar, som satisfierar villkor sådana att de utgör en stabil kärna av de förteoretiska postulat som görs inom nutidens vetenskap.

Vi kan nu formulera påståendet om begränsning av teoretisk underbestämning: ramverket, som definieras av den mängd förteoretiska postulat inom vetenskapen som vi just antagit är empiriskt adekvata, tenderar att avsevärt begränsa underbestämningen av teorikonstruktion inom de exakta naturvetenskaperna. Denna formulering flyttar diskussionen till en helt ny nivå, och utgör den avgörande begreppsliga innovationen i det föreslagna angreppssättet: de faktorer som förklarar vetenskaplig framgång söks inte i den faktiska världen och dess likheter med våra nuvarande teorier, utan i den bredare sfären av möjligheter. Vi efterfrågar information om totaliteten av vetenskapliga system som är kompatibla med de empiriska data vi just nu besitter, snarare än bara om den enda sanna teorin.

Analysen som presenterats här står inför det uppenbara problemet att vi inte vet vilken denna mängd av teoretiska möjligheter är. All kunskap om denna mängd måste ha sin grund i påståendet att avgränsningar av den krävs för en förklaring av vetenskapens lyckade förutsägelser av nya fenomen. Ett argument av detta slag utgör en abduktiv slutledning likt den som vetenskapliga realister förlitar sig på i det klassiska NMA: givet att antagandet om begränsningar i vetenskaplig underbestämning tillhandahåller den mest rättframma förklaringen av vetenskapens prediktiva framgång, och givet att ingen annan tillfredsställande framställning tycks finnas, följer det abduktivt att det existerar signifikanta begränsningar i vetenskaplig underbestämning inom de exakta vetenskaperna.

LITTERATUR

- Boyd, R. 1984. "The Current Status of Scientific Realism". I *Scientific Realism*, red. J. Leplin. Berkeley: University of California Press.
- Dawid, R. 2006. "Underdetermination and Theory Succession from the Perspective of String Theory". *Philosophy of Science*, 73/3, s. 298–322.
- . 2013. *String Theory and the Scientific Method*, Cambridge University Press.
- Dawid, R. och S. Hartmann. 2017. "The No-Miracles Argument without the Base Rate Fallacy". Först i *Synthese* online.
- Duhem, P. 1906. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press, 1954.

- Fine, A. 1986. "The Natural Ontological Attitude". I A. Fine, *The Shaky Game: Einstein, Realism, and the Quantum Theory*, s. 118–19. Chicago: University of Chicago Press.
- Howson, C. 2000. *Hume's problem: Induction and the justification of belief*. Oxford: The Clarendon Press.
- Laudan, L. 1981: "A Confutation of Convergent Realism". *Philosophy of Science* 48, s. 19.
- Lyons, T. D. 2003: "Explaining the Success of a Scientific Theory". *Philosophy of Science* 70, s. 891.
- Musgrave, A. 1985. "Realism versus Constructive Empiricism". I *Images of Science: Essays on Realism and Empiricism*, red. P. M. Churchland och C. A. Hooker. Chicago: University of Chicago Press.

VÄXELSEMANTIK (SWITCHER SEMANTICS)

1. INTRODUKTION

Vi börjar med en historia: historien om Peter, Paul och Mary. Mary vill gärna ha någonting kallt att dricka; hon hoppas att det finns öl i kylan. Paul säger till henne att det finns öl i kylan. Mary litar på Paul och tror att det han säger är sant. Hon går till köket och öppnar kylskåpet. Men hon blir besviken – det finns inget öl kvar. Hon ropar åt Paul: ”Nehej, du, det finns inget öl! Peter måste ha druckit upp allt i går kväll.” Ledmotivet i den här historien är att det finns öl i kylskåpet – det är det Mary hoppas, Paul påstår, Mary först tror men sedan upptäcker vara falskt. Det finns otaliga sådana saker, saker man kan hoppas, tror, tvivlar på eller önska vara sanna. Filosofer har fäst sig vid denna anmärkningsvärda enhetlighet i mångfalden av former i vilka vårt liv som tänkande och kommunicerande varelser manifesterar sig. Den centrala idén är att det finns en unik slags entitet som förenar fenomen *prima facie* så olika som en förhoppning och ett påstående: propositionen.

I historien om Peter, Paul och Mary är det propositionen att det finns öl i kylskåpet som påstås, hoppas, tros och inte tros vara sann. Men vad är då en proposition? Detta har filosofer givit olika svar på, men den klassiska modellen använder sig av en idé som finns redan hos Leibniz: idén om möjliga världar. Världen är som den är, men den hade kunnat vara annorlunda. En möjlig värld är ett sätt på vilket världen hade kunnat varit annorlunda. Närmare bestämt: en möjlig värld är en maximal möjlighet, ett sätt på vilket världen hade kunnat vara annorlunda där inget lämnas obestämt. Enligt den moderna modellen, som först används av Wittgenstein i *Tractatus*, delar en proposition mängden av möjliga världar i de där propositionen är sann och de där den är falsk (och kanske också de där den är varken sann eller falsk). Enligt den enklaste modellen är då propositionen själv en mängd av möjliga världar: mängden där den är sann.

Under senare tid har dock själva idén att det skulle finnas en förenande typ av entiteter bakom alla dessa fenomen ifrågasatts. Å ena sidan finns kritik mot, och problem med, den klassiska modellen av propositionen som en mängd av möjliga världar. Å andra sidan har en hel del alternativa modeller föreslagits. Dessa förslag har vanligtvis kommit

till för att lösa specifika semantiska problem. Problemen utgörs oftast av satser som verkar ha semantiska egenskaper som utesluter att deras betydelse är en klassisk proposition. Förslagen har lett till en veritabel djungel av entiteter av olika slag som används som propositioner i en eller annan bemärkelse. Utvecklingen förstärks av en växande konsensus att det helt enkelt inte finns en enda entitet som kan göra allt vi förväntar oss att propositionen gör.

Vårt projekt är att försvara den klassiska propositionen. Eftersom kritiken oftast är semantiskt motiverad, kommer försvaret också vara semantiskt. Vi har utvecklat en semantik – växelsemantik – som kan användas för att förklara problematiska semantiska egenskaper utan att överge den klassiska propositionsmodellen. I den här uppsatsen ska vi illustrera strategin genom att fokusera på ett exempel: temporala operatörer. Dem ska vi behandla i avsnitt 4. Men först ska vi säga några ord om propositioner i avsnitt 2. Sedan ska vi ge en formell definition av växelsemantik (avsnitt 3). Avslutningsvis ska vi nämna några fler potentiella tillämpningar av växelsemantik.

2. OM PROPOSITIONER

I introduktionen sa vi att det enligt en växande konsensus inte finns en enda entitet som kan göra allt vi förväntar oss att propositionen gör. Vad är då detta? Traditionellt förväntar vi oss åtminstone följande av propositioner: Att det är de som

- har sanningsvärden – dvs. att de är sanna eller falska (och kanske något tredje),
- är nödvändiga eller möjliga,
- är (mer eller mindre) sannolika,
- är innehållet i påståenden och
- är innehållet i de s.k. propositionella attityderna (dvs. mentala tillstånd som övertygelser, önskningar, hopp osv.).

Redan på 80-talet formulerar David Lewis en pessimistisk inställning till denna traditionella lista när han skriver: "Uppfattningen vi associerar med ordet 'proposition' utgörs möjligen av ett virrvarr av motstridiga desiderata" (Lewis 1986, s. 54).¹ Sedan dess har pessimismen bara spritt sig genom litteraturen.

1. Engelska originalet lyder: "The conception we associate with the word 'proposition' may be something of a jumble of conflicting desiderata."

Exemplet vi ska fokusera på är temporala operatorer, alltså uttryck som 'alltid', 'ibland', och 'aldrig'. Ta följande två satser:

- (1) a. Det regnar där jag är.
- b. Det regnar där jag är nu.

Vi verkar säga precis samma sak oavsett om vi använder oss av (1a) eller (1b). Om (1a) är sann när vi yttrar den, då är (1b) också sann. Men någonting intressant händer när vi bildar mer komplexa satser med hjälp av temporala operatorer:

- (2) a. Det är alltid så att det regnar där jag är.
- b. Det är alltid så att det regnar där jag är nu.

Dessa två satser betyder rätt så olika saker – fast det enda vi har gjort är att byta ut sådana komponenter mot varandra som verkade ha precis samma betydelse.

Detta strider mot den s.k. kompositionalitetsprincipen. Kompositionalitets-principen anses vara helt avgörande för semantiken för naturliga språk. Detta för att om man kan visa att ett naturligt språk följer principen då kan man åtminstone i princip ge en förklaring till hur talarna kan lära sig språket utan att behöva memorera oändliga listor av mer och mer komplexa uttryck, hur de lätt och effektivt kan tolka satser som de aldrig tidigare ha hört eller läst, osv. Senare ska vi ge en formell definition av kompositionalitetsprincipen, men just nu räcker det med en informell formulering. Det finns två versioner av principen: funktionsversionen och substitutionsversionen. De kan formuleras så här:

- (K_F) Betydelsen hos ett komplext uttryck är en funktion av betydelsen hos dess delar och deras sammansättning.
- (K_S) Om man byter ut en del av ett komplext uttryck mot en annan del med samma betydelse kommer det nya komplexa uttrycket ha samma betydelse som det ursprungliga.

Men detta tycks alltså inte gälla när det gäller satserna i (2). Det verkar finnas tre möjliga reaktioner: Vi kan antingen säga

- 1. att (1a) och (1b) inte har samma betydelse, eller
- 2. att kompositionalitetsprincipen inte gäller, eller
- 3. att (1a) och (1b) har samma betydelse när de förekommer oinbäddade men olika betydelser när de förekommer inom räckvidden för en temporal operator.

Framstående filosofer har valt olika här. David Kaplan t.ex. valde alternativ 1, medan Michael Dummett och David Lewis valde alternativ 3 (och därmed automatiskt också alternativ 2).

Enligt Kaplan uttrycker (1a) en s.k. temporal proposition, alltså en proposition som kan vara sann vid somliga tidpunkter och falsk vid andra. (1b) däremot uttrycker en evig proposition, dvs. en proposition som antingen är sann vid alla tidpunkter, eller falsk vid alla tidpunkter. Kaplan är därmed en företrädare för temporalism: uppfattningen att somliga propositioner (men inte nödvändigtvis alla) är temporala.

Kaplan argumenterar för temporalism bl.a. med hjälp av det s.k. operator-argumentet: "Det måste finnas innehåll som är tids-neutrala vilka temporala operatorer kan ta som argument. Annars skulle temporala operatorer vara överflödiga" (Kaplan 1989, s. 503).² Ta följande satser:

- (3) a. Det regnar vid tidpunkten *t*.
- b. Det är alltid så att det regnar vid tidpunkten *t*.

(3a) uttrycker en evig proposition, och här kan 'alltid' inte göra någon skillnad. Om det bara fanns eviga propositioner att använda 'alltid' på skulle 'alltid' aldrig kunna göra skillnad. Och då skulle vi egentligen inte behöva ett sådant uttryck i språket alls.

Men även om godtagandet av temporala propositioner möjligen löser problemet med satserna i (2) så medför det allvarliga nya problem. Det finns en pågående diskussion i ämnet, men här ska vi nämna bara ett problem som hittills inte har fått tillräcklig uppmärksamhet.³ Problemet är att temporala propositioner leder till ett orimlig mångfaldigande av övertygelser. Anta att propositioner är innehållet i övertygelser, och att Kathrin tror att (4) är sann:

- (4) Peter är i Stockholm just nu.

Enligt Kaplan betyder det att Kathrin håller en evig proposition för sann. Men om Kathrin tror att (4) är sann, då tror hon också att (5) är sann:

- (5) Peter är i Stockholm.

Enligt Kaplan uttrycker (5) en annan, temporal proposition. Men över-

2. Engelska originalet lyder: "There must be contents that are neutral with respect to time for temporal operators to take as arguments. Otherwise temporal operators would be redundant."

3. En tidig kritik riktades mot Kaplans temporalism av Mark Richard (1981), medan Kaplan försvarats mot Richards typ av kritik av bl a Francois Recanati (2007) och Brit Brogaard (2012).

tygelsen Kathrin kan uttrycka genom att yttra antingen (4) eller (5) verkar vara en och samma. Kaplan däremot måste säga att det Kathrin tror om var Peter befinner sig består av två övertygelser.

Konsekvenser som denna kan undvikas om vi i stället använder oss av Dummetts distinktion mellan hävdande-innehåll och ingrediensmening ("assertoric content and ingredient sense"; Dummett 1973, s. 446–47, Dummett 1991, s. 47–49). Hävdande-innehåll är betydelsen en sats har när den är oinbäddad, dvs. när den används för sig själv för att göra ett påstående. Satsens ingrediensmening däremot är dess bidrag till betydelsen hos mer komplexa satser den utgör en beståndsdel av. David Lewis (1980, s. 95–96) gjorde en liknande distinktion mellan det som sägs ("what is said") och semantisk värde ("semantic value"). Distinktionen utgör en avgörande beståndsdel i vår växelsemantik. I nästa avsnitt ska vi förklara vad en växelsemantik är, och i avsnittet därpå ska vi tillämpa semantiken på temporala operatorer.

3. VÄXELSEMANTIK

I detta avsnitt kommer vi att ge en formell definition av växelsemantik. Denna definition förutsätter en formell definition av kompositionalitet, vilken vi presenterar nedan. Vi kommer att arbeta med funktionsversionen av kompositionalitet. Det formella ramverket för denna definition kommer att vara en modifiering av det algebraiska ramverk för kompositionalitet som utarbetades av Wilfrid Hodges (2001).⁴ Vi kommer att kalla kompositionalitet, så definierad, för standard-kompositionalitet. Skälet är vi kommer att presentera en generalisering av standard-kompositionalitet, kallad "generell kompositionalitet". Generell kompositionalitet tillåter semantisk växling i reglerade former, och omvänt krävs av en växelsemantisk semantik att den är generellt kompositionell.

Hodges ramverk är algebraiskt: syntaxen har formatet av en *algebra*, dvs. en mängd språkliga *uttryck* samt en mängd *operationer* på dess uttryck. Operationerna kan kombinera uttryck och/eller modifiera uttryck, t.ex. morfologiskt. För att illustrera idén kan vi anta en operation σ'_p som motsvarar enkel predikation: sammansättning av en nominalfras i singularis, som ett egennamn, med ett verb till en sats i presens. Vi använder ' _ ' för att markera ordmellanrum:

4. Detta ramverk har använts bl.a. i Pagin 2003, Pagin och Westerståhl 2010a, 2010b, 2010c.

- (6) $\sigma'_p('Kalle', 'springa') = 'Kalle' +_+ 'springer' = 'Kalle\ springer'$

För ett enskilt språk L tänker sig Hodges en ändlig mängd A'_L atomära (enkla) uttryck i L , t.ex. enskilda ord eller ordstammar samt en ändlig mängd Σ'_L operationer av typen σ'_p . Tanken är att den totala mängden E'_L av grammatiska uttryck i språket, inklusive hela satser, är precis det som kan bildas genom att tillämpa operationerna i Σ'_L på uttrycken i A'_L , på resultaten av dessa tillämpningar, osv., dvs. på det *transitiva höljet* av Σ'_L tillämpat på A'_L . Vi har då en s.k. *uttrycksalgebra* $U_L = (E'_L, A'_L, \Sigma'_L)$, uttrycksalgebran för L . De syntaktiska operationerna är partiella: σ'_p är t.ex. inte definierad för paret av uttryck $\langle 'Kalle', 'Kalle' \rangle$.

Vi kan emellertid inte stanna vid uttrycksalgebran som syntaxen för ett språk. Skälet är att naturliga språk typiskt är semantiskt flertydiga. Dels har vi lexikal flertydighet, av typen 'bank' i meningen flodbank och 'bank' i meningen finansiell institution (eller kontor för en sådan). Dels har vi också syntaktisk flertydighet, av typen

- (7) Jag sköt en älg i min trädgård.

(Var det älgan som befann sig i trädgården när den blev skjuten, eller talaren som befann sig i trädgården när hen sköt?)

Eftersom tvetydigheter av detta slag ger upphov till olika tolkningar, så behöver vi en disambiguerad syntaktisk nivå, där alla flertydigheter är eliminerade. Flertydigheten i (7) beror på om prepositionsfrasen (PP) 'i min trädgård' är en direkt del av verbfrasen, dvs. hör samman med 'sköt', eller en del av nominalfrasen, dvs. hör samman med 'älg'/'en älg'. Om vi bygger upp (7) genom att först kombinera prepositionsfrasen inom nominalfrasen och sedan denna kombination 'en älg i min trädgård' med verbet, så får vi den första av de två tolkningarna.

Det hela beror alltså på vilka operationer som används på vilka argument, i användningen av uttrycksalgebran. Detta reflekteras i de meta-språkliga uttryck som denoterar dessa operationer. Ett sådan meta-språkligt uttryck hittar vi på vänster sida om '=' i (6), dvs. ' σ'_p ('Kalle', 'springa')'. Hodges kallar dessa "grammatiska termer". De grammatiska termerna kombineras till mer komplexa grammatiska termer, på ett sätt som reflekterar kombinationen av uttryck, förutom att dessa kombinationer aldrig är flertydiga. För kombinationer av grammatiska termer finns term-operationer som motsvarar uttrycksoperationerna i Σ'_L . Till sammans bildar dessa en *grammatisk termalgebra* $GT_L = (T_L, A_L, \Sigma_L)$ för ett språk L , där T_L är mängden av grammatiska termer, A_L är mängden

av atomära grammatiska termer, och Σ_L är mängden av operationer i termalgebran.

En av skillnaderna mot Hodges version är att vi här tar termalgebran som grundläggande. Det ansluter närmare till gängse syntaktisk teori och motsvarar ungefär *djupstruktur* i Chomskys *Principles and Parameters* (Chomsky och Lasnik 1993) och en kombination av *merge* och *copy* inom syntaktisk *Minimalism* (se Hornstein, Nunes och Grohmann 2005). De grammatiska termerna avbildas på uttrycken i språket av en funktion V . Vi har t.ex. $V(\sigma'p$ ('Kalle', 'springa')) = 'Kalle springer'. P.g.a. syntaktisk flertydighet kommer olika termer att avbildas på samma uttryck.

Termerna avbildas också på semantiska värden (meningar, betydelser) av en semantisk funktion μ . Exakt vad som utgör meningar spelar stor roll för våra tillämpningar av ramverket, men inte för det abstrakta perspektiv i vilket kompositionalitet och växelsemantik definieras. Vi antar en mängd M av meningar, samt en mängd R_M av operationer från meningstuplar (t.ex. meningspar) till meningar.

Vi har nu de ingredienser som behövs för en mer exakt formulering av den funktionella kompositionalitetsprincipen:

(KPF) För varje n -ställig operation $\sigma_i \in \Sigma_L$ finns en funktion $r_i \in R_M$ sådan att det för alla termer $t_1, \dots, t_n$, för vilka σ_i är definierad, gäller att

$$\mu(\sigma_i(t_1, \dots, t_n)) = r_i(\mu(t_1), \dots, \mu(t_n))$$

(om μ är definierad för $\sigma_i(t_1, \dots, t_n)$).

Här är det antaget att om en komplex term är meningsfull, så är också dess deltermer meningsfulla (Hodges kallar detta "Domänprincipen").

Detta är alltså funktionsvarianten av det vi kallar "standard-kompositionalitet". Själva grundidén för växelsemantik är dels att en semantik kan innehålla mer än en semantisk funktion, dels att den semantiska funktion som tillämpas på en underterm kan vara en annan än den som tillämpas på övertermen. T.ex. kan vi ha de semantiska funktionerna μ_1 och μ_2 och den komplexa termen ' $\sigma(t, u)$ ' för vilka det gäller att $\mu_1(\sigma(t, u)) = r(\mu_1(t), \mu_2(u))$. Vi ser här en växling i andra argumentet för σ . Värdet hos μ_1 för den komplexa termen beror på värdet hos μ_2 för undertermen ' u '. Detta är vad vi kallar "semantisk växling" (eller "semantiskt byte"). I växelsemantik beror växlingen på argumentplatsen, inte på vilken term som fyller platsen. Vi säger att paret $(\sigma, 2)$ är en växlare.

En *växelsemantik* är ett semantiskt system $S = (S, \mu_0, \Psi)$, där S är en mängd semantiska funktioner $\{\mu_0, \mu_1, \dots, \mu_n\}$, μ_0 är det *designerade*

elementet i S , och Ψ är en *växlings-funktion*. Det krävs att S har ändligt många element. Att μ_0 är det *designerade* elementet innebär att det alltid tillämpas på icke inbäddade termer. Dvs., semantisk tolkning med S börjar med tillämpning av μ_0 . $S(t) = \mu_0(t)$. Växlingsfunktionen Ψ anger växlingsschemat för S . Det är en funktion som tar som argument en trippel bestående av en medlem av S , en syntaktisk operation, samt en argumentplats, och ger som värde ett element av S : $\Psi : S \times \Sigma \times \mathbb{N} \rightarrow S$. I exemplet ovan har vi $\Psi(\mu_1, \sigma, 2) = \mu_2$.

Två saker till krävs för att ett semantiskt system S ska vara en växel-semantik. Det första är att semantisk växling faktiskt äger rum i S . Det vill säga att växelfunktionen Ψ måste vara så beskaffad att $\Psi(\mu_0, \sigma, i) \neq \mu_0$ för någon syntaktisk operation σ i Σ och någon argumentplats i .

Det andra kravet är att S ska vara *generellt kompositionellt*. Vi kan konstatera att en växelsemantik inte är standard-kompositionell. Detta beror på kravet att växling äger rum. I exemplet har vi $\mu_1(\sigma(t, u)) = r(\mu_1(t), \mu_2(u)) \neq r(\mu_1(t), \mu_1(u))$ (förutsatt att $\mu_1(u) \neq \mu_2(u)$). Vi kan nu definiera generell kompositionalitet:

(GFK) Låt (T_L, A_L, Σ_L) vara termalgebran för ett språk L och M en meningsdomän och låt $S = (S, \mu_0, \Psi)$ vara ett semantiskt system med domän T_L och värdeförråd M . S är generellt kompositionellt om det för varje $\sigma \in \Sigma_L$ och varje $\mu \in S$ finns en operation $r_{\sigma, \mu}$ sådan att om $\mu(\sigma(t_1, \dots, t_n))$ är definierad det gäller att

$$\mu(\sigma(t_1, \dots, t_n)) = r_{\sigma, \mu}(\mu_{j_1}(t_1), \dots, \mu_{j_n}(t_n))$$

där det gäller att $\mu_{j_i} = \Psi(\mu, \sigma, i)$, $1 \leq i \leq n$.

Man kan notera att detta är en generalisering av standard-kompositionalitet, för i det specialfall där $S = \{\mu_0\}$ erhåller vi standard-kompositionalitet som resultat.

Historiskt sett introducerades semantisk växling av Frege (1948). T.ex. föreslog Frege den relevanta semantiska funktionen i s.k. indirekta kontexter är innebörd (Sinn), inte betydelse (Bedeutung). Indirekta kontexter är t.ex. kontexter i indirekt tal. I satsen

(8) Trump säger att kol är rent.

befinner sig delsatsen 'kol är rent' i en indirekt kontext. Betydelsen (sanningsvärdet) för (8) beror på innebörden hos 'kol är rent', men inte på dess betydelse (sanningsvärde). Emellertid är Freges system inte en växelsemantik, av det skälet att Frege tycks ha tänkt sig (och i allmänhet

har tolkats som) att det finns en oändlig hierarki av indirekta innebörder, vilket strider mot ändlighetskravet för S .

Två generella observationer talar för generell kompositionalitet gentemot standard-kompositionalitet. Det första är att skäl som talar för standard-kompositionalitet (t.ex. möjligheten att förstå yttranden av nya satser eller att tolkningskomplexiteten är låg) också talar för generell kompositionalitet. Även i det generella fallet finns ett ändligt antal semantiska grundfakta att lära sig vilka tillsammans är tillräckliga för att (effektivt) räkna fram meningen hos nya uttryck. Antalet semantiska grundfakta att lära sig är fler, men antalet steg i en tolkningsprocess är detsamma.

Den andra observationen är att flera språkliga konstruktioner som inte har en standard-kompositionell semantik har generellt kompositionell semantik. Ett enkelt exempel är anföringskonstruktionen (citering av språkliga uttryck; se Pagin och Westerståhl 2010c). Det kan tilläggas att alla tillämpningar som hittills har gjorts (se avsnitt 5) i en viss mening har varit av enklast tänkbara slag. Om vi fixerar argument 2 och 3 till Ψ kommer vi att få funktioner av typ $\Psi_{\sigma,i}$ från S till S . I alla semantiska tillämpningar som hittills gjorts, för olika konstruktioner i naturliga språk, inklusive den i nästa avsnitt, har dessa funktioner varit antingen identitetsfunktionen ($\Psi_{\sigma,i}(\mu) = \mu$), eller en konstant funktion, t.ex. $\Psi_{\kappa,1}(\mu) = V$ när det gäller anföringsfunktionen (κ är anföringsoperatoren, som växlar till att ge uttrycket självt som semantiskt värde).

Växelsemantik är den systematiska behandlingen och generaliseringen av Dummetts och Lewis distinktion mellan "assertoriskt innehåll" och "ingrediensmening" (respektive "innehåll" och "semantiskt värde").

3. VÄXELSEMANTIK FÖR TEMPORALA OPERATORER

Vi ska nu tillämpa det växelsemantiska ramverket för att ge en generellt kompositionell semantik för temporala operatorer. Semantiken kommer att använda sig av klassiska propositioner som semantiska utgångsvärden (värden för den designerade funktionen).

Vi utgår från ett Kaplan-ramverk. Det innebär en möjlig-världs-semantik i två nivåer: den övre nivån är Kaplans "character" och den undre nivån är Kaplans "content". Vi översätter dessa termer med "mening" respektive "innehåll". Mening är då en funktion från yttrande-kontexter ("contexts of use") till innehåll. För Kaplan är *content* en funktion från

ett $\text{par}(w, t)$ av en möjlig värld och en tidpunkt till ett sanningsvärde: en sats s är sann i en viss kontext c omm dess innehåll i c är sant i världen för c och tidpunkten för c (Kaplan 1989, s. 522). Skälet till att tala om paret (w, t) oberoende av c är att man ibland vill ta hänsyn till andra världar (p.g.a. modala operatorer) eller till andra tidpunkter (p.g.a. temporal operatorer).

Att utvärdera ett innehåll vid ett par av en värld och en tidpunkt är precis att behandla detta innehåll som temporalt. Kaplans semantik är en semantik för temporal propositioner. Vi ska nu avvika från Kaplan och försöka ge en semantik som i stället utgår från klassiska (eviga) propositioner och därför utvärderar innehåll vid möjliga världar, oberoende av tidpunkt. Vi kommer att se var problemen dyker upp och vad vi kan göra för att undvika dem.

Vi använder ' μ ' som meningsfunktion motsvarande Kaplans *character*, och ' μ^c ' som innehållsfunktion vid yttrandekontexten c . Ett naturligt förslag för semantik för operatorn 'det är alltid fallet att', här förkortad 'A' är då följande:

$$(9) \quad \mu_w^c(Ap) = 1 \text{ omm det gäller för varje tidpunkt } t \text{ att } \mu_w^c(p) = 1.$$

Detta ger ett förslag till mening för operatorn 'A'. Men förslaget är helt klart felaktigt när det gäller att representera den verkliga betydelsen hos 'det är alltid fallet att'. För om ' p ' representerar en evig proposition, så är den sann vid varje tidpunkt eller falsk vid varje tidpunkt. Att lägga till 'alltid' ändrar ingenting: det blir en tom kvantifikation över tidpunkter, precis som Kaplan påpekade i sitt operator-argument.

För att komma runt detta prövar vi att skifta tidsparametern i kontexten, enligt ett känt mönster: 'Det är alltid sant att p ' är sann vid en tidpunkt t omm p är sann vid alla tidpunkter t' . Här använder vi en substitutionsnotation: ' $c[t/\tau]$ ' betyder den kontext c' som skiljer sig från kontexten c högst däri att tidpunkten för c' är t , medan tidpunkten för c är τ .

$$(10) \quad \mu_w^c(Ap) = 1 \text{ omm det gäller för varje tidpunkt } t \text{ att } \mu_w^{c[t/\tau]}(p) = 1.$$

Denna semantik är kompositionell, för vi kan definiera en funktion r_s från meningar till meningar sådan att

$$(11) \quad r_s(m)_w^c = 1 \text{ omm det gäller för varje tidpunkt } t \text{ att } m_w^{c[t/\tau]} = 1.$$

Nu kan vi sätta $\mu(A) = r_s$, vilket tillsammans med (10) och (11) ger

$$(12) \quad (Ap) = \mu(A)(\mu(p)) = 1 \text{ omm det gäller för varje tidpunkt } t \text{ att } m_w^{c[t/\tau]}(p) = 1.$$

I det här fallet består den funktion som tillämpas på $\mu(A)$ och $\mu(p)$ helt enkelt i funktionsapplikation: $\text{APPLY}(\mu(A), \mu(p)) = \mu(A)(\mu(p))$.

Vi har alltså en kompositionell, icke-tom semantik för 'alltid'. Men problem kvarstår. Semantiken (10) för 'alltid' gör uttrycket till ett *monster* i Kaplans mening: ett uttryck som opererar på mening (*character*) (Kaplan 1989, s. 511). Som Kaplan understryker ger monster-operatorer felaktiga resultat för indexikala uttryck inom deras räckvidd. Vi kan illustrera detta med följande sats:

- (13) a. Det regnar alltid där jag är nu.
b. $A(\text{det regnar där jag är nu})$.

En partiell formalisering av (13a) ges i (13b). Vi tillämpar (10) på (13b). Resultatet blir som följer:

- (14) (i) $\mu_w^c(\text{det regnar alltid där jag är nu}) = 1$, omm
(ii) $\mu_w^c(A(\text{det regnar där jag är nu})) = 1$, omm
(iii) det gäller för varje tidpunkt t att $\mu_w^{c[t]}(\text{det regnar där jag är nu}) = 1$, omm
(iv) det gäller för varje tidpunkt t att det regnar vid t där jag är vid t .

(Vi hoppar över många detaljer i steget från (iii) till (iv), och tillåter oss att använda 'jag' i meta-språket.)

Resultatet i (14iv) är fel. Det (13a) betyder är att det vid varje tidpunkt t regnar där talaren befinner sig när satsen yttras, dvs. i kontextens tidpunkt, inte där talaren befinner sig vid t . Detta är precis den egenhet som monster i Kaplans mening har i relation till indexikala uttryck. Uttryckens semantiska värde skiftas inte under temporala operatorer, och en semantik som behandlar dessa som monster ger precis det felaktiga resultatet att låta värdena skifta.

Problemet är fatalt för försöket att både bevara kompositionalitet och den klassiska propositionen som utgångsmening. För att få rätt tolkning av (13a) måste vi antingen låta 'nu' bidra på ett annat semantiskt sätt i

- (15) Det regnar där jag är nu.

än det gör i (13a), eller ändra uppfattning om hur 'nu' bidrar i (15). Kaplan själv valde det senare alternativet, vilket innebar att övergå till temporala propositioner. Genom att låta 'nu' göra semantisk skillnad också när den inte är inbäddad under en temporal operator, som 'alltid', innebär att det 'nu' kombineras med, dvs.

(16) Det regnar där jag är.

uttrycker en temporal proposition. (15) och (16) har olika innehåll, även när de används direkt för att göra hävdanden. Detta var ett resultat vi ville undkomma.

Det vi ska göra för att undslippa Kaplans resultat är att ge upp kompositionalitet, dvs. standard-kompositionalitet, till förmån för generell kompositionalitet. Vi ska tillämpa växelsemantiik och ersätta semantiken (10) med en växelsemantisk behandling. Den kommer att delvis vara analog med vår behandling av namn i modala kontexter i Glüer och Pagin 2006.⁵

Det första steget är att införa en distinktion mellan verklig yttrande-kontext och *skiftad* kontext, analogt med distinktionen mellan den aktuella världen och icke-aktuella möjliga världar. Vi kallar den verkliga yttrandekontexten ' c_a '. Strängt taget är detta inte adekvat, eftersom det finns oändligt många verkliga yttrandekontexter. En stringent behandling skulle kräva användningen av par $\langle c, c' \rangle$ av kontexter, analogt med två-dimensionell semantik för modala uttryck, men vi förenklar notationen här och låtsas som att vi kan identifiera en verklig yttrandekontext på detta sätt.

Det andra steget är att låta kontext-skiftande operatorer också vara semantiska växlare. Närmare bestämt, om σ_t är en temporal operator, där argumentet är en sats, så är $(\sigma_t, 1)$ en växlare.

Det tredje steget är att en sådan växlare byter till en *aktualistisk* semantisk funktion. Denna funktion, μ_a , är sådan att dess värde för ett indexikalt uttryck $\#$ vid en godtycklig kontext c är detsamma som värdet för den designerade funktionen μ_0 vid yttrande-kontexten c_a . Till exempel kommer $\mu(\text{'nu'})$ vid en godtycklig skiftad kontext c att vara $\mu(\text{'nu'})$, dvs. tidpunkten för yttrandet ifråga.

Med dessa ingredienser skapar vi ett växelsemantiskt system $S = \{\mu_0, \mu_a\}$, μ_0, Ψ , där växelfunktionen Ψ är definierad bl.a. av

$$(17) \quad \Psi(\mu_0, \sigma_a, 1) = \mu_a.$$

Här är σ_a den operator i termalgebra som genererar 'det är alltid fallet att' (förkortat 'A'). Argument till σ_a tolkas alltså med μ_a .

5. En annan växelsemantisk behandling än den vi presenterar här har givits av Sara Packalén (2016). Packaléns idé är att växla till temporal propositioner som semantiska värden inom räckvidden för temporal operatorer. Ett problem med denna metod är att särskilda undantag måste göras för satstyper där temporal uttryck meningsfullt förekommer i sådana språkliga kontexter, som t.ex. 'nu' i (13a).

Vi ersätter nu (10) med den nya semantiken. Vi ger här den del där det är μ_0 som appliceras på en σ_a -term:

$$(18) \quad \mu_{0,w}^c(Ap) = 1 \text{ omm det gäller för varje tidpunkt } t \text{ att } \mu_{a,w}^{c[t/\tau]}(p) = 1.$$

Detta är alltså den växelsemantiska klausulen för icke temporalt inbäddade "alltid"-satser. Att den ger rätt resultat kan lätt inses:

- (19) (i) $\mu_{0,w}^c(\text{det regnar alltid där jag är nu}) = 1$, omm
 (ii) $\mu_{0,w}^c(A(\text{det regnar där jag är nu})) = 1$, omm
 (iii) det gäller för varje tidpunkt t att $\mu_{a,w}^{c[t/\tau]}(\text{det regnar där jag är nu}) = 1$, omm
 (iv) det gäller för varje tidpunkt t att det regnar vid t där jag är vid τ .

(Vi hoppar liksom tidigare över många detaljer i steget från (iii) till (iv).)

(18) ger den korrekta tolkningen. Semantiken är generellt kompositionell, men inte standard-kompositionell. 'Alltid' behandlas som ett monster i Kaplans mening, men växelsemantiken undviker de problem som detta annars ger upphov till. μ_0 ger klassiska propositioner som semantiska värden till hela satser, vilket var ett av målen. Vi kan lägga till att även μ_a ger klassiska propositioner som semantiska värden, dock i många fall, som i (19), andra än dem som ges av μ_0 .

3. ANDRA TILLÄMPNINGAR

Den första tillämpningen av växelsemantik (Glüer och Pagin 2006, 2008, 2012b) gällde egennamn i modala kontexter, och syftade till att stödja de intuitioner om sanningsvärden för modala satser som Saul Kripke (1980) åberopat för tesen att namn är *rigida designatorer*. Semantiken lät namn ha icke-rigid tolkning i icke-modala kontexter men bidra med sin referens i den aktuella världen i modala kontexter, vilket ger samma sanningsvärden (i den aktuella världen) som Kripkes semantik. Sedan dess har semantiken också utvecklats för att hantera *tomma namn* och negativa existensutsagor (Glüer och Pagin 2014).

Vi har också tillämpat metoden för *generella termer* i modala kontexter (Glüer och Pagin 2012a), med syftet att semantiskt karakterisera termer för naturliga sorter (t.ex. 'guld', 'vatten'). Här är idén att en semantisk funktion tar ytliga egenskaper som värden (det som Hilary Putnam (1975b) kallar "stereotyper"), medan den andra semantiska funktionen tar underliggande egenskaper (det som Putnam kallar "essenser") som

värden. Liksom i fallet med egennamn är modala operatorer semantiska växlare.

Dag Westerståhl och Peter Pagin (2010c) har tillämpat metoden på *anföring* (citat), där en anföringsoperator växlar från den designerade funktionen till V , strängfunktionen.

I en opublicerad text har Pagin också tillämpat metoden för en semantik för övertygelsetillskrivningar ('Sture tror att jorden krymper'). Idén här är att klassiska propositioner utgör innehållet i övertygelser, men att dessa presenteras för subjektet på ett strukturerat sätt. Övertygelseoperatorn är då i sin andra argumentplats en växlare som från den designerade funktionen byter till en funktion som ger strukturerade meningar som värden. Med hjälp av växelsemantiken kan möjlig-världssemantik av klassiskt Hintikka-format (Hintikka 1962) kombineras med strukturerade meningar och ge intuitivt korrekta substitutionsvillkor.

Till sist kan vi framhålla att den växelsemantiska tekniken i flera fall kan visa att något som ofta uppfattas som ett *alternativ* till klassiska propositioner bättre kan ses som *komplement*. Detta kan enkelt illustreras med kvantifikatorer. Standardsemantiken för första ordningens språk, med kvantifikatorerna 'alla' och 'något', eller deras formella motsvarigheter ' $\forall$, $\exists$ ' följer Tarski (1935) i att tolka såväl öppna som slutna formler med hjälp av funktioner för tilldelning av värden till variabler (eller sekvenser av tilldelade objekt). Satser, dvs. formler utan fria variabler, blir helt enkelt ett specialfall. En sats är sann omm den satisfieras av alla tilldelningar. Sanning självt blir ett specialfall av satisfiering.

Ett problem med detta är att vi inte ser någon semantisk skillnad mellan en sann sats, som ' $1 + 1 = 2$ ' och en giltig formel med fria variabler, som ' $x = x$ '. Problemet förvärras i en intensionell utvidgning, då meningen hos en sats inte blir en mängd möjliga världar utan en mängd av *par* $\langle w, f \rangle$ av en värld och en tilldelning.

Dessa problem kan undvikas genom att behandla kvantifikatorer (och andra variabelbindande operatorer) som växlare: de växlar från sanning till satisfiering. I det extensionella fallet ser det ut så här:

$$(20) \quad \mu_0(\forall x Fx) = 1 \text{ omm det gäller för alla tilldelningar } f \text{ att } \mu_1(f, Fx) = 1.$$

Här är μ_1 (den karakteristiska funktionen för) satisfieringsrelationen, medan μ_0 tilldelar sanningsvärden, inte mängder av tilldelningsfunktioner. På det viset får vi en tydlig skillnad mellan semantiken för satser och semantiken för formler med fria variabler och behåller en begreppslig skillnad mellan sanning och satisfiering.

Vi har ett analogt resultat för *dynamisk semantik*, där dynamiska semantiska värden, som *kontextförändringspotential*, ofta setts som ersättare till sanningsvillkor. De dynamiska värdena behövs, men de behövs just för att hantera dynamiska konstruktioner, alltså för den inre semantiken. Återigen visar en växelsemantisk behandling att de dynamiska värdena kompletterar, inte konkurrerar med, klassiska propositioner som meningar för icke inbäddade satser. Utrymmet här tillåter inte en presentation av detaljerna. Men den allmänna strategin torde vara klar nog.

LITTERATUR

- Brogaard, Brit. 2012. *Transient Truths: An Essay in the Metaphysics of Propositions*. Oxford: Oxford University Press. Elektronisk resurs på SUB.
- Chomsky, Noam och Howard Lasnik. 1993. "Principles and Parameters Theory". I *Syntax: An International Handbook of Contemporary Research*, utg. A. Jacobs m. fl. Berlin: de Gruyter.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- . 1991. *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.
- Frege, Gottlob. 1948. "Sense and Reference". *Philosophical Review* 57, s. 209–30. Övers. Max Black. Även i Frege 1960.
- . 1960. *Translations from the Philosophical Writings of Gottlob Frege*, utg. Peter Geach och Max Black. Oxford: Basil Blackwell.
- Glüer, Kathrin och Peter Pagin. 2006. "Proper names and relational modality". *Linguistics & Philosophy* 29, s. 507–35.
- . 2008. "Relational Modality". *Journal of Logic, Language and Information* 17, s. 307–22.
- . 2012a. "General Terms and Relational Modality". *Nous* 46, s. 159–99.
- . 2012b. "Reply to Forbes". *Analysis* 72, s. 298–303.
- . 2014. "Vulcan Might Have Existed, and Neptune Not. On the Semantics of Empty Names". I *Empty Representation: Reference and Non-Existence*, utg. Genoveva Martí och Manuel García-Carpintero, s. 117–41. Oxford: Oxford University Press.
- Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca, NY: Cornell University Press.
- Hodges, Wilfrid. 2001. "Formal Features of Compositionality". *Journal of Logic, Language and Information* 10, s. 7–28.
- Hornstein, Norbert, Jairo Nunes och Kleantes K. Grohmann. 2005. *Understanding Minimalism: An Introduction to Minimalist Syntax*. Cambridge: Cambridge University Press.
- Kaplan, David. 1989. "Demonstratives". I *Themes from Kaplan*, utg. J. Almog, J. Perry och H. Wettstein. Oxford: Oxford University Press, s. 481–563. Elektronisk resurs på SUB.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

- Lewis, David. 1980. "Index, Context, and Content". I *Philosophy and Grammar*, utg. Stig Kanger och Sven Öhman, s. 79–100. Dordrecht: Reidel.
- . 1986. *On The Plurality of Worlds*. Oxford: Blackwell.
- Packalén, Sara. 2016. *Content and Composition: An Essay on Tense, Content, and Semantic Value*. Diss. Stockholm: Department of Philosophy, Stockholm University.
- Pagin, Peter. 2003. "Communication and strong compositionality". *Journal of Philosophical Logic* 32, s. 287–322.
- Pagin, Peter och Dag Westerståhl. 2010a. "Compositionality I: Definitions and Variants". *Philosophy Compass* 5, s. 250–64.
- . 2010b. "Compositionality II: Arguments and Problems". *Philosophy Compass* 5, s. 265–82.
- . 2010c. "Pure Quotation and General Compositionality". *Linguistics and Philosophy* 33, s. 381–415.
- Putnam, Hilary. 1975a. *Mind, Language and Reality: Philosophical Papers vol. 2*. Cambridge: Cambridge University Press.
- . 1975b. "The Meaning of 'Meaning'". I *Language, Mind and Knowledge*, utg. Keith Gunderson. Minnesota Studies in the Philosophy of Science VII. Minneapolis: University of Minnesota Press. Omtryckt i Putnam 1975a.
- Recanati, François. 2007. *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
- Richard, Mark. 1981. "Temporalism and eternalism". *Philosophical Studies* 39, s. 1–13.
- Tarski, Alfred. 1935. *Der Wahrheitsbegriff in den formalisierten Sprachen*. Leopoli: Studia Philosophica.

LOGIK OCH NORMATIVITET

1. INTRODUKTION

Logiken har ofta beskrivits som en normativ disciplin som specificerar regler för tänkande eller korrekt argumentation (Priest 1989, s. 2). Exempelvis karakteriserar Kant logik som bestående av "de absolut nödvändiga reglerna för tänkandet"¹ (A52/B76), vilka inte talar om för oss "hur förståelsen är och tänker" utan "hur den *bör* fortskrida" (1800/1974, s. 16, citerad i Steinberger 2017). Frege säger att "logik är en normativ vetenskap" likt etiken, eftersom dess mål är att föreskriva "regler för vårt tänkande och för att hålla något för sant" (1897/1997, s. 228). Till stöd för detta påstående hänvisar både Kant och Frege till det faktum att logiken uppenbarligen inte beskriver hur människor faktiskt resonerar, vilket är en uppgift för empirisk psykologi. Vi vänder oss till logiken, inte för att upptäcka hur vi *faktiskt* resonerar, utan hur vi *bör* resonera eller hur vi *får* resonera.

Samtidigt finns det ett problem. Vi kan tänka på en *logik* som en *teori om implikation*, det vill säga en specifikation av konsekvensrelationen på en mängd sanningsbärare. Som sådan är teorins påståenden om logisk konsekvens eller giltighet inte explicit normativa; de innehåller inga förekomster av termerna "bör", "får" eller "skäl". Således verkar logikens lagar vara deskriptiva; de verkar *beskriva* vad som följer från vad. En *normativ teori om slutledning* och *resonemang* är däremot en teori om hur man bör resonera, eller hur man får resonera, där resonemang förstås som ett välgrundat byte av åsikter (Harman 1968). Frege reflekterar över detta problem i sin diskussion av frågan huruvida logikens lagar är mer som moralens lagar eller som fysikens lagar:

Ordet "lag" används på två sätt. När vi talar om moraliska eller civila lagar, så avser vi föreskrifter, vilka bör efterlevas men med vilka faktiska företeelser inte alltid överensstämmer. Naturlagar är generella kännetecken hos det som händer i naturen, och företeelser i naturen överens-

1. Samtliga översättningar av citat från engelska till svenska är gjorda av Sara Packalén.

stämmer alltid med dem. Det är snarast i denna bemärkelse som jag talar om sanningens lagar. Här handlar det naturligtvis inte om det som händer utan det som är. Från sanningens lagar följer föreskrifter om hävdande, tänkande, omdömen, slutledningar. (Frege 1918/1997:325)

Det som Frege föreslår här har slagit an tonen för mycket av den samtida diskussionen om logikens normativitet: idén att logiken – bestående av ”sanningens lagar” som Frege uttrycker det – implicerar normativa påståenden om hävdande, tänkande, omdömen, slutledningar.² Alltsedan Frege har diskussionen om logikens normativitet tenderat att fokusera på bryggprinciper på formen (jfr Harman 1986; MacFarlane 2004; Steinberger 2016):

$$(P_1, \dots, P_n \models C) \rightarrow \Phi(P_1, \dots, P_n, C)$$

där Φ är en normativ operator (som *bör*, *får* eller *det finns skäl att*); $\rightarrow$ är materiell implikation; och $\Phi(P_1, \dots, P_n, C)$ är ett normativt påstående av något slag som rör trosföreställningar, som att om man tror på premisserna, så bör man tro på slutsatsen:

$$(1) \quad (P_1, \dots, P_n \models C) \rightarrow \text{Om } S \text{ tror var och en av } P_1, \dots, P_n, \text{ så bör } S \text{ tro } C.$$

Mycket av diskussionen om huruvida logiken är normativ utgår från antagandet att logiken är normativ endast utifall någon liknande bryggprincip är *sann*. Jag kommer dock här argumentera för att antagandet är förhastat. Att en sådan bryggprincip är sann är inte i sig tillräckligt för att visa att logiken är normativ. Som vi kommer att se krävs betydligt mer än att en bryggprincip som (1) är sann för att logiken ska vara normativ.

2. MATERIELL IMPLIKATION

Som vi har sett, förstås (1) vanligtvis som en materiell implikation. Likväl, även om (1) eller motsvarande är sann, så följer det inte att logiken är normativ. Materiell implikation är inte en tillräckligt stark länk mellan påståenden om giltighet och normativa påståenden för att logik därmed ska vara normativ. För att se varför, beakta den generella princip som antas här, nämligen:

2. Notera att Frege föreslår att logik är normerande för flera saker – hävdande, tänkande, omdömen, slutledningar. Jag kommer här att fokusera på idén att logik är normerande för slutledningar.

Materiell implikation: Om $A \rightarrow \Phi(B)$, så är A normativ.³

Materiell implikation säger att varje materiell implikation med en normativ eftersats har en normativ försats. Denna princip antas ofta i diskussioner om normativitet hos mening och innehåll.⁴ Till exempel karaktäriserar Boghossian (1989) tesen att mening är normativt på följande sätt:

Anta att uttrycket "grön" betyder *grön*. Det följer omedelbart att "grön" tillämpas *korrekt* enbart på *de här* tingen (de gröna) och inte på *de där* (de icke-gröna). Faktumet att uttrycket betyder något implicerar alltså en hel mängd *normativa* sanningar om mitt beteende med detta uttryck ... (Boghossian 1989, s. 513)

Boghossian föreslår här att om det faktum att jag menar något med ett uttryck implicerar något normativt, så är mening normativt.

Det är dock lätt att se att *Materiell implikation* är falsk. Överväg till exempel följande hedonistiskt utilitaristiska princip:

(2) Handling A maximerar njutning $\rightarrow S$ bör utföra A .

Det är uppenbart att (2) är en materiell implikation med en normativ eftersats. Trots det följer det inte från att eftersatsen är normativ att försatsen är normativ: Även om (2) är sann så följer det inte att njutningsmaximering är en normativ egenskap, och inte heller att det faktum att A maximerar njutning är ett normativt faktum. Just faktumet att A maximerar njutning är i allmänhet ansett som ett paradigmatiskt exempel på ett icke-normativt, naturligt faktum, det vill säga den typen av fakta som studeras i natur- och samhällsvetenskaperna. *Materiell implikation* är således falsk. Dessutom, även om (1) är sann, så följer det inte att logisk giltighet är en normativ egenskap eller att fakta om logisk giltighet är normativa fakta. I allmänhet följer det inte att försatsen i en materiell implikation är normativ från att eftersatsen är normativ.

3. Jag kommer i regel att anta att A och B är propositioner. På vissa ställen kommer jag växla mellan att tala om propositioner och att tala om fakta, i vilket fall jag menar de faktum som A och B motsvarar.

4. I enlighet med Boghossian och andra förstod jag påståendet att mening är normativt framför allt på detta sätt när jag argumenterade mot det (se Hattiangadi 2006, 2007).

3. METAFYSISKT NÖDVÄNDIG FÖLJD

Vi skulle kunna försöka förstärka påståendet att logiken är normativ till ett påstående om metafysisk nödvändighet:

- (3) Det är metafysiskt nödvändigt att: $(P_1, \dots, P_n \models C) \rightarrow$ Om S tror var och en av $P_1, \dots, P_n$, då bör S tro C .

Återigen ser vi dock begränsningarna med förslaget om vi betraktar den generella principen som antas (där $\Box$ är operatoren för metafysisk nödvändighet):

Metafysiskt nödvändig följd: Om $\Box(A \rightarrow \Phi(B))$, då är A normativ.

Metafysiskt nödvändig följd är den generella versionen av en idé som har florerat i många av de kontexter där normativitet har diskuterats. Exempelvis karaktäriserar Glüer och Wikforss en version av idén att innehåll är normativt som tesen att innehåll metafysiskt nödvändiggör normer; detta är synen som de kallar "innehållsskapad normativitet". (Glüer & Wikforss 2009)⁵ På ett liknande sätt karaktäriserar Broome (2013) synen att rationalitet är normativt som synen att det är metafysiskt nödvändigt att om rationalitet kräver att du gör Φ , så bör du eller har du pro tanto skäl att Φ . Även om Glüer, Wikforss och Broome kritiserar de versioner av normativitetstesen som karaktäriserar normativitet på detta sätt skriver många under på att något – må det vara mening, innehåll, tro, eller rationalitet – är normativt på grund av att det uppfyller *Metafysiskt nödvändig följd* (se Blackburn 1984; Kripke 1982; Millar 2004, Miller 1998; Verheggen 2011; Whiting 2007, 2009).

Vi kan dock se att *Metafysiskt nödvändig följd* är falsk, återigen genom att vända vår uppmärksamhet mot det moraliska fallet. För de första

5. Glüer and Wikforss gör en distinktion mellan meningssskapande normativitet och meningsbestämmande normativitet. I det senare fallet är i själva verket förhållandet mellan vad som är det mest metafysiskt grundläggande omvänt. Idén här är att $\Box(\Phi(B) \rightarrow A)$, så är A normativ. Exempelvis omfamnar Brandom (1994) en sådan syn: Normativa sanningar om slutledningar i en viss gemenskap bestämmer vilka som är giltiga i den gemenskapen. För en invändning mot Brandoms teori om meningsbestämmande normativitet, se Hattiangadi 2003. I det specifika fallet med logik, talar invändningen mot *Identitet* mot en sådan syn på logikens normativitet. Det vill säga, $\Box((S \text{ tror var och en av } P_1, \dots, P_n \rightarrow S \text{ bör tro } C) \rightarrow (P_1, \dots, P_n \models C))$ förefaller falsk eftersom det mycket väl kan vara sant att $(S \text{ tror var och en av } P_1, \dots, P_n \rightarrow S \text{ bör tro } C)$ av en massa olika skäl: för att $P_1, \dots, P_n$ ger induktivt stöd för C , för att personen kommer utsättas för sanktioner från sin gemenskap om hon inte sluter sig till C , och så vidare.

anser de flesta metaetiker att normativa moraliska fakta supervenierar på naturliga fakta.⁶ Det vill säga, om vi låter *D* vara en icke-normativ naturalistisk beskrivning av världen, så accepterar de flesta metaetiker att:

Minimal superveniens: $\Box(D \rightarrow \Phi(B))$.⁷

Det är uppenbart att även om *Minimal superveniens* är sann så följer det inte att *D* är normativ, tvärt emot vår initiala stipulation. Faktum är att om sanningen hos *Minimal superveniens* vore tillräcklig för att visa att *D* är normativ, så skulle det inte kunna vara sant att det normativa supervenierar på det icke-normativa! Om *Metafysisk nödvändig följd* vore sann så skulle de sanningar som ingår i superveniensbasen för en normativ sanning själva vara normativa, oavsett vilka de var. Denna poäng kan generaliseras till logiken. Om logiska sanningar är normativa, och vad det innebär att vara normativ fångas av *Metafysiskt nödvändig följd*, så följer det att de normativa sanningarna inte supervenierar på de icke-normativa naturliga sanningarna.

För att ta ett mer konkret exempel så anser många att fundamentala normativa principer är metafysiskt nödvändiga. Om det stämmer, och hedonistisk utilitarism är en fundamental normativ sanning, då är även (4) sann:

- (4) Det är metafysiskt nödvändigt att: Handling A maximerar njutning $\rightarrow S$ bör utföra A.

(4) är återigen en instans av *Metafysiskt nödvändig följd*. Men eftersom faktumet att en viss handling maximerar njutning är ett paradigmiskt exempel på ett naturligt faktum, så följer det inte, även om (4) är sann, att njutningsmaximering är en normativ egenskap eller att faktumet att A maximerar njutning är ett normativt faktum. Alltså är *Metafysiskt nödvändig följd* falsk.

4. IDENTITET

För att logiken ska vara normativ krävs en ännu starkare koppling mel-

6. Se McPherson 2015 för en översikt. Notera att jag inte delar denna syn. Se Hattiangadi, kommande.

7. Strikt talat så är det *Stark global superveniens* som i regel omfattas av metaetiker: För alla metafysiskt möjliga världar, *w* och *w'*, om *w* och *w'*, är desamma i alla icke-normativa avseenden, så är de desamma i alla normativa avseenden. *Minimal superveniens* följer dock av *Stark global superveniens*, så i den utsträckning som metaetiker accepterar *Stark global superveniens* måste de även acceptera *Minimal superveniens*.

lan det logiska och det normativa än metafysiskt nödvändig implikation. Kan identitet göra jobbet? Följande generella princip verkar *prima facie* lockande:

Identitet: Om $A = \Phi(B)$, så är A normativ.

Men till och med *Identitet* framstår som lite förbryllande när vi tittar på det moraliska fallet. Överväg, till exempel, (5):

- (5) Faktumet att handling A maximerar njutning = faktumet att S bör utföra A .

Det som är förbryllande med *Identitet* är att (5) vanligtvis antas implicera att faktumet att S bör utföra A är ett *naturligt* faktum, inte att faktumet att handling A maximerar njutning är ett *normativt* faktum. Det här beror på att naturliga fakta vanligtvis ses som fundamentala; alltså, att slå fast att något icke-fundamentalt faktum – så som ett moraliskt faktum eller ett mentalt faktum – är identiskt med ett naturligt faktum hjälper oss att reducera det icke-naturliga faktumet till ett naturligt faktum. Tvärtom verkar *Identitet* anta – vilket är kontroversiellt som bäst – att normativa fakta är mer fundamentala än naturliga fakta; alltså, att slå fast att ett visst naturligt faktum är identiskt med ett normativt hjälper oss att reducera det naturliga till det normativa.

En annan möjlighet är naturligtvis *Inget företräde*-tesen.⁸

Inget företräde: Om $A = \Phi(B)$, så är A och $\Phi(B)$ både naturliga och normativa.

Inget företräde undviker vissa av konstigheterna med *Identitet* eftersom den inte behandlar det normativa som *mer* fundamentalt än det naturliga. Likväl behandlar den det normativa och det naturliga som lika fundamentala i någon bemärkelse, vilket många kommer anse som osannolikt. Men inte heller *Inget företräde* tar oss särskilt långt när det kommer till logikens normativitet, eftersom själva idén att fakta om logisk giltighet är identiska med normativa fakta är osannolikt. Skälet är att om faktumet att $(P_1, \dots, P_n \models C)$ är identiskt med det normativa faktumet att $\Phi(P_1, \dots, P_n, C)$, så måste även den omvända implikationen, från $\Phi(P_1, \dots, P_n, C)$ till $(P_1, \dots, P_n \models C)$, hålla med metafysisk nödvändighet. Om vi håller fast vid instansen av $\Phi(P_1, \dots, P_n, C)$ som föreslås ovan, och fakta om giltighet är identiska med normativa fakta, så:

8. Wedgwood (2007) omfamnar en inget-företräde-tes om ett ömsesidigt beroende mellan det normativa och det mentala.

$$(6) \quad \Box((S \text{ tror var och en av } P_1, \dots, P_n \rightarrow S \text{ bör tro } C) \rightarrow (P_1, \dots, P_n \models C))$$

Likväl är (6) osannolik, eftersom $P_1, \dots, P_n$, skulle kunna utgöra icke-deuktiva skäl för varför S bör tro C . Exempelvis kan $P_1, \dots, P_n$ ge starkt induktivt stöd för C , så att om det är sant att S tror $P_1, \dots, P_n$, så bör S tror C , utan $P_1, \dots, P_n$ logiskt implicerar C .

5. ESSENTIELL KOPPLING

Ett annat sätt att förstå idén att logiken är normativ är att stava ut den som idén att egenskapen logisk giltighet har en normativ essens. Enligt *Modal essentialism* är det just så att metafysiska nödvändigheter måste grundas i essenser:⁹

Modal essentialism: Om det är metafysiskt nödvändigt att p , så finns det en samling entiteter X sådan att det är en del av essensen av X att p .

Man skulle då kunna argumentera för att (3), som är en metafysisk nödvändighet, är en del av essensen hos egenskapen giltighet. Detta antyder följande generella princip, där F denoterar en egenskap av något slag:

Essentiell koppling: Om det är en del av F 's essens att $\Box(A \rightarrow \Phi(B))$, så är F normativ.¹⁰

Essentiell koppling leder till ett liknande problem som *Metafysiskt nödvändig följd*. Som vi har sett accepterar många metaetiker *Minimal superveniens*. Tillsammans med *Modal essentialism* följer det från den principen att $\Box(D \rightarrow \Phi(B))$ måste vara en del av essensen hos något, låt oss anta egenskapen F . Om *Essentiell koppling* är sann, så följer det att F är en normativ egenskap. Det strider dock mot en välbekant form av etisk naturalism, enligt vilken faktumet att $\Box(D \rightarrow \Phi(B))$ är en del av essensen hos F implicerar att F är en *naturlig* egenskap. Exempelvis skulle en naturalist kunna argumentera för att det är en del av essensen hos egenskapen att vara god att det är nödvändigt att om något är njutbart, så är det gott. En sådan naturalist skulle mer generellt kunna anta att det finns en lång – kanske oändlig – disjunktion av naturliga egenskaper (N_i), sådan att:

$$(9) \quad \Box \forall x (Fx \leftrightarrow (N_1(x) \vee N_2(x) \vee N_3(x) \vee \dots)).$$

9. Rosen, kommande.

10. Se Wedgwood 2007.

Dessutom skulle hon kunna mena att (9) är en del av essensen hos F , där F är någon egenskap, så som egenskapen godhet eller riktighet. En sådan etisk naturalist skulle anse att detta visar, inte att F är *normativ*, utan snarare att F är *naturlig*. Återigen är det underliggande antagandet utan tvekan att naturliga egenskaper är mer fundamentala än normativa egenskaper, så att slå fast att det finns en metafysiskt nödvändig koppling mellan naturliga och normativa egenskaper, som är en del av essensen hos normativa egenskaper, är att reducera normativa egenskaper till det naturliga, inte vice versa. Alltså är *Essentiell koppling* som bäst kontroversiell, och som sämst falsk.

6. BEGREPPSLIGT NÖDVÄNDIG FÖLJD

Hartry Field (2009; 2015) föreslår ytterligare en kandidat till att vara den starkare kopplingen mellan det logiska och det normativa: att *begreppet* logisk giltighet är normativt. Mer precist menar Field att begreppet logisk giltighet spelar en normativ roll, vilken tar sig uttryck som en uppsättning regler för trosföreställningar, i synnerhet regeln att man inte bör tro $P_1, \dots, P_n$, utan att tro C , där C följer logiskt från $P_1, \dots, P_n$. Notera att Field inte menar att vi kan ge en normativ *definition* av betydelsen hos "giltig". Som Field argumenterar, skulle det "solka logikens renhet att definiera giltighet i normativa termer vilkas exakta innehåll inte är helt klart" (Field 2015, §3). Tanken är snarare att begreppet logisk giltighet spelar en normativ roll i våra mentala liv genom att vägleda oss eller motivera oss i hur vi argumenterar.¹¹

Även om det här är på rätt spår mot att specificera ett intressant sätt på vilket logiken kan sägas vara normativ, så behöver vi fortfarande vara aktsamma med vad exakt som menas med påståendet att begreppet giltighet har en normativ roll. Om vad det innebär för ett begrepp att spela en normativ roll är att det *faktiskt* reglerar trosföreställningar eller handlingar, så kommer alldeles för många klart icke-normativa begrepp till slut klassas som normativa. Beakta återigen den hedonistiska utilitaristen som anammar regeln att enbart agera på ett sätt som maximerar njutning. I det fallet spelar begreppet njutningsmaximering en normativ roll för den hedonistiska utilitaristen i Fields mening, även om begreppet njutningsmaximering inte i sig är ett normativt begrepp. Trots allt skulle en deontisk etiker till fullo kunna förstå begreppet njut-

11. Gibbard (2012) omfamnar en liknande tes om meningsnormativitet.

ningsmaximering samtidigt som hon förkastar den hedonistiska utilitaristens regel. För att ett begrepp ska vara normativt krävs snarare att det är ett villkor för att vara kompetent med begreppet att det spelar en normativ roll för en.

Om begreppet logisk giltighet är normativt i den bemärkelsen, så är det ett villkor för att förstå begreppet logisk giltighet att om man bedömer att ett argument är giltigt, så fäller man ett normativt omdöme av något slag. Det vill säga:

- (10) Det är begreppsligt nödvändigt att: $(P_1, \dots, P_n \models C) \rightarrow S$ bedömer att $\Phi(P_1, \dots, P_n, C)$

Återigen kan vi konstruera en generell princip:

Begreppsligt nödvändig följd: Om det är begreppsligt nödvändigt att: $A \rightarrow \Phi(B)$, så är A normativ.

Vid första anblick verkar denna princip tåla granskning. De flesta håller nämligen med om att exempelvis (2) inte är begreppsligt nödvändig, och inte heller några liknande substantiella normativa sanningar, så som att du bör hålla dina löften eller att mord är fel. Dessutom kvalificerar tjocka normativa begrepp, så som begreppet mod eller begreppet rättvisa, som normativa just på grund av att vissa enkla påståendesatser, så som "handling A är modig" eller "handling A är rättvis" verkar ha normativa följder, så som att "man har skäl att utföra A". Alltså, *Begreppsligt nödvändig följd* utgör det bästa sättet att förstå idén att logik är normativ.

7. SLUTSATS

Ovanstående är en föregångare till frågan "är logik normativ?", vilken jag inte har utrymme att diskutera här.¹² Det som vi har slagit fast här är att det kommer krävas mer för att visa att logiken är normativ än att enbart visa att någon viss bryggprincip är sann. Denna poäng har relevans långt utanför fallet logik, så som exempelvis vår förståelse för påståendet att mening är normativt, att rationalitet är normativt, att epistemologi är normativt, eller att estetik är normativt.¹³

12. Det diskuteras i Hattiangadi, kommande b.

13. Jag är mycket tacksam för Sara Packaléns översättning av uppsatsen från engelska till svenska. En tidigare version av denna uppsats presenterades på European Normativity Network konferensen i Stockholm. Tack till Kathrin Glüer för hennes kommentarer och till publiken för en lärorik diskussion.

LITTERATUR

- Blackburn, S. 1984. "The individual strikes back". *Synthese* 58, s. 281–301.
- Boghossian, P. A. 1989. "The rule-following considerations". *Mind* 98(392), s. 507–49.
- Broome, J. 2013. *Rationality Through Reasoning*. Oxford: Wiley-Blackwell.
- Field, H. 2009. "What is the normative role of logic?", *Proceedings of the Aristotelian Society* 83, s. 252–68.
- . 2015. "What is logical validity?". I *Foundations of Logical Consequence*, red. R. Coln, R. Caret och Ole T. Hjortland. Oxford: Oxford University Press.
- Frege, G. 1897/1997. "Logic". I *The Frege Reader*, utg. M. Beaney. Oxford: Blackwell.
- . 1918/1997. "Thought". I *The Frege Reader*, utg. M. Beaney. Oxford: Blackwell.
- Harman, G. 1984. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Glüer, K. och Å. Wikforss. 2009. "Against content normativity". *Mind* 118, s. 31–70.
- Gibbard, A. 2012. *Meaning and Normativity*. Oxford: Oxford University Press.
- Hattiangadi, A. 2003. "Making it implicit: Brandom on rule following". *Philosophy and Phenomenological Research* LXVI, s. 419–31.
- . 2007. *Oughts and Thoughts: Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.
- . 2017. "The normativity of meaning". I *A Companion to the Philosophy of Language*, utg. B. Hale, C. Wright och A. Miller. Andra upplagan. Oxford: Wiley Blackwell.
- . Kommande a. "Moral Supervenience". I *Representation and Evaluation, Special Edition of the Canadian Journal of Philosophy*, utg. M. Bedke.
- . Kommande b. "Logical Disagreement". I *Meta-Epistemology*, utg. C. McHugh och D. Whiting. Oxford: Oxford University Press.
- Kant, I. 1781/1787/1988. *Kritik Der Reinen Vernunft*. Hamburg: Meiner.
- Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Cambridge MA: Harvard University Press.
- MacFarlane, J. 2004. "In what sense (if any) is logic normative for thought?", Unpublished.
- McPherson, T. 2015. "Supervenience in Ethics". *The Stanford Encyclopedia of Philosophy* (Winter 2015 edition). E.N. Zalta, ed. URL = <https://plato.stanford.edu/archives/win2015/supervenience-ethics/>.
- Millar, A. 2002. "The normativity of meaning". I *Logic, Thought and Language*, utg. A. O'Hear, s. 57–73. Cambridge: Cambridge University Press.
- Miller, A. 1998. *Philosophy of Language*. Montreal och Kingston: McGill-Queens University Press.
- Priest, G. 1989. "Two Dogmas of Quineanism". *Philosophical Quarterly* 117, s. 289–301.
- Ridge, M. 2007. "Anti-reductionism and supervenience". *Journal of Moral Philosophy* 4(3), s. 330–48.

- Rosen, G. 2010. "Metaphysical dependence: grounding and reduction". I *Modality: Metaphysics, Logic, and Epistemology*, utg. B. Hale och A. Hoffman, s. 109–39. Oxford: Oxford University Press.
- . Kommande. "Normative Necessity". I *Metaphysics, Meaning and Modality: Themes from Kit Fine*, utg. M. Dumitru. Oxford: Oxford University Press.
- Steinberger, F. 2016. "Explosion and the normativity of logic". *Mind* 125(498), s. 385–419.
- . 2017. "The Normative Status of Logic". *The Stanford Encyclopedia of Philosophy* (Spring 2017 edition). E. N. Zalta, utg. URL = <https://plato.stanford.edu/archives/spr2017/entries/logic-normative/>.
- Verheggen, C. 2011. "Semantic normativity and naturalism". *Logique & Analyse* 216, s. 553–67.
- Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.
- Whiting, D. 2007. "The normativity of meaning defended". *Analysis* 67, s. 133–40.
- . 2009. "Is meaning fraught with ought?". *Pacific Philosophical Quarterly* 90, s. 535–55.

FÖRSTÅELSE UR ETT KUNSKAPSTEORETISKT PERSPEKTIV

INLEDNING

Det säger sig självt att kunskapsteorin teoretiserar om kunskap. Men är kunskap det enda tillstånd kunskapsteoretiker bör intressera sig för? Om vi i stället benämner denna del av filosofin epistemologi, som ju ämnet kallas på åtskilliga språk, uppstår frågan om inte också andra tillstånd utöver kunskap bör undersökas. Ett mer tekniskt svar på frågan vad denna disciplins undersökningsområde närmare bestämt består i är att kunskapsteorin, eller epistemologin, undersöker *epistemiska* tillstånd. Då inställer sig härnäst förstås frågan vad som utmärker sådana tillstånd. I denna artikel kommer jag att begagna mig av William Alstons svar som lyder

[Det epistemiska] perspektivet definieras som målsättningen att maximera sanning och minimera falskhet. (Alston 1989, s. 83. Alston gör tillägget att målen gäller för en tillräckligt stor mängd övertygelser.)

Epistemiska tillstånd ska alltså kunna analyseras teleologiskt i relation till sanning (eller falskhet). Kunskap uppfyller detta villkor, men det är en öppen fråga om inte också andra tillstånd gör det.

Inom den analytiska kunskapsteorin har *förståelse* under de senaste 25 åren seglat upp som en kandidat till att just utgöra ytterligare ett epistemiskt tillstånd som bör undersökas. Förståelse har ju annars främst varit föremål för filosofisk reflektion inom hermeneutiken. Inom denna tradition har förståelse betraktats som ett utmärkande mål för humanvetenskaperna (i kontrast till naturvetenskaperna som i stället skulle sträva efter förklaringar). Som vi ska se ställer inte den diskussion inom den analytiska kunskapsteorin som denna artikel behandlar förståelse och förklaring mot varandra på detta sätt. Denna diskussion har i stället målsättningen att just undersöka förståelse som ett epistemiskt tillstånd med den klassiska analysen av kunskap som välgrundad sann övertygelse som utgångspunkt. Deltagarna försöker bland annat besvara frågan hur motsvarande analys ser ut för förståelse.

Just jämförelsen med kunskap är föremål för intensiv diskussion:

utgör förståelse ett epistemiskt tillstånd som är *distinkt* från kunskap, som *separatister* hävdar, eller utgör förståelse i stället en viss typ av kunskap som *unionister* anser? (Termerna är mina egna.) En rad kandidater till såväl förenande som särskiljande egenskaper har föreslagits och några av dem kommer att behandlas i denna artikel.

Jag kommer att ge en översikt av förslag till nödvändiga och tillräckliga villkor för förståelse. Dessa förslag strukturerar jag efter villkor som det råder enighet om i sektion 2 och därefter, i sektion 3, villkor som det i stället råder oenighet om. Allra först i denna inledande sektion presenterar jag dock kort bakgrunden till hur förståelse kommit att diskuteras inom den analytiska kunskapsteorin samt urskiljer de olika former av förståelse filosofer skiljer mellan.

De gemensamma drag som förståelse har i den nuvarande diskussionen har två källor som jag ser det. Den första utgörs av vetenskapsteorin som ofta karakteriserar vetenskapens epistemiska mål just som förståelse. (Se till exempel Strevens 2006.) Som jag påpekade ovan kontrasteras inte förståelse här mot förklaring, som i den hermeneutiska traditionen, utan betraktas i stället förenklat som det psykologiska tillstånd forskaren uppnår genom att ta del av en förklaring. Forskaren förstår alltså en förklaring. Anledningen till att termen förståelse används här snarare än kunskap är att det epistemiska tillstånd forskaren når i dessa fall är intellektuellt mer sofistikerat än åtminstone många typiska fall av kunskap. Komplicerade vetenskapliga förklaringar kräver avancerade intellektuella förmågor för att subjektet ska kunna tillägna sig dem epistemiskt.

Filosofer är inte ensamma om att betrakta förståelse som ett intellektuellt mer sofistikerat tillstånd. Inom flera pedagogiska teorier (som hos till exempel Säljö 1979 och Carlgren 1992), placeras förståelse som ett av de högre tillstånden av lärande där personen förstår det som ska läras in till skillnad från de lägre tillstånden där hen enbart inhämtar kunskap om fakta som hen memorerar och sedan kan reproducera. (Se också Peter Gärdenfors bok med den talande titeln *Lusten att förstå*, Gärdenfors 2010.)

För det andra så har denna syn på förståelse som ett intellektuellt sofistikerat epistemiskt tillstånd i sin tur inspirerat flera deltagare i de senaste decenniernas intensiva dialektik kring epistemiskt värde vilken utgör en ytterligare källa ur vilken diskussionen om förståelse sprungit. Epistemiskt värde är ett ämne för en egen artikel, men väldigt kortfattat har flera filosofer hävdat att kunskap inte är så värdefull som vi för-teoretiskt är benägna att tro utan i stället är det just förståelse som vi bör vär-

desätta ur ett epistemiskt perspektiv. Jonathan Kvanvigs omdiskuterade bok *The Value of Knowledge and the Pursuit of Understanding* (2003) utgör kanske det mest kända försöket. Förståelse skulle därmed utgöra ett högre, mer värdefullt, epistemiskt tillstånd än kunskap just i kraft av att i högre grad vara intellektuell sofistikerat. Ovan anförda exempel med kunskap om fakta utgör ett tacksamt jämförelseobjekt som illustration. Denna bakgrund förklarar delvis det fokus som den kunskapsteoretiska diskussionen av förståelse kommit att få.

Om vi nu övergår till de olika formerna av förståelse, så har just Kvanvig dragit en inflytelserik distinktion mellan "propositional" och "objectual" förståelse (Kvanvig 2003). "Propositional" förståelse består i förståelse av enskilda påståenden: ett subjekt S förstår till exempel att "Dessa meriter räcker inte för att bli astronaut" eller S förstår att "Det är artigt att bjuda igen på lunch". "Propositional" förståelse kontrasterar mot "objectual" förståelse, eller som jag kommer att kalla denna form här: *ämnesförståelse*. Ämnesförståelse varierar inte över enskilda påståenden utan över större enheter av påståenden. Sådan förståelse är holistisk. Precis hur dessa enheter eller ämnen närmare bestämt ska avgränsas är en delikat fråga, men exempel utgörs av organisk kemi, Afrikas 1800-talshistoria, finsk grammatik eller kunskapsteori. Motsvarande formulering blir "S förstår organisk kemi" etc. Sådan förståelse tillåter grader, så följaktligen kan S ha en bättre förståelse av Afrikas 1800-talshistoria än G även om G också förstår detta ämne.

Om denna form av förståelse skriver nu Kvanvig följande:

Till skillnad från kunskap kräver förståelse ett inre begripande eller insikt om hur olika element i en mängd av information är relaterade till varandra i termer av förklaringar, logiska relationer, sannolikhetssamband samt övriga relationer koherentister betraktat som konstitutiva för rättfärdigande. (Kvanvig 2003, s. 192–93)

Detta citat är representativt för en rad egenskaper som kännetecknar ämnesförståelse och som jag ska utveckla vidare i nästa sektion.

Ytterligare en form av förståelse som figurerar i diskussionen är orsaksförståelse (eng. "understanding-why"). Ett subjekt förstår exempelvis *varför* många träd faller sina löv på hösten. Svaren varierar på frågan om denna form av förståelse kan inordnas under någon av de båda ovan nämnda formerna eller utgör en egen form. Jag kommer i det följande att koncentrera mig på ämnesförståelse, så det är denna form av förståelse jag avser så länge något annat inte uttryckligen anges.

2. TRE MINDRE OMTVISTADE EGENSKAPER

I denna sektion kommer jag att lyfta fram tre epistemiska egenskaper citatet ovan förmedlar och som det råder viss enighet om.

Den första egenskapen är att förståelse väsentligen består i att (*be-*)*gripa* (engelska "grasp") *relationer* som *relata* inom ämnesområdet står till varandra, alltså påståenden inom ämnet i fråga. I citatet ovan räknar Kvanvig upp en rad sådana epistemiska relationer, men han gör troligen inte anspråk på att ge en uttömmande lista. Med hjälp av Alstons ovanstående citat går det att skilja ut genuint epistemiska relationer från sådana relationer som inte är det.

Vi kan vidare ytterligare precisera dessa epistemiska relationer som *slutledningar*. När ett subjekt begriper dessa relationer, så begriper hen just att vissa påståenden inom ämnet utgör premisser medan andra påståenden utgör slutsatser. De förra ger stöd till de senare. Genom att begripa vilka relationer olika *relata* står till varandra ges därmed ämnet en inre struktur för subjektet från att inledningsvis utgjort isolerade fragment (i linje med de pedagogiska modeller som nämndes ovan). Det är lätt att förstå att Kvanvig i ovanstående citat nämner standpunkten koherentism eftersom det är just på detta sätt som en koherentist anser att ett subjekt erhåller, samt stärker, sitt rättfärdigande av en övertygelse.

En andra egenskap, kopplad till den första, som citatet förmedlar är att förståelse har en *internalistisk* aspekt. Internalism är en term som förekommer inom flera filosofiska discipliner. Inom kunskapsteorin betecknar denna term standpunkten att ett nödvändigt villkor för rättfärdigande är att rättfärdigandet är *tillgängligt* eller åtkomligt för subjektet. Rivalen *externalism* förnekar följaktligen detta villkor. För en internalist räcker det därmed inte att ett subjekt är pålitlig i något epistemiskt avseende, till exempel med avseende på sin perception. För att subjektet också verkligen ska vara rättfärdigad i att tro något på basis av denna förmåga så måste pålitligheten dessutom vara tillgänglig för subjektet. Subjektet måste välgrundat tro att hen är pålitlig i detta avseende.

Det förefaller naturligt att en sådan internalistisk komponent finns hos begripande. Att begripa vilken relation två *relata* står till varandra innebär just att inse att ett *relatum* ger stöd åt eller följer från ett annat *relatum*. Alternativet vore att ett subjekt helt enkelt grundar en övertygelse på en annan övertygelse utan att vara medveten om att övertygelsen grundats på detta sätt. Även om jag inte själv vill förneka att sådana övertygelser kan vara välgrundade, så är det i sådana fall likväl inte fråga

om ett begripande. Åtkomlighet är ett undflyende villkor som medger grader, så det är en stor utmaning att mer precist ange *hur* åtkomlig relationen måste vara för att detta villkor ska vara uppfyllt. Hursomhelst lär detta krävande intellektuella villkor utesluta icke-mänskliga djur och små barn från att kunna nå förståelse. (Måhända en smula illvilligt misstänker jag att ytterligare ett skäl till det ökade intresse förståelse kommit att åtnjuta inom kunskapsteorin är att förståelse erbjuder en fristad för internalister.)

Det tredje villkor som kan utvinnas ur citatet från Kvanvig är att förståelse utgör ett *förmedlat* (eng. "non-basic") epistemiskt tillstånd i betydelsen *icke-grundläggande* tillstånd. (Grundläggande motsvaras alltså på engelska av "basic".) Vi uppnår *inte* förståelse genom att stå i direkt förbindelse med verkligheten som i fallet med övertygelser som vi erhåller genom perception. I stället uppnår vi förståelse genom att begripa de relationer olika *relata*, som vi redan grundat just perceptuellt (eller genom andra grundläggande epistemiska förmågor), står till varandra. Dock föreligger förstås möjligheten att ur dessa befintliga *relata* sluta sig till nya påståenden.

Förståelsens epistemiska status är därmed beroende av dessa *relata*. En skillnad mellan åtminstone min syn på förståelse och koherentismen är att den epistemiska statusen hos dessa *relata* inte nödvändigtvis ökar i styrka genom att de kommer att ingå i dessa relationer. (Andra filosofer som Ernest Sosa (2009) tycks ha en avvikande syn och tillåter att den epistemiska statusen alltid stärks i sådana lägen.) Rättfärdigande kvarstår som en asymmetrisk relation: om p rättfärdigar q, så kan inte q rättfärdiga p. Sådan epistemisk cirkularitet föreligger följaktligen inte hos förståelse (eller något annat epistemiskt tillstånd heller för den delen).

Det råder alltså viss konsensus om att förståelse utgör ett förmedlat tillstånd av begripande som inbegriper en väsentlig internalistisk aspekt av tillgänglighet. I nästa stycke ska vi lyfta fram ett par punkter kring vilken oenigheten är desto större.

3. DE MER OMTVISTADE EGENSKAPERNA

En central fråga är om förståelse är ett så kallat *faktivt* begrepp precis som kunskap är. Om ett subjekt *vet* att p, så följer det att p är sant. Gäller motsvarande förhållande för förståelse? Eftersom objektet för ämnesförståelse består av en större mängd påståenden snarare än ett enskilt sådant, så skulle faktivet här motsvaras av att de *relata* som

bildar ämnet är sanna. Kvanvig förordar standpunkten att visserligen kan några mer perifera *relata* vara falska, men merparten, och i synnerhet de mer centrala *relata*, måste vara sanna för att förståelse ska föreligga (Kvanvig 2003). Om jag förstår Afrikas 1800-tals historia, så är merparten av de påståenden som ingår i min förståelse sanna och i synnerhet de mer centrala av dem. Något enstaka årtal eller namn kan jag dock missta mig om.

Fördelen med en sådan standpunkt är att förståelse därmed på ett självklart sätt uppfyller Alstons villkor för att utgöra ett epistemiskt tillstånd. I likhet med kunskap når subjektet förståelse när målsättningen att erhålla sanna övertygelser uppfylls (tillsammans med de övriga nödvändiga villkoren för förståelse naturligtvis).

Nackdelen med denna standpunkt är att förståelse därmed tycks begränsas till sanna teorier. Men om förståelse väsentligen består i ett begripande av de relationer en mängd *relata* står till varandra, så tycks ett sådant begripande kunna äga rum oberoende av vilka sanningsvärden dessa *relata* har. Låt mig illustrera med två exempel (hämtade från Janvid 2014):

1. Vid en medeltida disputation har en respondent till uppgift att försvara en given tes gentemot en opponent. Under disputationen visar både respondenten och opponenten att de begriper ett stort antal relationer med tesen som ett centralt *relatum*, men en av dem intar en falsk standpunkt gentemot tesen. Är det verkligen därmed så att den av dem som gör det nödvändigtvis har en sämre förståelse av ämnet? Kan inte själva begripandet vara lika god hos dem båda? Dennes epistemiska status är förvisso totalt sett sämre, men är det verkligen just förståelsen som påverkas negativt och inte snarare den epistemiska statusen hos (åtminstone) tesen vilket medför att deltagaren inte vet detta påstående?

2. I det andra exemplet går vi från en intersubjektiv till en intrasubjektiv jämförelse: den skeptiska ledaren av den platonska akademien, Karneades, orsakade en skandal under sitt besök i Rom som medlem av en atensk diplomatisk delegation år 146 f.Kr. Han chockade den romerska publiken genom att först ge ett föredrag om rättvisa som han dagen efter vederlade i ett följande föredrag. Trots att den centrala tesen i respektive tal är oförenliga, så kan väl Karneades förstå ämnet lika väl i båda talen och också totalt sett? Kanske förstår han till och med ämnet bättre just genom att han var i stånd att hålla båda talen?

Båda dessa exempel illustrerar att förståelse kan föreligga trots att några av de mer centrala *relata* inom ämnet är falska. De visar åtmins-

tone att frågan om sanningen hos dessa *relata* utgör ett nödvändigt villkor för förståelse är mer komplicerad än i fallet med kunskap. Som ytterligare ett exempel lyfter Catherine Elgin i sin kritik av att förståelse är faktiskt fram idealiserade vetenskapliga modeller som ger oss förståelse trots att modellerna strängt taget är falska (Elgin 2007 och 2009).

Det intrasubjektiva fallet kanske i själva verket är vanligen förekommande, även om det oftast inte tar sig samma uttryck som hos Karneades. (Kanske borde vi dock följa hans exempel och suspendera omdömet i sådana fall?) I vår ämnesförståelse ingår det ofta oförenliga teorier och ståndpunkter inom detta ämne. Ta min egen förståelse av kunskaps-teori till exempel och i synnerhet min förståelse av förståelse. Till skillnad från Karneades suspension av omdömet, så har jag en viss teoretisk preferens inom ämnet eftersom jag lutar åt en separatistisk ståndpunkt. Skulle min förståelse med nödvändighet vara sämre om det visar sig att jag har fel i detta? Jämför med en nykomling inom ämnesområdet som inte tagit del av diskussionen i samma utsträckning och inte begripit lika många relevanta relationer, men som när en fast övertygelse om unionismens riktighet. Anta nu att hen har rätt och därmed vet något som jag inte vet, förstår hen dock nödvändigtvis därmed mer?

Notera här en tvetydighet kring ämnesförståelsens objekt: omfattar subjektets förståelse strängt taget enbart själva teorin eller innefattar förståelsen dessutom den verklighet som ämnet teoretiserar om? Om det förra alternativet avses så går det att förstå en falsk teori som teori betraktat. Följaktligen kan subjektet då förstå falska teorier och ståndpunkter inom ämnet. Subjektet begriper relationer som teorins delar står till varandra. Faktiviteten består i så fall i att teorin verkligen innehåller de komponenter hen förstår. På det sättet kan vi förstå Aristoteles uralstringsteori trots att den är falsk. Om förståelsen i stället också innefattar den verklighet som teorin teoretiserar om, så går det bara att förstå sanna teorier. Jag förordar själv det första alternativet. Den överensstämmelse mellan teori och verklighet som föreligger vid förståelse enligt det andra alternativet verkar inte vara en sådan relation som ingår i själva förståelsen. Relationen mellan teori och verklighet är inte något som subjektet begriper givet de egenskaper jag presenterade i den förra sektionen, bland annat eftersom verkligheten själv i så fall ingår som ett *relatum* i överensstämmelsen enligt det alternativet.

Frågan om faktivitet ställer också förståelsens oberoende av kunskap i fokus. Om förståelse inte utgör ett faktiskt tillstånd, så utgör inte förståelse heller en typ av kunskap. Däremot följer det inte att om förståelse är

faktivt, så utgör förståelse en typ av kunskap eftersom andra skillnader kan föreligga.

Kvanvigs ståndpunkt utgör ett exempel på det senare alternativet. Han betonar förståelsens faktivit, men är ändå separatist eftersom han hävdar att förståelse är förenlig med tur vilket kunskap inte är. Kvanvig anser alltså att tursam förståelse kan föreligga, men däremot inte tursam kunskap. Enligt en vanlig tolkning är det i kraft av den oförenligheten som så kallade Gettier-fall fungerar som motexempel till den klassiska analysen av kunskap som sann, välgrundad övertygelse. I fallet förståelse ger Kvanvig ett omdiskuterat exempel på ett subjekt som *råkar* välja en vederhäftig bok om något ämne och därmed når förståelse genom att läsa boken. (Kvanvig 2003) Det tursamma består i att subjektet lika gärna hade kunnat välja en annan bok skriven av en charlatan. (Subjektet befann sig på ett dåligt sorterat bibliotek.) På grund av de tursamma omständigheterna saknar subjektet kunskap, men likväl förstår hen ämnet i fråga i kraft av den tursamt valda, men likväl vederhäftiga, boken. Kvanvig har inte fått stå oemotsagd. (Se till exempel översikterna i Baumberger, Beisbart & Brun 2017 och Grimm 2011 samt Janvid 2012.)

Jag avslutar denna artikel med att peka på en mindre uppmärksammas skillnad mellan övertygelse och kunskap med avseende på subjektets övertygelser. Om förståelse verkligen består i att begripa de relationer olika *relata* står till varandra inom ett ämne, så går det att förstå ett ämne utan att tro på någon av ämnets komponenter. Närmare bestämt kan ett subjekt förstå ett ämne utan att tro på någon av de ingående *relata*. Subjektet kan förstå att *relata* x och y ger stöd åt, eller förklarar, z utan att tro att vare sig x, y eller z. Karneades i exemplet ovan är ett exempel på detta och de båda deltagarna i den medeltida disputationen kan mycket väl vara det. Kanske är det dessutom möjligt att ha falska övertygelser rörande dessa *relata* eftersom begripandet av de relationer de står till varandra inte behöver påverkas av deras sanningsvärden?

Med dessa frågor avslutar jag min översikt av förståelse: ett område inom vilken diskussionen lär fortsätta att vara intensiv!

LITTERATUR

- Alston, W. 1989. "Concepts of Epistemic Justification", återutgiven i *Epistemic Justification*, s. 81–114. Ithaca: Cornell University Press.
- Baumberger, C., Beisbart, C. & Brun, G. 2017. "What is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science".

- I *Explaining Understanding. New Perspectives from Epistemology and Philosophy of Science*, red. S. Grimm, C. Baumberger och S. Ammon, s. 1–34. New York och London: Routledge.
- Carlgrén, I. 1992. *Kunskap och lärande*. Läroplanskommitténs betänkande Skola för bildning, SOU 1992:94, Stockholm.
- Elgin, C. 2007. "Understanding and the Facts". *Philosophical Studies* 132, s. 33–42.
- . 2009. "Is Understanding Factive?". I *Epistemic Value*, red. A. Haddock, A. Millar och D. Pritchard, s. 322–30. Oxford: Oxford University Press.
- Grimm, S. 2011. "Understanding". I *The Routledge Companion to Epistemology*, red. S. Bernecker och D. Pritchard. New York och London: Routledge, s. 84–94.
- Gärdenfors, P. 2010. *Lusten att förstå: Om lärande på människans villkor*. Stockholm: Natur & Kultur.
- Janvid, M. 2012. "Knowledge versus Understanding: The Cost of Avoiding Gettier". *Acta Analytica* 27, s. 183–97.
- . 2014. "Understanding Understanding: An Epistemological Investigation". *Philosophia* 42, s. 971–85.
- Kvanvig, J. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
- Sosa, E. 2009. *Reflective Knowledge. Apt Belief and Reflective Knowledge, Volume II*. Oxford: Oxford University Press.
- Strevens, M. 2006. "Scientific Explanation". I *The Encyclopedia of Philosophy*, red. D. Borchert. Andra upplagan. New York: Macmillan.
- Säljö, R. 1979. "Learning about Learning". *Higher Education* 8, s. 443–51.

HUMES SENTIMENTALISM – INTE NON-KOGNITIVISM

1. INLEDNING

Det har länge varit en spridd uppfattning att David Hume (1711–1776) var en tidig non-kognitivist om moral. På senare tid har den uppfattningen kommit att bli föremål för kritik som enligt min mening är berättigad. Non-kognitivistiska tolkningar av Humes metaetik bygger ofta på missförstånd och på att man inte gjort erforderliga metaetiska distinktioner. Jag ska i denna uppsats argumentera för att det finns mycket lite som talar *för* non-kognitivistiska tolkningar av Hume och ganska mycket som talar *mot* dem.¹

Jag ska börja med att göra några förtydliganden om non-kognitivism i avsnitt 2. Därefter ska jag i avsnitt 3 försöka förklara Humes sentimentalistiska teori om moralomdömen. Därefter ska vi i avsnitt 4 och 5 se närmare på två argument som ofta anförs för non-kognitivistiska tolkningar av Hume, nämligen det s.k. *motivationsargumentet* och *Humes lag*. Slutligen ska vi i avsnitt 6 se att det också finns god evidens mot non-kognitivistiska tolkningar.

2. NON-KOGNITIVISM: PSYKOLOGISK OCH SEMANTISK

Man säger ofta att non-kognitivism är en teori om *moraliska omdömen*. Men detta är inte tillräckligt precist, eftersom ”moraliskt omdöme” är ett tvetydigt uttryck. För det första kan det beteckna det psykologiska tillstånd man befinner sig i när man hyser en moralisk uppfattning om vad som är rätt eller fel, dygdigt eller odygdigt, etc. Att anse t.ex. att tortyr är fel är, enligt non-kognitivism om detta psykologiska tillstånd, att hysa en icke-kognitiv attityd gentemot tortyr, t.ex. *ogillande*. Låt oss kalla denna typ av non-kognitivism *psykologisk non-kognitivism*. Enligt psykologisk *kognitivism* hyser man en övertygelse (med moraliskt innehåll) om tortyr då man tror att tortyr är fel.

1. Non-kognitivistiska tolkningar av Humes metaetik har försvarats av bland andra Anthony Flew (1963), Robert Fogelin (1979), John Brice (1996) och mera nyligen av Fabio Shecaira (2011). De har kritiserats av bland andra Elizabeth Radcliffe (2006), Rachel Cohon (2008), Nicholas Sturgeon (2008) och Charles Pigden (2010). Jag har tidigare diskuterat Humes metaetik, inklusive non-kognitivistiska tolkningar av den, i Olson 2011.

För det andra kan "moraliskt omdöme" beteckna yttrandet av en moralisk sats, till exempel "Tortyr är fel". Enligt non-kognitivism om moraliska yttranden ska meningen hos ett sådant yttrande inte förstås i termer av dess sanningsvillkor eller i termer av en proposition som det uttrycker, för enligt traditionella versioner av non-kognitivism om moraliska yttranden har det inga sanningsvillkor och uttrycker ingen proposition. Detta beror på att yttrandet konventionellt uttrycker en icke-kognitiv attityd, såsom gillande och ogillande. Låt oss kalla denna typ av non-kognitivism *semantisk non-kognitivism*. Låt oss kalla uppfattningen att yttranden av moraliska satser är hävdanden som uttrycker moraliska propositioner *semantisk kognitivism*.

Det är viktigt att skilja mellan psykologisk och semantisk non-kognitivism eftersom de är logiskt oberoende. Till exempel kunde man anse att åsikten att tortyr är fel består i ogillande av tortyr, men att yttrandet av satsen "Tortyr är fel" inte konventionellt *uttrycker* ogillande av tortyr, utan att det konventionellt *rapporterar* att talaren ogillar tortyr. På detta sätt kan psykologisk non-kognitivism kombineras med semantisk kognitivism. För frågor om hur Hume ska tolkas är detta relevant eftersom evidensen som ofta framförs för non-kognitivistiska tolkningar stödjer tolkningar av Hume som *psykologisk non-kognitivist*, men inte som *semantisk non-kognitivist*. (Hume och hans samtida hade i själva verket mycket lite att säga om moralens semantik.)

3. HUMES SENTIMENTALISM

Den metaetiska frågeställning Hume är mest intresserad av är den *epistemologiska* frågan om hur vi skiljer mellan dygd och last, rätt och fel och gott och ont. I synnerhet vill han undersöka "huruvida det är möjligt att genom förnuftet ensamt skilja mellan moraliskt gott och ont" (A 3.1.1.4).² Han argumenterar framför allt för att vi inte kan göra detta med hjälp av det blotta förnuftet (*reason*); ytterst beror våra moraliska omdömen på känslor (*sentiments*). Med andra ord argumenterar Hume mot *rationalism* och för *sentimentalism* om moralomdömen.

I sin *Enquiry Concerning the Principles of Morals* (1998 [1751]) definierar Hume moralisk dygd som "vilken som helst mental handling eller egenkap som hos en betraktare väcker den angenäma känslan av gillande, och

2. I mina hänvisningar till Hume står "A" för Avhandling om den mänskliga naturen. Siffrorna som följer anger i tur och ordning bok, del, sektion och stycke.

last det motsatta". (EPM App. 1.10, min övers, Humes kursivering.)³ Det som i grund och botten ger oss insikter om dygder och laster är alltså våra känslor av gillande och ogillande av olika karaktärsdrag. Blotta förnuftet kan å sin sida bara ge oss insikter om två saker: relationer mellan idéer (ungefär det som i modernt filosofiskt språkbruk kallar analytiska och logiska sanningar) och orsaksrelationer. Förnuftet kan inte säga oss vilka karaktärsdrag som gillas eller ogillas då vi intar "ett visst gemensamt perspektiv" ("*the common point of view*") (A 3.3.1.30), ur vilket vi bortser från våra personliga preferenser, smaker och böjelser; vi måste själva uppleva dessa känslor av gillande eller ogillande genom att inta detta gemensamma perspektiv. Men när vi väl har upplevt sådana känslor kan vi med förnuftets hjälp inse att liknande karaktärsdrag ger, eller skulle ge, upphov till liknande känslor hos liknande betraktare. Detta eftersom "lika orsaker alltid åstadkommer lika verkningar" (A 1.3.15.8). Det är i denna mening moraliska distinktioner inte härrör (enbart) från förnuftet (T 3.1.1).

Eftersom Humes sentimentalism handlar om moraliska omdömens epistemologi har den ingen direkt implikation för moraliska omdömens semantik. Men det kan vara en frestande tanke att Humes sentimentalism passar bättre ihop med psykologisk non-kognitivism än med psykologisk kognitivism. Men det finns goda skäl att motstå den frestelsen. Begrunda till exempel den psykologiskt kognitivistiska teorin att uppfattningen att ett visst karaktärsdrag, t.ex. generositet, är en dygd består i övertygelsen att en betraktare som intagit det gemensamma perspektivet gillar, eller skulle gilla, generositet. Förespråkaren av denna teori kunde hävda att övertygelsen att en betraktare som intar det gemensamma perspektivet gillar, eller skulle gilla, generositet måste baseras på en känsla av gillande man själv har när man betraktar karaktärsdraget utifrån det gemensamma perspektivet. På detta sätt passar Humes sentimentalism väl ihop med åtminstone någon form av psykologisk kognitivism.

Emellertid brukar de som försvarar non-kognitivistiska tolkningar av Hume inte i första hand hänvisa till hans sentimentalism, utan till vad som kommit att kallas "Humes motivationsargument" och till Humes lag. Låt oss härnäst se närmare på motivationsargumentet, vilket också är bekant från den moderna metaetiska debatten.

3. I mina hänvisningar till Hume står "EPM" för *Enquiry concerning the Principles of Morals*. Siffrorna som följer anger i tur och ordning sektion och stycke i den utgåva som är redigerad av T. Beauchamp. ("App." står för bokens appendix.)

4. MOTIVATIONSARGUMENTET

Enligt Hume kan moralen motivera oss att handla (A 3.1.1.6). Förnuftet är däremot "svält och distanserat" ("*cool and disengaged*") och "utgör inget motiv för handling" (EPM App. 1.21). Humes motivationsargument brukar framställas så här (baserat på A 3.1.1.6):

- (M) "Moralen väcker passioner och framkallar eller förhindrar handlingar."
- (F) "Förnuftet i sig självt är i detta avseende fullständigt vanmäktigt."
- (S) "Moralens regler är således inte slutsatser av vårt förnuft."

Men som många kommentatorer har påpekat är det långt ifrån klart att argumentet är giltigt. För att få ett giltigt argument brukar man omformulera det på ungefär följande vis:

- (M*) Moraliska omdömen väcker passioner och framkallar eller förhindrar handlingar.
- (F*) Slutsatser av vårt förnuft kan inte i sig själva väcka passioner eller framkalla eller förhindra handlingar.
- (S*) Moraliska omdömen är således inte slutsatser av vårt förnuft.

Men argumentet är fortfarande inte uppenbart giltigt. Man kunde ju tänka sig att moraliska omdömen är förnuftsomdömen som väcker passioner och framkallar eller förhindrar handlingar därför att vi människor är disponerade att motiveras att handla i enlighet med våra moraliska omdömen, kanske därför att vi bryr oss om moralen och generellt *vill* vara moraliska. Utan denna disposition skulle moraliska omdömen inte vara motiverande eftersom förnuftsomdömen inte i sig själva kan motivera. På detta sätt skulle (M*) och (F*) kunna vara sanna och (S*) falsk.

Den vanliga manövern för att få till ett giltigt argument är att tolka premisserna (M) och (M*) som att de uttrycker *motivationell internalism* om moralomdömen.⁴ Motivationell internalism är tesen att det är nödvändigt att om en person accepterar ett moralomdöme så är hen i någon grad motiverad att handla i enlighet med omdömet. Internalism av denna typ är en kontroversiell uppfattning. Faktum är att vi inte behöver tillskriva Hume denna uppfattning för att argumentet ska vara giltigt. Det räcker med att tillskriva honom den modalt svagare tesen att moraliska omdömen *kan* vara handlingsmotiverande i sig själva, dvs. utan

4. Se t.ex. Shecaira 2011.

assistans av ett oberoende psykologiskt tillstånd, såsom en disposition att göra det man tror är rätt, eller en vilja att vara moralisk. Låt oss kalla denna tes den modalt svagare internalismen. Huruvida Hume omfattade någon form av internalism är också kontroversiellt, men låt oss för argumentets skull gå med på att tillskriva honom åtminstone den modalt svagare versionen av internalism. Ger detta stöd åt non-kognitivistiska tolkningar?

Låt oss först undersöka huruvida internalism ger stöd åt *semantisk* non-kognitivism. Enligt min mening föreligger det ingen direkt koppling mellan internalism och semantiska teorier om moralomdömen. Detta beror på att den föregivna kopplingen mellan moralomdömen och motivation att handla råder mellan det psykologiska tillståndet att hysa en moralisk uppfattning och motivation att handla. Det vore inte rimligt att hävda att kopplingen råder mellan (uppriktiga) yttranden av moraliska satser och motivation att handla, för i många situationer accepterar vi ju moraliska omdömen och är motiverade att handla i enlighet med dem, utan att yttra några moraliska satser.

Kanske kunde man hävda att internalism passar bättre ihop med semantisk non-kognitivism än med semantisk kognitivism. Men det är oklart vad det innebär att en teori om moralomdömens psykologi och deras koppling till handlingsmotivation passar bra eller dåligt ihop med teorier om moralomdömens semantik. En rimligare tanke är att internalism passar bättre ihop med psykologisk non-kognitivism än med psykologisk kognitivism, åtminstone givet vissa populära teorier om vilka typer av psykologiska tillstånd som är handlingsmotiverande. Men som vi strax ska se accepterade Hume inte dessa teorier.

Den psykologiska non-kognitivismen kan enkelt förklara den starka och eventuellt nödvändiga kopplingen mellan att acceptera ett moralomdöme och att vara motiverad att handla i enlighet med omdömet. För enligt psykologisk non-kognitivism är ju moralomdömen attityder av gillande eller ogillande och sådana attityder är ju i sig själva handlingsmotiverande.

Men denna förklaring av kopplingen mellan moralomdömen och motivation ger stöd åt psykologisk non-kognitivism endast under förutsättning att inget kognitivt tillstånd – dvs. ingen övertygelse (*belief*) – är, eller kan vara, i sig motiverande. Det är en vanlig uppfattning inom modern metaetik och handlingsfilosofi att övertygelser (*beliefs*) är en typ av tillstånd som i sig inte kan motivera till handling, medan begär (*desires*) är en typ av tillstånd som är i sig handlingsmotiverande. Enligt vad som

ofta kallas den "Humeanska motivationsteorin" krävs det att en agent har ett begär och en relevant övertygelse om medel för att tillfredsställa begäret, för att hen ska vara motiverad att handla. Denna teori bygger vanligtvis på funktionalistiska teorier om övertygelser och begär. Tanken är att övertygelser och begär har olika funktioner eller *directions of fit*: övertygelsers funktion är att *representera* världen (*mind-to-world direction of fit*), medan begärens funktion är att *förändra* världen (*world-to-mind direction of fit*).⁵

Accepterade Hume den Humeanska motivationsteorin? Ironiskt nog är svaret troligen nej. Åtminstone accepterade han inte tanken att inga övertygelser kan vara i sig själva handlingsmotiverande. Hume försvarade den mycket mer begränsade tesen att ingen övertygelse som helt och hållet är grundad i förnuftet kan vara motiverande. Men den tesen är alldeles för begränsad för att ge stöd åt non-kognitivism.

Varför har man då trots att Hume accepterade den Humeanska motivationsteorin och i synnerhet tanken att ingen övertygelse kan vara i sig handlingsmotiverande? En möjlig förklaring är att man tänkt sig att Hume ansåg att endast förnuftet genererar övertygelser. Men detta kan knappast ha varit Humes ståndpunkt. Till exempel är min övertygelse att jag nu är hungrig grundad i ett sinnesintryck och inte resultatet av ett förnuftsresonemang. Detsamma gäller många övertygelser om omvärlden som är baserade på perception.

Hume tänkte sig som sagt inte att vilka sorters övertygelser som helst kan motivera till handling. Det är främst, och kanske bara, övertygelser om njutning och smärta som kan vara handlingsmotiverande. Innan vi ser närmare på två belysande citat måste jag säga något om Humes medvetandefilosofi. Enligt Hume är allt medvetandehåll perceptioner. Dessa är av två slag – *idéer* och *intryck* – som skiljer sig åt fenomenologiskt och etiologiskt. Intrycken är fenomenologiskt livliga medan idéerna är stillsamma kopior av intrycken. Övertygelser är ett slags idéer och begär är ett slags (reflektions)intryck.⁶ Betrakta nu Humes illustration av hur ett (sensoriskt) intryck genererar en idé, vilken i sin tur genererar ett (reflektions)intryck som är handlingsmotiverande:

Ett intryck träffar först våra sinnen och får oss att varsebli hetta eller köld, törst eller hunger, lust eller smärta av det eller andra slaget. Från

5. Se t.ex. Smith 1994, kap. 4, för en inflytelserik diskussion.

6. Hur skiljer sig idéer som är övertygelser från idéer som inte är övertygelser? Även här hänvisar Hume till fenomenologiska skillnader: idéer som är övertygelser upplevs som livligare (A 1.3.7.5) eller stadigare och fastare (A, Tillägg 3).

detta intryck gör [medvetandet] en kopia som kvarstår efter att intrycket upphört, och detta kallar vi en idé. Denna idé om lust eller smärta åstadkommer, då den återvänder till själen, de nya intryck av begär och motvilja, hopp och rädsla, vilka med rätta kan kallas reflektionsintryck, eftersom detta är deras härkomst. (A 1.1.2.1)

Hume säger också att

erfarenheten lär oss att *idéerna* om de [njutbara eller obehagliga] objekt om vilka vi tror antingen att de existerar eller att de kommer att existera, åstadkommer *samma verkan*, om än i lägre grad, som de vilka är omedelbart närvarande i sinneserfarenheten och varseblivningen (A 1.3.10.3, mina kursiveringar).

Notera att Hume säger att det är *idén* om lust eller smärta som åstadkommer nya intryck av begär och motvilja, och att *idéer* om njutbara eller obehagliga objekt om vilka vi tror antingen att de existerar eller att de kommer att existera, åstadkommer samma verkan, om än i lägre grad, som de *intryck* vilka är omedelbart närvarande i sinneserfarenheten och varseblivningen. Hume tycks alltså inte mena att övertygelser om njutning och smärta kan motivera till handling endast om de kombineras med relevanta och oberoende begär. Han tycks mena att övertygelser om njutning och smärta själva ger upphov till handlingsmotivation genom att generera intryck av begär och motvilja.

Eftersom Hume menar att det är njutbart att vara dygdig och att det normalt är njutbart att observera andras dygdighet skulle han kunna mena att moraliska övertygelser kan motivera till handling på liknande sätt, dvs. genom att ge upphov till begär och aversioner. Moraliska övertygelser, eller moraliska idéer, *är*, enligt denna tolkning, kopior av intryck av vissa typer av gillande och ogillande.

En sådan teori är förenlig med den modalt svaga formen av internalism, enligt vilken moralomdömen kan vara handlingsmotiverande i sig själva. För tanken är inte att moraliska övertygelser kan motivera till handling endast givet andra oberoende psykologiska tillstånd; tanken är att moraliska övertygelser, precis som övertygelser om vad som är njutbart och vad som är obehagligt mer generellt, kan generera handlingsmotivation i kraft av sina speciella innehåll.

Innehållet i moraliska övertygelser, och i övertygelser om vad som är njutbart respektive smärtsamt, är speciellt eftersom det enligt Hume finns ”inplanterat i det mänskliga medvetandet en varseblivning av ...

smärta och lust som är den huvudsakliga källan och motiverande principen för all dess verksamhet.” (A 1.3.10.2) En konsekvens av detta är att när vi tror att ett visst handlingsalternativ kommer att vara njutbart eller kommer att leda till njutning, så är vi motiverade att utföra det. På motsvarande sätt är vi motiverade att inte utföra handlingen om vi tror att den kommer att vara obehaglig eller leda till smärta. Andra typer av övertygelser, vilkas innehåll inte handlar om vad som är njutbart eller obehagligt, genererar inte handlingsmotivation på detta sätt. I synnerhet kan inte övertygelser som är resultat av rena förnuftsresonemang generera handlingsmotivation.

Humes motivationsargument är alltså riktat mot moralisk rationalism och det utgör inget hinder mot att tolka honom som psykologisk kognitivist. Vi har härmed underminerat stödet som motivationsargumentet sägs ge åt non-kognitivistiska tolkningar. Innan vi begrundar passager som på ett mer direkt vis talar mot non-kognitivistiska tolkningar ska vi undersöka huruvida Humes lag ger stöd åt dem.

5. HUMES LAG

Det brukar sägas att Humes lag är tesen att inget *bör* följer från *är*. Tesen verkar rimlig eftersom logiken är konservativ; man får inte ut i slutsatsen vad man inte stoppat in i premisserna. Det finns tekniska problem med detta intuitiva sätt att formulera Humes lag men vi ska inte fördjupa oss i dem. Låt oss i stället undersöka vad Hume vill åstadkomma med sin diskussion om *är* och *bör*. Till att börja med är det värt att notera att diskussionen dyker upp som en avslutande reflektion i Humes argumentation mot rationalismen, där han försöker visa att moraliska omdömen inte är rena slutsatser av vårt förnuft. Låt oss se på passagen i sin helhet:

Till dessa resonemang kan jag inte låta bli att foga en iakttagelse som kanske är av viss betydelse. I varje morallära jag hittills stött på har jag kunnat notera att författaren ett tag håller sig till det vanliga sättet att resonera, och fastställer att en Gud existerar eller gör observationer beträffande mänskliga angelägenheter. Men sedan upptäcker jag plötsligt till min förvåning att i stället för de vanliga bindeleden i satser, *är* och *är inte*, så binds alla satser samman genom *bör* eller *bör inte*. Förändringen sker omärkligt, men är likväl av största betydelse. För eftersom detta *bör* eller *bör inte* uttrycker en ny relation, eller ett nytt hävdande, skulle det behöva uppmärksammas och förklaras, samtidigt som man måste ge ett skäl för

något som tycks helt obegripligt: hur denna nya relation kan härledas från andra som är av ett helt annat slag. Men eftersom författare vanligen inte vidtar denna försiktighetsåtgärd, tillåter jag mig i stället att rekommendera den till läsarna, och jag är övertygad om att denna obetydliga eftertanke skulle räcka för att undergräva alla gängse moralläror och låta oss se att skillnaden mellan last och dygd inte är grundad på de blotta relationerna mellan föremål eller uppfattas av förnuftet. (A 3.1.1.27).

De som i denna passage tycker sig finna stöd för non-kognitivismen brukar hävda att non-kognitivismen på ett elegant sätt förklarar varför inget *bör* kan härledas från är. Förklaringen sägs vara att eftersom *bör*-satser inte uttrycker några propositioner så uttrycker de ingenting som kan stå i logiska relationer till propositioner som uttrycks av *är*-satser. Detta visar, har somliga menat, att det finns en oöverbryggbar klyfta mellan fakta och värde. Vad denna klyfta består i förklaras också av non-kognitivismen: *bör*-satser uttrycker non-kognitiva attityder medan *är*-satser inte gör det.

Men för det första menade Hume nog inte att det finns en klyfta mellan fakta och värde. Det hävdas ibland att Hume förnekade att moralen, inklusive dygden och lasten, består i något som är en fråga om fakta. Men Hume försvarade de mer kvalificerade uppfattningarna att "moralen inte består i något som är en *fråga om fakta*, som kan upptäckas genom förståndet" (A 3.1.1.26, Humes kursivering), samt att "last och dygd inte är en fråga om fakta, vars existens vi kan sluta oss till genom förnuftet" (A 3.1.1.26). Hume förnekar alltså inte att det finns moraliska fakta eller värdefakta. Hans poäng är, som vi tidigare sett, att de inte kan inses genom blotta förnuftet. Moraliska fakta och värdefakta är fakta om känslor och vi kan inte inse dem endast med hjälp av förnuftet.

För det andra behöver man ju, som vi också sett, inte åberopa non-kognitivismen för att förklara varför *bör* inte följer från är. Det räcker med att påpeka att logiken är konservativ. Men då kvarstår ju frågan vad Hume ville säga med *är/bör*-passagen.

För att ge ett adekvat svar kan vi börja med att beakta Annette Baiers påpekande att ordet "*deduce*" på Humes tid hade en vidare betydelse än det normalt har i dagens fackfilosofiska språkbruk.⁷ I 1700-talsengelskan, liksom i modern vardaglig engelska, kan "*deduce*" helt enkelt betyda att man från ett antal premisser drar en slutsats om vad som är sant eller sannolikt. Det behöver inte vara fråga om logisk härledning. Detta talar för

7. Baier 2010, s. 52.

att Humes påpekande åtminstone inte *bara* handlar om att det inte går att logiskt härleda *bör* från *är* (som Humes lag säger), utan att det *tycks* helt obegripligt (*"seems altogether inconceivable"*) hur man skulle kunna sluta sig till vad som *bör* vara fallet endast från premisser om vad som *är* fallet. Det kan med andra ord tyckas omöjligt att i vid mening dra slutsatser om vad som *bör* vara fallet från premisser om vad som *är* fallet.

Men Humes teori visar till att börja med hur det är både möjligt och rimligt att dra moraliska slutsatser från premisser om vad som *är* fallet. För enligt Humes teori är egenskapen att vara en dygd egenskapen att vara ett karaktärsdrag eller en mental egenskap som väcker gillande, eller skulle väcka gillande, hos en betraktare som intagit det gemensamma perspektivet. Denna definition förklarar varför det är rimligt att från premissen att en viss person är välvillig, skarpsinnad, modig, etc., sluta sig till att personen *är*, åtminstone i dessa avseenden, dygdig: det är rimligt eftersom en betraktare som intagit det gemensamma perspektivet gillar, eller skulle gilla, dessa karaktärsdrag. Och dygder är ju definierade som karaktärsdrag som väcker gillande, eller skulle väcka gillande, hos en betraktare som intagit det gemensamma perspektivet. Om vi dessutom tänker oss att dygder är karaktärsdrag som vi *bör* ha och om vi i linje med Humes teori om dygder också tillskriver honom teorin att vad det innebär att vi *bör* ha ett visst karaktärsdrag är att det väcker gillande, eller skulle väcka gillande, hos en betraktare som intagit det gemensamma perspektivet, så har vi en förklaring till hur vi från premissen att ett visst karaktärsdrag är sådant att det väcker, eller skulle väcka, sådant gillande, kan sluta oss till att vi *bör* ha karaktärsdraget i fråga.⁸ Vi kan göra detta via en premiss som utgörs av Humes teori om vad det innebär att ett visst karaktärsdrag är sådant att vi *bör* ha det.

Som vi har sett säger Hume att *bör* eller *bör inte* uttrycker en ny relation eller hävdande, men med detta menar han nog bara att när en författare yttrar sig om vad man *bör* göra, så uttrycker dessa *bör* eller *bör inte* en relation som är ny i förhållande till det som författaren tidigare sagt om Guds existens, eller om mänskliga angelägenheter eller om vad det må vara. Enligt Hume behöver dessa *bör* eller *bör inte* uppmärksammas och

8. Hume uttalar sig sällan explicit om vad vi *bör* göra eller om hur vi *bör* vara. Men det är värt att notera att han skiljer mellan naturliga och moraliska förpliktelser (obligations). Att en person har en naturlig förpliktelse att bete sig på ett visst sätt innebär att det gynnar, eller skulle gynna, hennes egenintresse att bete sig på detta sätt. Att en person har en moralisk förpliktelse att bete sig på ett visst sätt innebär att hon väcker, eller skulle väcka, gillande genom att bete sig på detta sätt, eller att hon väcker, eller skulle väcka, ogillande genom att inte bete sig på detta sätt. Se t.ex. A 3.2.2.23.

förklaras. Och Humes teori erbjuder en sådan förklaring, men rationalisten har ingen rimlig eller begriplig förklaring att erbjuda. Rationalisten har inget upplysande eller ens begripligt att säga om vad dygder är eller om vad det innebär att ett visst karaktärsdrag är sådant att vi bör ha det.

Passagen om *är* och *bör* kan alltså förstås som en konkluderande reflektion, eller sammanfattning, av Humes genomgående argumentation i början av den tredje boken i *Avhandling om den mänskliga naturen*. Den går ut på att den moraliska rationalismen är oklar och föga förklarande. I den sektion som omedelbart följer argumenterar Hume för att hans sentimentalistiska teori erbjuder större förklaringskraft och är jämförelsevis enklare och mera begriplig (A 3.1.2). Det förtjänar att påpekas att Hume inte fann det värt att inkludera passagen om *är* och *bör* i *Enquiry concerning the Principles of Morals*, som är en komprimerad omarbetning av den tredje boken i *Avhandling om den mänskliga naturen*. Kanske ansåg han att den var en överflödig reflektion eller sammanfattning. Detta är en smula ironiskt med tanke på den omfattande debatt den har genererat. Mycket talar för att Humes poäng om *är* och *bör* har ganska lite att göra med det som kommit att kallas "Humes lag".

6. EVIDENS MOT NON-KOGNITIVISTISKA TOLKNINGAR

Hume säger i inledningen till *Enquiry concerning the Principles of Morals* att han vill diskutera "huruvida vi tillägnar oss kunskap om [moraliska distinktioner] genom en kedja av argument och induktion eller genom en omedelbar känsla och känsligare inre sinne". (EPM 1.3, min övers.) Denna formulering är en tydlig antydning om att Hume tog för givet att vi kan ha moralisk kunskap; frågan är hur vi tillägnar oss den. Det rimmar illa med den psykologiskt non-kognitivistiska teorin att moraliska omdömen är icke-kognitiva attityder eller passioner. Passioner är, enligt Hume, icke-representationella tillstånd; de är "ursprungliga existenser", dvs. de är inte kopior av intryck, och de kan därför inte vara sanna eller falska (A 2.3.3.5). Eftersom kunskap implicerar sanning tyder detta på att vetenskapen att en person är dygdig, eller att ett karaktärsdrag är en dygd, består i en livlig idé om personen som dygdig eller om karaktärsdraget som en dygd. Denna idé kan härledas till ett intryck av gillande av personen eller karaktärsdraget.

Det är också värt att återvända till Humes definition av dygd som "*vilken som helst mental handling eller egenskap som hos en betraktare väcker den angenäma känslan av gillande*, och last det motsatta." (EPM App. 1.10, min övers, Humes kursivering.) Den ger vid handen att upp-

fattningen, liksom utsagan, att ett visst karaktärsdrag är en dygd har propositionellt innehåll och sanningsvillkor. Men traditionella former av non-kognitivism förnekar ju att moraliska uppfattningar och utsagor har propositionellt innehåll och sanningsvillkor.

Av allt detta drar jag slutsatsen att non-kognitivistiska tolkningar av Humes metaetik inte håller streck. Men jag upprepar att inget av detta undergräver Humes sentimentalism. Min tolkning av Hume som psykologisk och semantisk kognitivist är helt förenlig med hans tes att det är omöjligt att "genom förnuftet ensamt skilja mellan moraliskt gott och ont" (A 3.1.1.4). Det är i denna mening som moral är en fråga om känsla snarare än om omdömen (A 3.1.2.1).⁹

LITTERATUR

- Baier, Annette 2010. "Hume's Own 'Ought' Conclusions". I *Hume on Is and Ought*, red. C. Pigden. Basingstoke: Palgrave Macmillan, s. 49–64.
- Bricke, John 1996. *Mind and Morality*. Oxford: Clarendon Press.
- Cohon, Rachel 2008. *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press.
- Flew, Anthony 1963. "On the Interpretation of Hume", *Philosophy* 38, s. 178–82.
- Fogelin, Robert 1979. "Review of Jonathan Harrison, *Hume's Moral Epistemology*", *Noûs* 13, s. 523–25.
- Hume, David 1998 [1751]. *An Enquiry concerning the Principles of Morals*, red. T. L. Beauchamp. Oxford: Oxford University Press.
- Hume, David 2004. *Avhandling om den mänskliga naturen*, tre böcker, övers. R. Callergård och J. Jakobsson. Stockholm: Thales.
- Olson, J. 2011. "Förnuft och känsla: Humes metaetik". I *Förnuft, känsla och moral: Perspektiv på David Hume*, red. R. Callergård. Stockholm: Thales, s. 197–216.
- Pigden, Charles 2010. "If Not Non-Cognitivism, Then What?" I *Hume on Motivation and Virtue*, red. C. Pigden. Basingstoke: Palgrave Macmillan, s. 80–104.
- Radcliffe, Elizabeth 2006. "Moral Internalism and Moral Cognitivism in Hume's Metaethics", *Synthese* 152, s. 353–70.
- Shecaira, Fabio 2011. "Hume and Non-Cognitivism", *History of Philosophy Quarterly* 28, s. 267–87.
- Smith, Michael 1994. *The Moral Problem*. Oxford: Blackwell.
- Sturgeon, Nicholas 2008. "Hume's Metaethics: Is Hume a Moral Non-Cognitivist?" I *A Companion to Hume*, red. E. S. Radcliffe. Oxford: Wiley-Blackwell, s. 513–28.

9. Uppsatsen bygger på två föredrag hållna vid Filosofidagarna i Linköping 2015, respektive Filosofiska föreningen vid Linköpings universitet i november samma år. Jag tackar åhörarna för hjälpsamma kommentarer och diskussioner.

NOTISER

FÅNGARNAS DILEMMA

Spelteorins mest välkända exempel på en klurig valsituation är det som brukar kallas "fångarnas dilemma". Detta är i alla fall det vanligaste svenska namnet. Det ursprungliga amerikanska namnet var däremot "prisoner's dilemma". Singularis, alltså. Det vill säga "fångens dilemma".

Man kunde kanske tro att det inte spelar någon roll om man väljer singularis eller pluralis, eftersom spelet är helt symmetriskt. De bägge spelarna ställs var för sig inför helt likartade problem.

Men ett annat sätt att se på saken är att "fångarnas dilemma" kan syfta på det problem som de två spelarna ställs inför som en kollektiv aktör, vilket inte är detsamma som de ställs inför var för sig.

I själva verket är detta den språkligt mest korrekta tolkningen. Om man syftade på de enskilda individernas problem borde det ju snarare heta "fångarnas *dilemman*", i genomgående pluralis, där alltså varje aktör har sitt dilemma, även om de är helt likartade.

Uttrycket "fångarnas dilemma" antyder däremot att de bägge fångarna har ett *gemensamt* dilemma, inte var sitt. Men det har de ju inte. Ur deras gemensamma synpunkt finns inget dilemma. För kollektivet är samarbete optimalt.

FILOSOFISOCIOLOGI

I mitten av nittonhundrafemtioalet hade sociologin helt nyligen blivit ett eget ämne i Sverige. Tidigare var det en del av den praktiska filosofin. På senare tid har filosofin tvärtom blivit ett studieobjekt för sin egen frigjorda avkomma. Det har uppstått något som kallas "filosofisociologi" (*sociology of philosophy*) – analogt med t.ex. familjesociologi och utbildningssociologi (och redan tidigare fanns förstås ett vidare område som hette "kunskapssociologi").

Kanske kan man också säga att det finns något som kunde kallas "sociologifilosofi" (*philosophy of sociology*), men det är i så fall detsamma som, eller en del av, samhällsvetenskapernas filosofi (*the philosophy of social science*). Filosofen Kathy Wilkes i Oxford brukade säga att det

finns en filosofi för varje vetenskap utom geografi, men antagligen har även geografin någonstans i världen gjorts till objekt för filosofiska undersökningar. Liksom varje verksamhet har sin sociologi. (Inte minst, får man anta, sociologisociologi, sociologisociologisociologi, och så vidare.)

I en nyutkommen bok, *Sverige och filosoferna: Svensk 1900-talsfilosofi i sociologisk belysning*, författad av Carl-Göran Heidegren, Henrik Lundberg och Klas Gustavsson (Studentlitteratur 2018), definieras filosofisociologi på följande sätt:

Filosofisociologi är studiet av filosofisk verksamhet som en socialt organiserad aktivitet förankrad i olika historiska och sociala kontexter, en aktivitet som inbegriper produktionen av filosofisk kunskap, det vill säga påståenden och argument som reser anspråk på giltighet. (s.10)

Detta skulle väl också, med en mer väletablerad term, kunna rubriceras som "idéhistoria", men filosofisociologi är förstås ett smalare område och det ska kanske också kännetecknas av en lite mer speciell metod och begreppsapparat, som är samhällsvetenskaplig snarare än humanistisk.

Filosofisociologins utgångspunkt är Marx uttalande "Det är inte människornas medvetande som bestämmer deras vara, utan tvärtom deras samhälleliga vara som bestämmer deras medvetande" (jfr s. 15). Syftet tycks alltså vara att förklara filosofernas påståenden och argument – vilket ju kan vara nog så knepigt – med hänvisning till deras materiella situation i det samhälle de tillhör. För en filosof är det närliggande att associera till Bertrand Russells program i boken *Västerlandets filosofi*, där han placerade filosoferna i sina sociala sammanhang, som produkter av sin miljö och tid. Men Russells bok kan knappast betraktas som en avhandling i sociologi. Man får väl snarare tänka sig att filosofihistoriska studier kan ordnas på en glidande skala från ett rent filosofiskt till ett rent sociologiskt angreppssätt.

Men filosofisociologer gör också psykologiska antaganden. Somliga, som Pierre Bourdieu, anser att filosofer strävar efter makt och inflytande över sina kolleger. Andra menar att filosoferna strävar efter att "göra teoretiska val som ligger i linje med deras intellektuella självuppfattning". Diskussionen om vem som har rätt anses ha stor betydelse för den framtida utvecklingen av filosofisociologin (jfr s. 21). Man kunde ju annars tänka sig att bägge typerna av motivation förekommer. Och kanske fler än dessa två.

FILOSOFISK KUNSKAP

Det sägs ibland att man aldrig kan nå fram till någon verklig kunskap i filosofi. Filosofi skulle med andra ord bara vara en massa tyckande. Tomt prat.

Mot detta pessimistiska påstående har man ibland invänt på ungefär följande sätt.

Låt oss betrakta påståendet att man inte kan veta något i filosofi. Detta påstående är ju självt ett filosofiskt påstående. Om man kan veta att det är sant, så är det alltså falskt (eftersom det just säger att man inte kan veta något i filosofi). Man kan alltså inte veta att det är sant. Alltså kan man veta något i filosofi, nämligen att man inte kan veta att man inte kan veta något i filosofi.

Är detta ett övertygande resonemang?

Inte riktigt. Ty om man kan veta att påståendet är sant, så är det ju sant (eftersom man inte gärna kan veta att ett givet påstående är sant, om det inte är sant).

Situationen är alltså osäker och lite förvirrande. Det verkar som att påståendet är både sant och falskt, om man kan veta att det är sant. Detta tyder på att man inte kan veta att det är sant (eftersom inget påstående är både sant och falskt). Alltså kan man veta något i filosofi, nämligen att man inte kan veta att man inte kan veta något i filosofi.

Vilket stämmer med vad vi nyss tycktes kunna visa.

Men resonemanget förutsätter att inget påstående är både sant och falskt. Även detta är ett filosofiskt påstående. Och om vi kan veta det, så är ju saken redan klar. Då kan vi veta något i filosofi.

Men kan vi verkligen veta att inget påstående är både sant och falskt? Den som menar att man inte kan veta något i filosofi, skulle antagligen förneka det.

VIKTIGA FILOSOFER

En enkät bland filosofer i början av 2000-talet om vilka som har varit de viktigaste filosoferna under de senaste tvåhundra åren gav följande resultat (källa: Brian Leiter): 1. Ludwig Wittgenstein; 2. Gottlob Frege; 3. Bertrand Russell; 4. John Stuart Mill; 5. W. V. O. Quine; 6. G. W. F. Hegel; 7. Saul Kripke; 8. Friedrich Nietzsche; 9. Karl Marx; 10. Sören Kierkegaard; 11. Rudolf Carnap; 12. John Rawls; 13. David K. Lewis; 14. G. E. Moore; 15. Donald Davidson; 16. Martin Heidegger; 17. Edmund Husserl; 18. Hilary Putnam;

19. William James; 20. Charles Sanders Peirce; 21. Alfred Tarski; 22. J. L. Austin; 23. P. F. Strawson; 24. Karl Popper; 25. G. E. M. Anscombe; 26. Jean-Paul Sartre; 27. John Dewey; 28. Wilfrid Sellars; 29. Arthur Schopenhauer; 30. Henry Sidgwick; 31. Alfred North Whitehead; 32. Michel Foucault; 33. Bernard Williams; 34. Gilbert Ryle; 35. Maurice Merleau-Ponty; 36. Franz Brentano; 37. Michael Dummett; 38. Jürgen Habermas; 39. Hannah Arendt; 40. Simone de Beauvoir.

NYA BÖCKER

Filosofiprofessorn Jacques Rancière diskuterar filmmediets gränser i boken *Filmens spelrum*, översättning av Alfred Matérn, Lindelöws bokförlag 2018.

Peter Asp, som är justitieråd i Högsta domstolen och tidigare professor i straffrätt, har utgivit *Truman & The Fat Man: En text om doktrinen om dubbel effekt*, Thales 2017.

Emma Beckman har doktorerat med avhandlingen *Mistaken Morality? An Essay on Moral Error Theory*, Umeå universitet 2018.

Den engelske idéhistorikern och journalisten Peter Watson har utkommit på svenska med en omfångsrik bok om *Den gudlösa tidsåldern*, Fri Tanke förlag 2018. Den handlar om hur sekulära idéer har utvecklats och präglat det senaste århundradets konst, litteratur vetenskap och filosofi.

Anders Hansson, som tidigare har utgivit en bok om Aristoteles etik, argumenterar i en ny bok, *Tre slags etiska teorier?*, Thales 2018, för att de moralläror som framförts av Aristoteles, Immanuel Kant och Henry Sidgwick egentligen har mycket mer gemensamt än man vanligen föreställer sig.

Medarbetarna i detta nummer av *Filosofisk tidskrift* tillhör alla den Filosofiska institutionen vid Stockholms universitet: Erik Angner är docent i praktisk filosofi, Krister Bykvist är professor i praktisk filosofi, Richard Dawid är professor i teoretisk filosofi, Kathrin Glüer och Peter Pagin är professorer i teoretisk filosofi, Anandi Hattiangadi är professor i filosofi, Mikael Janvid är docent i teoretisk filosofi och Jonas Olson är professor i praktisk filosofi.

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