

Hjerner i et kar - Hilary Putnam

noter af Mogens Lilleør, 1996

Historien om 'hjerner i et kar' tjener til: 1) at rejse det klassiske, skepticistiske problem om den ydre verden og 2) at diskutere forholdet mellem bevidsthed og verden.

Konklusionen er ifølge Putnam: Kan vi hvis vi er hjerner i et kar overhovedet sige eller tænke, at vi er hjerner i et kar? Svaret er Nej! Påstanden, at vi er hjerner i et kar, er selvgendrivende. Er den sand, kan den ikke hævdes uden at vise sig falsk. Man kan med andre ord ikke benægte den ydre omverdens eksistens uden at gendrive selv samme påstand.

Påstanden, at vi er hjerner i et kar, viser sig nemlig at have karakter af et paradoks, dvs. *en påstand hvis sandhed implicerer sin egen falskhed*, f.eks. 'alle generelle påstande er falske'. Hvis denne generelle påstand er sand er den falsk, ergo, den er falsk. En påstand kan også siges at være selvgendrivende, hvis det er selve dette, *at påstanden udtrykkes, der implicerer dens falskhed*, f.eks. 'jeg eksisterer ikke'. Påstanden er falsk, hvis jeg tænker den eller udtrykker den.

Putnam ønsker at vise, at påstanden 'vi er hjerner i et kar' har denne paradokse egenskab: *Hvis vi kan overveje, hvorvidt 'hjerner i et kar'- påstanden er sand eller falsk, så er den ikke sand. Ergo, den er ikke sand.* Eller én gang til: Selv om folk i 'kar-verdenen' kan tænke og sige ethvert ord, som vi kan tænke og sige, så kan de ikke referere til, hvad vi kan referere til, når vi tænker og siger de samme ord. De kan i særdeleshed ikke tænke eller sige, at de er hjerner i et kar, og mene det på samme måde, som vi kan mene det. Hvis jeg er solipsist, kan jeg ikke hævde, jeg er solipsist med samme mening, som hvis en ikke-solipsist hævdede det, for jeg vil kun kunne referere til min solipsist verden. Og det er ikke nok til at konstituere sandhed. Jeg ville derfor hævde noget, der ikke var sandt.

Der er tale om et argument mod **radikal skepticisme**, der tager udgangspunkt i, at sanseerfaring er den eneste kilde til erkendelse: Hvis vi er fanget af vore sanser, så er påstanden, at det eneste vi kan erkende er forestillingsbilleder baseret på sanseerfaring, falsk, fordi det eneste vi henviser til med påstanden er os og virkeligheden som et forestillingsbillede, og ikke os og virkeligheden givet uafhængigt af sanseerfaringer som kendsgerning.

Følgen er, at vi ikke ud fra sanseerfaring kan bestemme, hvad vi ikke kan erkende ved hjælp af sanseerfaring; det ville kræve at vi kunne 'træde ud af vort eget hovede', dvs. se tingene uafhængigt af vore sanser. Det kan vi ikke, og vi kan

derfor ikke hævde, at der ikke gives en ydre verden eller ikke findes andre bevidstheder. Det vil ligefrem være falsk at hævde det.

D) Repræsentation/reference

Hvordan er intentionalitet - reference - mulig?

En ting eller et billeds lighed er ikke tilstrækkelig til at konstituere repræsentation,- lighed er end ikke nødvendig. Normalt vil man kræve, at der skal være en **intention** om, at noget skal repræsentere noget andet. Men det forudsætter, at man er i stand til at tænke på det, der skal repræsenteres, dvs. at tankeformer kan repræsentere noget. Men hvordan kan de det? Hvilken forbindelse er der mellem tankeformer og fysiske ting? Ingen fysisk ting kan i kraft af sig selv repræsentere andre fysiske ting, men tankeformer kan tilsyneladende ikraft af deres særlige ikke-fysiske struktur: **intentionaliteten**. Men hvordan er intentionalitet (reference) mulig?

To svar: *a) Den magiske reference-teori*: Navne har en nødvendig forbindelse med deres bærere; at kende en tings sande navn giver en magt over tingen. *b) Den konventionelle teori om reference*: Navne har kun en kontekst-afhængig, konventionel og kontingent forbindelse med deres bærere (Putnams eget synspunkt). Hvad der gælder for fysiske billeder gælder også for mentale billeder eller repræsentationer såsom navne: Hverken fysiske- eller mentale billeder har en nødvendig forbindelse til det, de repræsenterer. At hævde det modsatte er at forfægte en magisk reference-teori. Reference er konventionel i den forstand, at den udspringer af praksis, dvs. af vores ikke-verbale omgang med tingene. Man kan ikke referere til bestemte slags ting, hvis man slet ikke har nogen kausal-omgang med dem - direkte eller indirekte.

Transcendental analyse

Dette er en undersøgelse af forudsætningerne for reference og dermed for tanken, forudsætninger som er bygget ind i bevidstheden, men som ikke er uafhængige af erfaringen. Samtidig er det en diskussion af det klassiske, skepticistiske problem om den ydre verden. Metoden afhænger af empirisk-farvede antagelser og er således ikke a priori. Undersøgelsen bygger på to præmisser: 1) Magiske referenceteorier er falske. 2) Man kan ikke referere til bestemte slags ting, hvis man slet ikke har nogen kausal-omgang med dem.

a) Begreber er evner eller færdigheder

Mentale fænomener består af ord, billeder, sansninger og følelser. Men at besidde et begreb er ikke et mentalt fænomen som f.eks. dette at besidde et billede: Billeder er ikke efterfulgt af nogen evne til at handle på en bestemt måde. Man kan have et hvilket som helst system af billeder uden at have evnen til at bruge sætninger på den i situationen korrekte måde. Man kan have et billede uden at

vide hvad man skal stille op med det. Billeder er blot billeder. **Begreber** er slet ikke mentale fænomener; begreber er **tegn brugt på en bestemt måde**. Tegnene kan være offentlige, private, mentale eller fysiske. Men tegnene (f.eks. navne) taget for sig er ikke noget begreb; tegn har ikke intrinsisk reference, de refererer ikke i sig selv. Forskellig reference afhænger af **konteksten**, dvs. det der bestemmer reference er socialt, og ikke noget individuelt som f.eks. psykologiske modi. Meninger er ikke i hovedet. Det, der udgør kriteriet på om man besidder et begreb, er om man kan anvende de relevante sætninger på en måde som er korrekt i situationen. Med andre ord: **Færdighed og begreb korrelerer**.

b) Begreb/færdighed og forståelse hænger nøje sammen

Det er ikke bevidsthedsfænomener, der konstituerer **forståelse**, men snarere tænkerens evne til at anvende fænomenerne, dvs. til at fremkalde de rette fænomener under de rette og relevante omstændigheder. Forståelsen - ens egen forståelse af ens egne tanker - er ikke et mentalt fænomen eller -begivenhed, men en **evne eller færdighed**. Intet sæt af mentale begivenheder er nødvendige for forståelse. Mentale billeder kan være tilstede i en mand, som ikke har nogen begreber (som intet forstår), og være fraværende i en mand, som forstår en masse, dvs. som ikke tænker i billeder, men handler (i begreber). Mentale genstande såsom ord, billeder og følelser har ingen nødvendig, intrinsisk, reference til virkelige ting. Og at postulere specielle mentale størrelser med en indbygget reference til ydre ting er at begå en logisk fejl (kategorifejl). Det er nemlig at forveksle evner med begivenheder, aktivitet med passivitet. Det er det den magiske referenceteori gør.

c) Forståelse og billede hænger ikke sammen

Det er både logisk og fysisk muligt at forestille sig en situation (som ganske vist er ret usandsynlig), hvor en person **tænker ord**, som faktisk beskriver træer, i et sprog, og samtidig har de **mentale billeder**, der svarer til ordene, men som **hverken forstår ordene eller ved, hvad et træ er**. Alt det, som passerer forbi bevidstheden, kan være kvalitativt identisk med, hvad der passerer gennem en talende dansker, der virkeligt forstår sig på træer, uden at noget af det refererer til træer. Dette illustrerer den vigtige begrebslige sandhed, at ingen repræsentations-systemer har intrinsisk forbindelse med det de repræsenterer. Der findes ingen magisk forbindelse mellem billede og ting uafhængig af talerens og tænkerens dispositioner. Tankeformer repræsenterer ikke i sig selv det, de er om. Man kan have tanke og mentale billeder uden forståelse, og man kan have forståelse uden tanke og mentale billeder.

Lad mig prøve at uddybe dette: Hvordan skulle man f.eks. afgøre om 'nn' er en robot eller et menneske? Man kunne gennemføre en test med henblik på at fastslå **fælles reference**. Altså undersøge om 'nn's ord refererer til det samme som vores

ord gør. Testen kunne bestå i at gennemføre en konversation mellem en 'cleared' kontrolperson, der refererer som vi gør, og 'nn'. Hvis vi ikke kan skelne mellem kontrolpersonen og 'nn', har sidstnævnte bestået testen: 'nn' er så ikke en robot. Men desværre er testen ikke definitiv, fordi det ikke er logisk umuligt, at en eller anden kunne passere testen uden at referere til noget som helst.

En maskine som er konstrueret til at tale pænt om alt, hvad vi kan tale om, kan ikke siges at referere, fordi den vil ikke være i stand til at genkende et æbletræ eller et æble, et bjerg eller en ko, en mark eller et tårn, hvis den var foran noget sådant. Maskinen er blot et instrument til at producere sætninger, som svar på sætninger, men ingen af sætningerne er nødvendigt forbundet med den ydre verden. Maskinerne refererer ikke mere til den ydre verden end den 'tegning' af Winston Churchill sporet af en myres tilfældige vej laver i sandet. To maskiner koblet sammen ville 'samtale i al evighed', selv om den øvrige verden forsvandt. Maskinen er ufølsom over for ændringer i den kontekst, der udgør grundlaget for dens programmer. Derfor refererer den ikke, selv om dens programmører gør. Det følger heraf, at ord (sprogeregler og programmer for sprogbrug) ikke har nogen nødvendig reference til det, som ordene indeholder eller er om.

Det er vores **konvention om reference**, der tilskrives maskinen og myren. Men vores tale om æbler og marker etc. er snævert forbundet med vores ikke-verbale omgang med æbler og marker. Der er konventionelle adgangs-sprogeregler, der tillader os at gå fra erfaringer med æbler til udsagn om æbler, og der er konventionelle udgangs-sprogeregler, der tillader os at gå fra udsagn om æbler til handlinger, som ikke er talehandlinger. Både maskinen og myren mangler disse sprogeregler og er derfor uden reference. Maskinens sprog er blot **syntaks**, dvs. sætninger som ligner rigtig refererende sætninger og diskussioner, men som er **uden semantisk indhold**. Den nævnte test for reference er ikke tilstrækkelig, fordi den ikke logisk kan udelukke, at en maskine der er lavet specifikt til at bestå testen også vil bestå den.

II) Hjerner i et kar

Det er dette, at ord ikke har nogen intrinsisk, nødvendig, reference til det som ordene indeholder eller er om, der er afgørende for konklusionen: at hjerner i et kar ikke kan referere til noget som helst ydre, og dermed ikke kan sige, at de er hjerner i et kar.

Hjerner i et kar refererer ikke til ydre genstande, f.eks. eksisterende træer. Ordet træ bruges af hjernerne på den måde, disse hjerner nu gør, de tænker på den måde de gør, har de billeder de har, uanset om der er træer eller ej. Deres 'billeder' er kvalitativt identiske med billeder, ord etc. som repræsenterer virkelige træer, men dette er ikke nok. Når hjernerne tænker 'der er et træ foran mig', så tænker de ikke på noget virkeligt træ, men på et 'kar-træ'. Det faktum, at deres sprog er forbundet med et program med sanseinputs, som ikke i sig selv eller ved noget

ydre repræsenterer træer (eller noget som helst andet ydre), kan ikke bevirke, at hele 'repræsentations'-systemet - sproget i anvendelse - faktisk refererer til eller repræsenterer træer eller noget ydre overhovedet. Hele systemet af sansedata har ikke mere forbindelse med træer end myrens 'tegning' af Churchill har med Churchill. Kvalitativ identitet mellem de tanker hjerner i et kar gør sig, og dem mennesker i den virkelige verden gør sig, implicerer ikke samme reference. Der mangler **ikke-verbal omgang** med tingene. Det samme gælder rigtige maskiner: Selvom maskinerne var bevidste, ville det ikke være tilstrækkeligt til at sikre, at de refererer til det, vi refererer til.

Konklusionen

Kan vi, hvis vi er hjerner i et kar, overhovedet sige eller tænke, at vi er hjerner i et kar? Nej! Påstanden, at vi er hjerner i et kar, er selvgendrivende. Er den sand, kan den ikke hævdes uden at vise sig falsk! Selv om folk i 'kar verdenen' kan tænke og sige ethvert ord, som vi kan tænke og sige, så kan de ikke referere til, hvad vi kan referere til, når vi tænker og siger de samme ord. De kan i særdeleshed ikke tænke eller sige, at de er hjerner i et kar, og mene det på samme måde, som vi kan det. Heraf følger, at hvis vi virkelig er hjerner i et kar, så ville påstanden at vi er hjerner i et kar henvise til 'mentale hjerner i et kar' eller 'billede af hjerner i et kar' (den intentionale genstand eller forestillingen). Men en del af hypotesen, at vi er hjerner i et kar, er at vi ikke er 'mentale hjerner i et kar', men faktiske hjerner i kar. Så hvis vi faktisk er hjerner i et kar, så er sætningen 'vi er hjerner i et kar' falsk (den refererer kun til det mentale billede). Hvis vi er hjerner i et kar må vi nødvendigvis sige noget, der er forkert, nemlig at vi er et mentalt billede af hjerner i et kar, for dette er det eneste vi kan henvise til. Men det er selvfølgelig falsk givet præmissen, at vi faktisk er hjerner i et kar. Hvis det altså er sandt at vi er hjerner i et kar (og altså ikke et billede), så er det (selvfølgelig) falsk at hævde vi blot er et billede. Hvis vi er hjerner i et kar, så er påstanden, at vi er hjerner i et kar falsk i og med at den (qua dens mangelfulde referenceevne) hævder, at vi er et billede af hjerner i et kar.

Hvis jeg er en hjerne i et kar og hævder, jeg er en hjerne i et kar, så henviser jeg ikke til noget udenfor karret, fordi jeg som hjerne i et kar ikke kan have kausal omgang med tingene, jeg har blot billeder. Jeg kan derfor ikke henvise til kendsgerningen, at jeg er hjerne i et kar, men kun til billedet af 'hjerne i et kar', hvilket svarer til, at når jeg siger kar, så betyder det 'kar i kar'.

Eftertanke

Der er her tidligere nævnt tale om et argument mod **radikal skepticisme**, der tager udgangspunkt i, at sanseerfaring er den eneste kilde til erkendelse: Hvis vi er fanget af vore sanser, så er påstanden, at det eneste vi kan erkende er forestillingsbilleder baseret på sanseerfaring, falsk, fordi det eneste, vi henviser til

med påstanden, er os og virkeligheden som et forestillingsbillede, og ikke os og virkeligheden givet uafhængigt af sanseerfaringer som kendsgerning. Følgen er, at vi ikke udfra sanseerfaring kan hævde, at der ikke gives en ydre verden eller ikke findes andre bevidstheder. Det ville kræve at vi kunne 'træde ud af vort eget hovede', dvs. se tingene uafhængigt af vore sanser. Det kan vi ikke.