

ER BEVIDSTHEDEN ET FINIT SYSTEM?¹

Niels Ole Finnemann

Diskussionen af forholdet mellem det fysiske, biologiske og psykiske har i tidens løb motiveret en lang række paradigma-tiske fornyelser i videnskabelig teori og filosofisk tænkning. I nyere tid har spørgsmålet især været taget op fra naturvidenskabelig side. Artiklen vil søge at indkredse spørgsmålets betydning og diskutere de grundlæggende antagelser i nogle af disse videnskaber – med særlig vægt på udviklingen af ideen om den finite bevidsthed i den mekaniske teoriehistorie i det 20. århundrede.

1. Indledning

Det er velkendt, at vi i forståelsen af det univers, vi befinder os i, har et særligt prekært problem med forståelsen af vores egen evne til at forstå, eller med en lidt bredere term: forståelsen af den menneskelige bevidsthed, idet denne på en gang optræder som subjekt og objekt – som beskrivelsens aksiomatiske forudsætning og som dens genstand.

Som følge af dette sammenfald har den bevidsthedsteoretiske overvejelse et særligt problem med at etablere og beskrive afstanden til sit objekt, men da denne vanskelighed angår forståelsen af bevidstheden, går spøgelse igen i enhver beskrivelse af ethvert objekt, da også disse beskrivelser har denne bevidsthed som forudsætning.

Vi har i moderne videnskab mindst tre hovedtraditioner for fortolkningen af dette problem.

For det første en naturvidenskabelig tradition for beskrivelse af naturen gennem bestemmelse af fænomenernes form og udstrækning og af de invariante matematiske relationer mellem hver især variable størrelser.

For det andet en modsvarende tradition for beskrivelse af bevidstheden som bærer af et fornuftsmæssigt rationelt subjekt.

Op for det tredje en tradition for beskrivelse af bevidstheden gennem introspektion, som vi eksempelvis kender det fra dele af psykologien og ikke mindst psykoanalysen.

Niels Ole Finnemann er Dr. Phil. og arbejder som ekstern lektor på Institut for informations- og medievidenskab i Århus.

Artiklen er til dels baseret på et foredrag holdt i Selskab for Filosofi og Psykologi, Københavns Universitet, 6.10 1994.

De tre anskuelser er ikke altid skarpt afgrænset. Det er f. eks. muligt at beskrive den cartesianske fornuftsopfattelse som en introspektiv opfattelse, idet den fornuftsmæssige erkendelse af sandheden i sidste ende beror på en intuitivt konciperet indsigt i, at noget ikke kan være anderledes. Og det er omvendt muligt at beskrive den freudianske bevidsthedsteori som en deskriptiv teori om fænomener – og ekspliciterede – objekters indbyrdes relationer. Der er i begge tilfælde, som også hos Kant, tale om bevidsthedsobjekter. Men hvor disse objekter hos Descartes siger noget om verdens indretning, som den er givet af Gud, hos Kant om verdens indretning, som den må fremtræde for fornuften, siger de hos Freud noget om psykens indretning. Det introspektive foreligger her i 2. grad – betragtning af bevidsthedens form gennem betragtning af ekspliciterede indholdsmanifestationer.

Selv om jeg i det følgende hovedsagelig vil beskæftige mig med den første og anden betragtning, vil jeg starte med at pege på et par af de problemstillinger, der knytter sig til den tredje.

Det grundlæggende problem for denne betragtning består, som flere har påpeget, i vores gensidige vanskelighed ved at se ind i hinandens indre. Selv om der givetvis er mange muligheder for at genkende andres selvbeskrivelse, rummer enhver selvbeskrivelse (og andres eventuelle genkendelse og anerkendelse) et uundgåeligt element af individuel tolkning, der er utilgængelig for andres kontrol, og som måske lige så meget repræsenterer den enkeltes personlige overbevisning, eller periodens tids- eller kulturbestemte fordomme.

Den introspektive betragtnings dilemma kan måske anskueliggøres i det forhold, at dens sandhed i psykoanalysen for eksempel beror på patientens accept, idet denne er forudsætning for, at den diagnostiske analyses forskrifter for helbredelse kan blive testet.

Når det er sagt, må man samtidig medgive, at den introspektive betragtning ikke desto mindre har et præg, nemlig, at den giver adgang til bevidsthedens indhold, hvor alle objektive og ydre betragtninger alene kan angå bevidsthedens mulige udtryksformer.

Der kan endda være tale om endnu et præg, idet den introspektive betragtning måske også er bedre egnet som kilde til belysning af bevidsthedens individualitet, altså af bevidsthedstræk, der ikke har nogen yderligere generaliserbarhed, træk, der ikke blot er eksemplarer af det typiske.

Min hensigt med disse indledende betragtninger er ikke at forfølge disse spørgsmål, men blot at markere et kontrapunkt til de bevidsthedsbeskrivelser, der tager udgangspunkt i en objektiv betragtning af bevidsthedens manifestationsformer – idet jeg vil koncentrere mig om spørgsmålet om bevidsthedens lokalisering og udstrækning.

2. Bevidstheden som genstand – Bevidsthedens afgrænsning

I den introspektive bevidsthedsbetragtning melder spørgsmålet om bevidsthedens udstrækning sig først på et meget sent trin i betragtningen, hvis det overhovedet melder sig. Bevidstheden fremtræder her som et uendeligt hav, der tilsyneladende er uden endelige grænser.

Denne forestilling om den ubegrænsede bevidsthed er ikke forbeholdt den introspektive betragtning, den svarer tværtimod ganske godt til den cartesianske subjektfilosofi, udmøntet i begrebet *res cogitans*, den tænkende substans, der modstilles den udstrakte verdens substans, *res extensa*. Forestillingen stammer efter alt at dømme også fra de samme filosofiske og religiøse kilder, idet vi både genfinder dette element i forståelsen af bevidstheden i Platons idélære, i den kristne lære og i andre religioners lære. Bevidstheden fremstår her som det i forhold til den endelige, udstrakte og/eller stofflige verden forskellige, hvorigennem individerne kan få del i det universelt guddommelige og ubegrænsede.

Når vi alligevel kan skelne mellem flere forskellige opfattelser af denne uendelige bevidsthed, beror det ikke mindst på, at den uendelige bevidstheds forankring i verden bliver fortolket på forskellig måde. Jeg skal her blot nævne 3 af de vigtigste:

For det første: reinkarnationslæren, hvor bevidstheden er udødelig og fortsætter i stadig nye materialisationer (herunder også i andre levende organismer) indtil det eventuelt måtte lykkes at nå den definitive immaterielle og tidløse sluttilstand.

For det andet: den kristne lære om det evige liv, hvor bevidstheden efter døden eksisterer i al evighed i en immaterialiseret form uden for tid og rum. I hvert fald i det lykkelige tilfælde. Her har bevidstheden således én og kun en inkarnation, og denne kan kun være menneskelig. Bevidstheden inkarnerer sig ikke i andre dyr endsige planter.

For det tredje: den moderne lære om jeg'ets tilblivelse og død som en med organismens tilblivelse og død forbundet og samtidig begivenhed.

Disse opfattelser lever som bekendt i mere eller mindre skøn samdrægtighed og i mange blandingsformer i vores egen kultur i dag.

Det er klart, at den sidste af de tre opfattelser adskiller sig fra de to andre ved at knytte bevidsthedens eksistens direkte sammen med organismens eksistens, og at den biologiske død dermed bliver udpeget som en grænsebetingelse for bevidsthed. Det springende punkt er her skiftet fra at forstå døden som grænsebetingelse for bevidsthedens inkarnation til at forstå døden som grænsebetingelse for bevidstheden – eller i det mindste som grænsebetingelse for de bevidsthedsfænomener, vi kan udtale os meningsfuldt om.

Det er her værd at notere sig, at der til denne sidste opfattelse hører en asymmetri i forståelsen af forholdet mellem det fysiske og biologiske, versus forholdet mellem det biologiske og det bevidste (eller psykiske),

idet dødsvilkåret indebærer, at den levende organisme regredierer til sine uorganiske fysiske bestanddele, der altså fortsat eksisterer, hvorimod bevidstheden forsvinder med livets ophør.

Men den moderne opfattelse adskiller sig også på et andet punkt, for hvor der til de to første opfattelser hører *et begreb om sjælen med reference til det guddommelige*, hører der til den sidste *et begreb om subjektiv identitet og selvreference*, der i et vist omfang i hvert fald er forbundet med det, vi med et andet vanskeligt begreb kan kalde subjektets mulighed for selvbestemmelse.

Bindingen til den biologiske organisme indebærer omvendt en ny grænsebetingelse for denne selvbestemmelse, idet bevidstheden ikke blot bliver underlagt organismens død, men også – på en eller anden måde, der kan tolkes forskelligt – bundet til organismens virkemåde.

Endelig indebærer lokaliseringen af bevidstheden i organismen også en udskillelse og isolering af bevidstheden fra andre bevidstheder.¹ En sådan udskillelse og isolering optræder derimod ikke som særskilt tema i de ikke-moderne forestillinger – idet sjælens biologiske isolation i organismen blot er en inkarnation eller instantiering af et fænomen, der er forankret i et transcendentalt og universelt åndeligt enhedsprincip.

Hvor den moderne forestilling om bevidstheden *som et til organismen bundet, lokaliserbart fænomen* således tager udgangspunkt i en afgrænsning af bevidstheden med den biologiske død som grænselinje, opgiver den samtidig en anden grænselinje, nemlig mellem det psykiske og biologiske, eller mere præcist: denne grænse, der tidligere blev trukket som en grænse mellem det transcendentale og det immanente, melder sig nu som en uklart forstået grænse i det immanente. Jfr. eksempelvis vores vanskeligheder med at komme til begreb om psykosomatiske relationer.

Vi kan forestille os biologiske fænomener, der ikke har psykiske egenskaber (evne til at skabe symboler, selvreference, emotionel og intellektuel abstraktionsevne mv) men ikke psykiske fænomener, der ikke er biologisk forankret, ganske som vi kan forestille os fysiske fænomener, der ikke har biologiske egenskaber (som dødelighed, drift og seksualitet), men ikke biologiske fænomener, der ikke også er fysisk realiserede.

Denne forestillingskreds synes at indebære et billede af verden som en pyramide med en bred fysisk basis, der udfolder sig i den universelle kosmologiske tidsdimension (endelig eller uendelig) ovenpå hvilken, der befinder sig en biologisk sfære, der udfolder sig i en mere afgrænset (og i hvert fald endelig) periode, hvorovenpå der endelig eksisterer en psykisk og symbolsk (kommunikativ og samfundsmæssig) sfære, der først tager sin begyndelse på et sent trin i den biologiske tidssfære, og derfor endnu smallere, men som dog ikke nødvendigvis afsluttes før denne, men senest med denne.

Mit formål med dette billede er at give et anskueligt indtryk af det forhold, at et moderne billede af bevidstheden både må omfatte en redegørelse for bevidstheden eller psyken som særskilt fænomen, og for dette

fænomens relationer til de underliggende biologiske og fysiske fænomener, idet vi således ikke blot kan tage det ene eller andet af disse momenter som en givet forudsætning for en isoleret betragtning.²

Modellen demonstrerer samtidig, at forholdet mellem det fysiske, biologiske og psykiske i den moderne forståelse både opfattes som et naturhistorisk udviklingsforhold og som et forhold mellem samtidigt sameksisterende niveauer.

Jeg har i det hidtil sagte afbrudt historien i slutningen af forrige og begyndelsen af dette århundrede, og vil nu følge et bestemt bevidsthedsteoretisk spor i det 20. århundrede, nemlig det spor, der går fra den termodynamiske fysik fra sidste del af forrige århundrede frem til de neo-cartesianske eller informationsteoretiske bevidsthedsteorier, der ligger bag begreberne kunstig intelligens og Cognitive Science.

Det er ikke min hensigt at gennemgå denne udvikling i sin helhed, men derimod at forfølge spørgsmålet om den forståelse af bevidsthedens grænser, der er indeholdt i disse teorier.

Som afslutning på denne gennemgang vil jeg vende tilbage til spørgsmålet, hvorledes vi muligvis kan tænke eksistensen af bevidstheden som et fænomen, der på en gang udgør et selvstændigt univers eller system, der samtidig er nedsænket i, betinget af og måske betingende for processer i de underliggende biologiske og fysiske niveauer. Der vil her blive argumenteret for det synspunkt, at forholdet mellem et biologisk og bevidst system kan forstås som en grænseflade, hvori det neurofysiologiske system optræder som et redundanssystem for bevidstheden, hvilket indebærer, at det bevidste systems forhold til de neurofysiologiske systems processer kan omfatte:

- at den samme neurofysiologiske proces kan udnyttes med skiftende betydning.
- at samme betydning kan processeres ad forskellige neurofysiologiske veje/gennem forskellige neurofysiologiske processer.
- at bevidste processer kan fremkalde nye neurofysiologiske processer og give allerede eksisterende processer nye virkninger og betydninger, herunder svække modificere eller suspendere deres betydning.
- at neurofysiologiske processer til stadighed kan fremkalde (nye) virkninger i det bevidste system.

3. Fra termodynamik til mekanisk bevidsthedsteori

Når jeg slår ned på den termodynamiske fysik, skyldes det ikke, at man der kan finde begyndelsen til mekaniske forklaringer af det biologiske og psykiske. Sådanne forklaringer indgår som bekendt også i adskillige ældre teorier, der ikke har stillet sig tilfreds med den cartesianske dualisme.

Min begrundelse er derimod, at man med en vis rimelighed kan sige, at den termodynamiske teori både direkte og indirekte giver anledning til udviklingen af nye forestillinger og teorier om forholdet mellem det fysiske, biologiske og psykiske – sammenlignet med de forestillinger, der grunder sig direkte på den klassiske newton'ske mekanik.

Det gør den

- For det første fordi termodynamikken i sig selv repræsenterer en teoretisk fornyelse i den mekaniske fysiks naturbillede – og dermed i forståelsen af begrebet om det fysisk-mekaniske. Til disse fornyelser hører forestillingerne om tidens irreversibilitet og den termodynamiske varmedød, men også en knapt så ofte fremhævet overgang fra den newton'ske beskrivelse af naturen som en stor maskine til en beskrivelse af naturen som en mængde af mange, små maskiner udtrykt i beskrivelsen af afgrænsede, lukkede systemer.
- For det andet fordi en række af de nye elementer også får en mere generel paradigmatiske betydning udenfor fysikken. Det gælder bl.a. begrebet om det lokale, afgrænsede finitte rum, etableringen af den matematisk-statistiske beskrivelse, det være sig af fysiske eller ikke-fysiske processer – eller hvad vi kan kalde opkomsten af moderne funktionalistiske beskrivelsesprincipper (til dels modsvarende strukturalismen i lingvistikken) – hvilket altsammen igen, negativt sagt, også indebærer et brud med den naturalistiske tolkning af den klassiske mekanik.
- For det tredje fordi den termodynamiske fysik indebærer en forskydning af iagttagelsesproblematikken, knyttet til diskrepansen mellem den makrofysiske orden og den mikrofysiske (molekylære) uorden. I kølvandet på denne forskydning udvikles sidenhen ideen om at inddrage iagttagelsesprocessen i beskrivelsen af det iagttagne og ideen om at betragte iagttagelsesprocessen som fysisk realiseret proces, hvor informationen om et system indvindes gennem – eller indeholdes i – den energi systemet afgiver til iagttageren.

Jeg skal ikke her gå nærmere ind på disse temaer i den termodynamiske fysik.

En del er behandlet i Witt Hansen (1985) og jeg har selv prøvet – fra en anden synsvinkel – at beskrive, hvordan disse temaer behandles af den tyske fysiker Ludwig Boltzmann, der regnes for at være ophavsmanden til den første udgave af den moderne statistiske og matematiske termodynamik.³

I det følgende vil jeg derimod komme ind på nogle af de disse temaer i forbindelse med den betydning, de har haft for senere forsøg på at tolke forholdet mellem det fysiske, biologiske og psykiske på mekanisk grund-

lag, og jeg vil her især hæfte mig ved to momenter, nemlig 1) bidraget til udviklingen af begrebet om et finitte rum og 2) beskrivelsen af iagttagelsesprocessen som en tids- og rumligt udfoldet, fysisk relation mellem objektet og iagttageren.

Disse to momenter udgør to af de centrale forudsætninger for en betydelig del af den erkendelsesteoretiske udvikling i det 20. århundrede, idet de foranlediger en række forsøg på at etablere et erkendelsesteoretisk grundlag, der kan træde i stedet for den klassiske mekaniks subjekt-objekt-model.

Den erkendelsesteoretiske krise i termodynamikken spiller eksempelvis en direkte og eksplicit udtalt rolle for Charles Peirce's arbejde med at udvikle en almen tegnteor. Og hvis den ikke spiller en direkte rolle, så danner den i hvert fald kontekstuel baggrund for G. Freges tilsvarende forsøg på at formulere en generel logisk symbolteori.

Der er i begge disse tilfælde tale om forsøg på at reformulere erkendelsesteorien på symbolteoretiske præmisser, fordi – kunne man tilføje – den symbolske reference til det ikke-symboliske ikke længere kunne tages som et sikkert erkendelsesfundament.

Hvor billedet af den fysiske natur bogstaveligt talt smuldrer under den analytiske findelings ætsende virkninger, søger videnskaberne des stærkere mod en fundering af erkendelsen ikke i metafysikken, men i metasymbolikken, i formelle teorier om formelle beskrivelser. (Frege, Russel-Whitehead, Hilbert, Gödel, Turing, hhv. Wittgensteins billedteori om verdens struktur, der »viser sig« i sætningens).

Set i dette perspektiv kan termodynamikkens begreb om det mere eller mindre afgrænsede finitte beskrivelsesrum betragtes som et første trin på vejen i en teoretisk abstraktionsprocedure, der fører frem til forestillingen om det formelle, aksiomatisk afgrænsede og – synes det – fuldt beskrivelige logiske rum.

Når jeg taler om det *mere eller mindre afgrænsede* termodynamiske rum, skyldes det, at termodynamikken, der etablerer begrebet om det finitte fysiske rum, paradoksalt nok også demonstrerer, at denne afgrænsning ikke lader sig fuldføre; dels fordi det teoretisk set er umuligt at operere med begrebet om et finit, fysisk rum, idet man ikke fuldstændigt kan udsondre den kosmiske baggrundsstråling; dels fordi observationen forudsætter, at den overførte information overføres i en fysisk medieret proces, hvori systemet afgiver energi til iagttagelsesmidlet, der i sidste ende altid inkluderer det menneskelige sansesapparat og hjernens aktivitet. Der er altså ikke blot tale om et fysisk forstået rum, hvori der foregår fysiske processer, men om et rum, hvorfra der i det mindste må være en forbindelse til noget udenfor dette rum, nemlig til den uden for dette rum værende iagttager.

Den termodynamiske teori demonstrerer derigennem i en objektiviseret form iagttagelsesproblemets kompleksitet. Behovet for at inddrage iagttageren i beskrivelsen af iagttagelsen anbringer uvægerligt det erken-

dende subjekt i dobbeltrollen af både at være subjekt og objekt på en og samme tid.

En sådan forbindelse spores også i Peirce's tegnteori, der netop forankres i relationen mellem det iagttagne objekt, den iagttagelige repræsentation for dette objekt og den interpretantfunktion, der definerer relationen mellem objektet og dets iagttagelige repræsentation.

En dertil svarende interpretantfunktion synes derimod helt fraværende i Frege's, Russel's og senere Hilbert's forskellige formelle teorier om formelle systemer. Men det er iøvrigt ikke afgørende i denne sammenhæng, hvor det kun drejer sig om at registrere en udviklingslinje, der fører fra fysikkens iagttagelsesproblem over formelle symbolteorier frem til de senere formuleringer af formelle bevidsthedsteorier.

Det er – i betragtning af det indre spændingsforhold mellem begrebet om det finite rum og iagttagelsesproblemet – ikke overraskende, at der dannes – mindst – to forskellige typer af løsningsstrategier. På den ene side strategier og teorier, der som de førnævnte, tager udgangspunkt i begrebet om det finite beskrivelsesrum, på den anden side strategier og teorier, der tager udgangspunkt i beskrivelsen af iagttagelsesprocessen som en fysisk proces.

Jeg skal ikke her forsøge at give noget samlet billede af disse to strategier og de teoridannelser, der er frembragt heraf, men derimod koncentrere mig om et ganske bestemt møde mellem disse to forskellige spor i den mekaniske teorihistorie, nemlig mødet mellem

1) Alan Turing's teori om den universelle computer (der har sine rødder i det spor, der går over den matematisk-logiske teoriudvikling, knyttet til forsøget på at udvikle formelle beskrivelser af formelle processer) og

2) Claude Shannon's matematiske teori om kommunikation, der har sine rødder i det spor, der udgår fra det termodynamiske begreb om ubestemthed og entropi, idet Shannon bl.a. definerede sit informationsbegreb med den samme matematiske formel, som Boltzmann brugte til at definere entropien.

Når jeg siger et bestemt møde mellem disse teorier, beror det på, at det er svært at dokumentere, hvornår og hvordan et sådant møde første gang fandt sted. Mens vi har et nogenlunde klart billede af den teorihistorie, der er forudsætningen for Turing's teori, tegner der sig et mere broget billede, når det gælder de teorihistoriske forudsætninger for Shannon's. Selv refererer han primært tilbage dels til Boltzmann, dels til Leo Szilard's beskrivelse af informationsoverførslen fra et system til en iagttagelse som en fysisk energiprocess, dels til nogle mere ingeniørmæssigt motiverede overvejelser over fysiske definitioner af symbolske udtryksstørrelser, men han har også selv i 30-erne beskæftiget sig med muligheden for at give en logisk beskrivelse af elektriske kredsløb.

Der er imidlertid intet i Shannon's teori, der peger direkte tilbage til Turing's teori fra 1936, men derimod en hel del, der peger på kendskab

til den mellemliggende – især amerikanske – udvikling af computerteorien. Men også her er det mere end uklart hvor meget eller lidt, man kendte til Turing's teori. Det er kun dokumenteret, at John von Neumann havde stiftet bekendtskab med teorien sidst i 30-erne eller begyndelsen af 40-erne.

Selv om Turing's teori utvivlsomt har haft en eller anden direkte betydning for von Neumann, er det vanskeligt at sige hvilken bestemt betydning og endnu vanskeligere at sige, hvordan denne betydning eventuelt er videreformidlet i den daværende amerikanske computerteori, der var meget mere direkte orienteret mod principperne for løsningen af de praktiske konstruktionsproblemer.

Mens der således er nogen uklarhed om den konkrete teorihistories forløb, tegner der sig ikke desto mindre to linjer, der begge refererer til den fælles grundforestilling om det finite rum, der er defineret ved et endeligt antal distinkte tilstande. I det ene spor optræder denne forestilling som et matematisk-logisk rum, med et finit sæt af operative skridt fra problem til løsning, mens den i det andet optræder som et fysisk realiseret rum.

Hvor Turing imidlertid arbejder med transformationen af den finite, formelle procedure til en fysisk-mekanisk udførbar procedure (reduktion af logik til skridtvis mekanik), tager bevægelsen hos Shannon den modsatte retning, idet han forsøger at bruge logiske beskrivelser i konstruktionen af effektive fysiske kredsløb (fra fysisk mekanik til symbolsk logik).

Vi skal imidlertid nok helt frem til 1956, før vi med sikkerhed kan datere et møde mellem disse teoridannelser, i det mindste hvis vi vil tale om en teoretisk bearbejdning, der står i gæld til begge sider. Det møde, jeg her taler om, lader sig i hvert fald datere med sikkerhed, og det lader sig også identificere som en teoretisk innovation af den mekaniske teori, idet vi her finder den første beskrivelse af det, vi i dag omtaler med begrebet Kunstig Intelligens.

Mens termen Kunstig Intelligens ved den lejlighed blev formuleret af John McCarthy, blev termens indhold formuleret af Herbert A. Simon, Allen Newell og Cliff Shaw, der selv talte om en beskrivelse af principperne for konstruktionen af en General Problem Solver – en generel problemløser.

En sådan indretning skulle ikke blot kunne udføre en hvilken som helst finit mekanisk procedure – som den universelle computer – den skulle i princippet kunne løse problemer af enhver art og kompleksitet, svarende til de problemer, den menneskelige fornuft kan møde i virkeligheden. Og det skulle den kunne gøre gennem mekanisk udførlige procedurer.

Til forskel fra de teorier, vi idag finder i Cognitive Science, sigtede Simon, Newell og Shaw alene på at formulere en teori om menneskelig intelligens, hvorved de forstod evnen til at finde løsninger på komplekse problemer, herunder ikke mindst evnen til at finde sådanne løsninger i

»nye situationer«. Hvorvidt dette efter deres opfattelse var identisk med begrebet intelligens som sådan, eller blot en bestemt form for intelligens, berøres ikke eksplicit, men det fremgår implicit, at det i hvert fald måtte forstås som en meget vigtig, hvis ikke den vigtigste, form for intelligens. Når teorien alligevel har et mere omfattende bevidsthedsteoretisk indhold, beror det på, at de baserer denne evne til at løse problemer på en generel model for fysisk-mekanisk symbolprocessing.

Selv om teorien ikke påstår, at bevidstheden kun består i en fysisk-mekanisk maskine, indebærer den altså en påstand om, at en central del af bevidstheden er en sådan maskine

At forestillingen om, at bevidstheden er et mekanisk-fysisk system, eller maskine på linje med alle andre maskiner, har provokeret andre end mig, kræver næppe nærmere dokumentation, men det adskiller for så vidt ikke denne teori fra så mange andre videnskabelige teorier.

Når jeg alligevel opholder mig lidt ved det provokatoriske moment, skyldes det to forhold. For det første, at det provokatoriske er blevet brugt som argument, der taler for teorien. De, der lader sig provokere over denne tanke, reagerer – sagde først Turing og siden andre som f.eks. Bruce Mazlish – blot med krænket selvfølelse, eller de sidder blot fast i en gammel religiøst funderet forestilling om åndens evige immaterialitet.⁴ For det andet, at det provokatoriske også er blevet lagt til grund for afvisningen af teorien, idet man simpelthen har affærdiget den på grund af dens mekaniske reduktionisme.

Selv om teorien både er mekanisk og reduktionistisk, er den det imidlertid på en ny måde, der giver anledning til at genoverveje forholdet mellem det fysiske, biologiske og psykiske.

Jeg vil her koncentrere mig om 3 grundlæggende aksiomer, der ligger bag forestillingen om kunstig intelligens, hvad enten vi tænker på den klassiske AI eller de senere formuleringer indenfor Cognitive Science, nemlig at

- at den menneskelige bevidsthed finder sted i tid og rum.
- at den menneskelige tænknings udfoldelse i tid og rum altid har en fysisk realisation. Så langt i hvert fald, at vi ikke kender nogen tankeproces, der ikke har en fysisk realisation.
- at der indgår fysisk realiserede mekaniske procedurer i menneskelige tænkning og intelligens. Dette kan muligvis ikke blot hævdes i forhold til bevidsthedsindholdet, det kan *måske* også hævdes for den fysiologiske *udførelse* af visse former for tænkning.

Når jeg nævner disse tre aksiomer beror det på, at de står uantastet tilbage, selv om man forkaster teoriens mekaniske og reduktionistiske udformning, ligesom de også – med deres rod i klassisk kontinuitetsfilosofi rummer en udfordring til alle de – ikke mekaniske – bevidsthedsteorier, der er funderet i en forestilling om en i sig selv beroende bevidsthedsfære.

Jeg vil nu opridsse nogle konsekvenser af disse aksiomer med vægt på tre ting.

For det første, at de tilsammen er uforenelige med forestillingen om, at den menneskelige tænkning – eller bevidstheden som hele – kan beskrives gennem mekaniske teorier. Jeg vil i den forbindelse skitsere et nyt kriterium for identifikationen af et bevidst system.

For det andet, at de også må indebære, at man forlader tanken om, at vi kan beskrive den fysiske natur med mekaniske principper, hvad enten vi vil beskrive naturen som en stor maskine eller som en mængde af mange små.

Og endelig for det tredje, at de tre aksiomer, hvori jeg følger teorierne om kunstig intelligens, ikke synes at kunne forenes indenfor rammerne af noget kendt begreb om forholdet mellem det fysiske, biologiske og bevidste – eller om forholdet mellem stof- og form- og regeldannelse.

Hvad jeg altså sætter mig for, er ikke at fremlægge en ny bevidsthedsteori eller ny erkendelsesteori, men snarere at bringe nogle allerede formulerede bevidstheds- og erkendelsesteoretiske dilemmaer en lille smule længere frem i forgrunden og skitsere nogle mekanismer, der muligvis kan forklare, hvordan bevidstheden »selv« arbejder med nogle af disse dilemmaer.

De tre konklusioners uforenelighed med mekanisk bevidsthedsteori.

Jeg skal nu prøve at skitsere min begrundelse for disse påstande.

Jeg vil her tage mit udgangspunkt i Herbert Simon's og Allen Newell's definition af et fysisk symbolsystem, hvorved de forstår et system, der består af 3 bestanddele:

a) et sæt af entiteter eller symboler, der er defineret ved et givet – og invariant – fysisk mønster.

b) en række mulige kombinationer af sådanne entiteter, indbyrdes ordnet i et samlet mønster, eksempelvis ved at en række symboler er anbragt i en rækkefølge. En sådan rækkefølge har således også en for den enkelte kombination bestemt fysisk form, der betegnes som et udtryk eller en sætning.

c) en række regler for hvordan en sådan kombination af entiteter kan omdannes til en anden, dvs. ny sætning, der igen danner en bestemt fysisk konstellation af fysiske enheder.⁵

Hvad vi her ser, er en ret let genkendelig reformulering af grundelementerne i Newton's mekaniske teori. Det enkelte symbol er en udelelig mindsteenhed på linje med definitionen af partikler i den newton'ske mekanik. Kombinationen af enkeltpartikler udgør tilsvarende en given tilstand i et samlet mekanisk system, og endelig modsvarer de regler, der kan transformere et udtryk til et andet i Newell og Simon's teori, ganske nøje Newton's begreb om de bevægende kræfter, der flytter rundt på tingene i det mekaniske univers.

Overensstemmelsen er endda så stor, at det i begge teorier kun er entiteterne og konstellationerne af entiteter, der er tænkt som fysiske størrelser, mens bevægelseskrafterne i begge tilfælde er immaterielle. De har ikke selv fysisk form, de virker bare på de fysiske former. Jeg skal ikke her gå i detaljer med spørgsmålet, om de hos Simon og Newell virker ligesom tyngdekraften – dvs overalt og samtidigt eller kun lokalt gennem tid og rum – blot minde om, at Newton's teori opererede med begge former.

Når vi møder en sådan fysisk symbolteori, altså en teori, der tager udgangspunkt i symbolernes fysiske form, bliver det første spørgsmål, hvad det er, der adskiller en fysisk form, der kun er fysisk form, fra en fysisk identisk form, der også optræder som symbolsk entitet i (eller for) bevidstheden?

Umiddelbart lyder spørgsmålet ret enkelt, for den ene form er jo blot en fysisk form, mens den anden også er en symbolsk form.

Nu var Newton's teori en teori om det fysiske univers, der forudsatte, at de enkelte bestanddele forelå fra begyndelsen. Den samme forudsætning er også tilstede i Simon's og Newell's teori, men her antager den et andet indhold, fordi man beskriver den menneskelige intelligens og andre fysiske symbolsystemer som et afgrænset finit fænomen i den fysiske verden. Det fysiske symbolsystem forudsætter således eksistensen af de fysiske størrelses definition som symboler. Denne evne til at definere fysiske størrelser som symbolske enheder i systemet indgår ikke i symbolsystemets egne egenskaber. Definitionen er forudsat, men den kan ikke afledes af systemets fysisk-mekaniske egenskaber.

Vi står her overfor ét blandt ganske mange eksempler på et dilemma, der følger med hver gang man forsøger at bruge den newton'ske model på beskrivelsen af afgrænsede fænomener, nemlig at man dermed smugler en transcendental forklaring ind som præmis for beskrivelsen af et fænomen, der er opstået længe efter det, vi med naturvidenskabelige termer kan kalde universets oprindelse, hvad der så end kan forstås herved.

Jeg kan ikke se andet, end at brugen af sådanne transcendentale forudsætninger i beskrivelsen af bestemte naturfænomener er uforenelig med den generelle videnskabelige grundsætning, hvorefter vi afviser mirakuløse begivenheder i naturen.

Vi må derfor, forekommer det mig, enten henføre eksistensen af fysisk manifesterede symbolsystemer til samme uklare begyndelsestilstand som resten af universet, eller også må vi opfatte disse symbolsystemers eksistens som et resultat af en naturhistorisk tilblivelse og udvikling.

Nu vil det på den anden side være vanskeligt at hævde, at vi slet ikke med nogen ret kan tage eksistensen af fysiske symbolsystemer for givet, eftersom denne påstand jo har denne eksistens som forudsætning. Vi kan altså antage, at der på en eller anden – i dag ukendt – måde er opstået, hvad jeg vil kalde en symbolgenerativ kompetence i naturen.

Det vil være vanskeligt at kræve, at en teori om kunstig intelligens kun

kan være legitim, hvis den kan besvare et spørgsmål, som ingen andre teorier har kunnet besvare, nemlig spørgsmålet, hvordan denne kompetence er opstået, altså spørgsmålet om de fysiske symbolsystemers tilblivelse.

Det man imidlertid med rimelighed kan kræve af en teori om fysiske symbolsystemer er derimod, at de overhovedet tillader denne problemstilling, og det vil, når det drejer sig om en mekanisk teori om et afgrænset fænomen, også sige en redegørelse for hvordan sådanne nydannelser 1) kan forenes med beskrivelsen af det fysiske system, hvori de dannes, og 2) opnår en sådan stabilitet, at de ikke bryder sammen eller forandrer sig i den videre proces. Dette er endda et specielt påtrængende problem for en mekanisk teori, idet disse forudsætter, at systemets bestanddele er karakteriseret ved et sæt af konstante egenskaber, der ikke forandrer sig – eller opløses i sine ikke-symboliske bestanddele – under påvirkningen af de kræfter, der bevæger systemet.

Man kan synes, det er strenge betingelser, jeg her opstiller, men det er ikke mig, der opstiller dem, det er betingelser, der indgår i definitionen af den mekaniske teori. Don't kill the messenger.

Hvis vi nu et øjeblik betragter denne problemstilling i lyset af Darwin's teori, eller rettere i lyset af den afstand denne teori skaber mellem universets og menneskets tilblivelse, og – for tankeeksperimentets skyld – fastholder en mekanisk beskrivelse af den fysiske natur, står vi med to centrale begyndelsesproblemer, der begge hører til et pænt stykke tid efter det, vi kan forestille os ved universets oprindelse. For det første, spørgsmålet om de levende organismers dannelse, og for det andet, spørgsmålet om dannelsen af den symbolgenerative kompetence og den menneskelige bevidsthed, intelligens og kultur.

Jeg siger udtrykkeligt dannelse og ikke udvikling, for så længe vi kun taler om udviklingsteori, taler vi om en udvikling af hhv. liv og bevidsthed, der allerede forudsætter disse fænomeners eksistens. Men når vi konfronterer disse udviklingsteorier med en mekanisk beskrivelse af den uorganiske natur, får vi et problem, eftersom mekaniske beskrivelser af den uorganiske natur ikke tillader dannelsen af nye lovmæssigheder eller fænomener med nye kvaliteter.

En mulig løsning på dette kunne være at beskrive liv og bevidsthed som produkter af nøjagtig samme universelle bestanddele og lovmæssigheder, vi hævder for andre fysiske fænomener – eksempelvis ved at forklare hvilke bestemte fysiske betingelser, der med nødvendighed frembringer de fænomener, vi betegner som liv og bevidsthed, herunder den teoretiske aktivitet, hvorigennem teorien derom selv frembringes.

En anden mulighed kunne være at supplere den mekaniske beskrivelse af den uorganiske natur med et enkelt eller måske to senere mirakler, nemlig dannelsen af livet og eventuelt også af bevidstheden.

Og en tredje mulighed kunne så endelig være, at drage den konklusion, at den mekaniske beskrivelse ikke kan gøres gældende som generel na-

turbeskrivelse, eftersom vi i denne natur finder fænomener med biologiske og symbolske egenskaber, der ikke er indeholdt i eller mulige at aflede fra nogen af de kendte mekaniske teorier om det fysiske univers.

Jeg bør nok her også tilføje, at der ikke i denne henseende er nogen væsentlig forskel på, om vi her tænker på den klassiske Newton-mekanik, den termodynamiske, relativitetsteoretiske eller kvantemekaniske beskrivelse, idet de alle bygger på aksiomerne om de grundlæggende entiteter og lovmæssighedens invarians. At disse teorier står i et indbyrdes spændingsforhold til hinanden, er blot et yderligere problem – for den mekaniske teori.

Spørgsmålet er nu, om alt dette har nogen betydning for bedømmelsen af de mekanisk funderede teorier om kunstig intelligens, og jeg skal være den første til at indrømme, at disse teorier jo meget vel kan være rigtige, selv om den mekaniske teori om den fysiske natur er forkert.⁶

Jeg kan også formulere spørgsmålet lidt anderledes, idet man med rimelighed kan sige, at spørgsmålet, hvordan vi forklarer bevidsthedens dannelse i naturen, især har teoretisk interesse, hvis svaret på dette spørgsmål har betydning for forståelsen af bevidsthedens virkemåde, når den først en gang er dannet. Man kunne jo tænke sig, at bevidstheden, når den nu engang var dannet, havde den egenskab, at den lukkede sig så effektivt fra sine omgivelser, at den kunne fungere som et finit, afgrænset fysisk realiseret symbolsystem, der udelukkende beroede på sine egne en gang for alle etablerede selvstændige virkningslove. Det er, så vidt jeg kan se, den konsekvens, Luhmann drager.

Men det efterlader så et billede af det fysiske, biologiske og symbolske (eller psykiske i vid forstand) (eller hos Luhmann, et billede af det biologiske, psykiske og kommunikative) som tre skarpt afgrænsede domæner – eller hvad vi kan kalde naturen som en dynamisk lagkage, hvor de kræfter, der virker i de enkelte lag ikke kan virke forandrende ind på de øvrige lags strukturer og virkemåder. En tidligere udgave af denne forestilling er den psykofysiske parallelisme, hvor det fysiske og psykiske er to forbundne, parallelle procesniveauer, med det psykiske som en overbygning, der respekterer og udnytter den underliggende fysiske mekanik efter sine egne love, hvordan de så end skulle være kunne opstået.

Denne tankegang er ikke fremmed i moderne vestlig tænkning. Vi kan finde en sådan forestilling om selvskabende og selvprocesserende systemer, der udskiller sig med et sæt af interne bevægelseslove, i Marx's begreb om kapitalen (eller i hvert fald i nogle tolkninger heraf) og den passer på sæt vis også med vores almindelige opdeling af videnskaben i enkeltvidenskaber, der dækker hvert sit afgrænsede domæne.

Hos Luhmann er der dog tale om en vis raffinering, fordi han ikke blot når frem til nye skillelinjer mellem systemerne, (herunder den noget overraskende – og for mig at se uholdbart skarpe – distinktion mellem bevidste og kommunikative systemer) men også har nogle interessante overvejelser over relationen mellem systemet og omverdenen.

Når jeg alligevel er noget skeptisk overfor Luhmann's model, skyldes det ikke blot en vis mistillid til hans biosociologiske fusion af Marx og Maturana, men også en mere specifik skepsis overfor den diskrete lethed, hvormed han passerer over det teoretiske svælg, der hos Marx kom til udtryk i beskrivelsen af overgangen fra den lovløse, kaotiske og vilde oprindelige akkumulation til en ubønhørligt lovmæssigt virkende akkumulationstvung. Altså, spørgsmålet om hvorledes der kan dannes nye niveauer for lovmæssigheder.

Problemet her er ikke, tror jeg, selve forestillingen om dannelsen af nye former for rationalitet, men snarere det, at man forsøger at beskrive disse nye rationalitetsformer som afgrænsede lukkede systemer, hvis virkemåde alene er begrundet i systemets interne struktur.

Hos Marx, hvor der er tale om sociale systemer, indebærer denne forestilling ikke blot en tese om en overgang fra en mere kaotisk tilstand til en mere ordnet – omend stadig konfliktfyldt – kulturtilstand, men også at de kræfter, der virkede i den tidligere tilstand, tilsyneladende fuldstændig ophører at virke i (eller subsummeres under) den senere.

Mens en sådan tanke måske ikke er helt umulig, når det drejer sig om et forhold mellem forskellige historiske epoker, ser det unægtelig anderledes ud, når det drejer sig om forholdet mellem forskellige samtidigt eksisterende systemer, idet de omgivelser, hvori systemet opstår, måske nok ændres, men ikke forsvinder eller ophører at virke, fordi der dannes et nyt system.

Tanken om verden som en mængde af finitte, lukkede og internt definerede systemer, der selv bestemmer vilkårene for udveksling med andre systemer, er en fælles basis for en væsentlig del af efterkrigstidens nye paradigmer. Fra Wiener's 1. ordens kybernetik over Maturana og Varela's frøer med det autopoietiske syn, til Luhmann's 2. ordens kybernetik, fra Turing's beskrivelse af den universelle computer, Shannon's kommunikationsteori til Simon, Newell og Shaw's 1. generationsteori om Kunstig intelligens og videre til 2. generationsteoriene i Cognitive Science. Fra den klassiske strukturalisme i lingvistikken til Chomsky's transformationsgrammatik.

Mens alle disse og mange andre teorier indbyrdes er uenige om det meste af det, de siger om verden, er de fælles om en grundlæggende tavshed om, hvordan de regelstrukturer og grundenheder, der virker i det enkelte system, er opstået.

Spørgsmålet er derfor om den ubestemthed, der gælder for tilblivelsen af et system og dets elementer, har nogen betydning for beskrivelsen af systemets virkemåde i fortsættelsen?

Mit svar på dette spørgsmål er ja, forestillingen om et finit, afgrænset helt selvregulerende rum er en forestilling, der kun findes som tankeindhold.

Nu er et sådant svar ikke noget særlig stærkt argument, og da det samtidig er en påstand om, at der altid indgår visse ubestemtheder i ethvert systems forhold til omgivelserne, er det heller ikke let at levere eksakte bevisligheder i sædvanlig forstand.

Jeg vil derimod demonstrere, hvordan der indgår noget, som jeg foreløbig vil kalde indbygget strukturel ubestemthed i den form for fysisk informationsprocessering, der danner model for disse bevidsthedsteorier og vise, at der her findes evidens for eksistensen af forskellige former for strukturel ubestemthed, der angår afgrænsningen af det enkelte system fra andre systemer, hvorved selve afgrænsning forbliver ufuldstændig og uafsluttet.

Jeg vil med det samme præcisere, at jeg ikke her kun taler om den ubestemthed, der defineres som en mængde af frie valgmuligheder i et givet system, men om ubestemthed, der også både angår systemets regelstruktur dvs. reglernes indhold og virkemåde, og stabiliseringen af systemets enheder. Det vil fremgå, at denne ubestemthed i nogle tilfælde har at gøre med relationen mellem et systems stabilisering i forhold til det, Luhmann kalder dets omgivelser, hvad der blot er et andet navn for under- eller overliggende systemer.

Jeg har foreløbig brugt betegnelsen strukturel ubestemthed, men de fænomener og relationer, jeg tænker på, er bl.a. de fænomener, der i en række sammenhænge omtales med termen *redundans*. Det er ikke termen, der er afgørende, det er derimod spørgsmålet den strukturelle ubestemthed eller *redundans*'s funktion og derigennem naturligvis også spørgsmålet, om vi kan definere dette begreb på en måde, der kan give ny viden?

Før jeg går nærmere ind på dette begreb, vil jeg knytte en bemærkning til *redundans*begrebets betydning i almindelig sprogbrug, hvor det ofte anvendes om noget, der er overflødigt, noget betydningsløst eller betydningssvagt, som vi enten kan se helt bort fra eller simpelthen smide væk. Vi anvender det almindeligvis ikke om noget, der blot forekommer en enkelt gang eller tilfældigt, men derimod om noget, der forekommer med en vis regelmæssighed, dvs. gentagelige, mere eller mindre overflødige mønstre. Eller genkommende mønstre, der ikke har nogen påviselig, bestemt funktion og/eller betydning. Her bruges begrebet altså i direkte modsætning til begrebet om det regelbestemte, hvorved vi forstår genkommende mønstre, der har en påviselig – regulativ – funktion og/eller betydning. Det interessante er det fælles element, der er forudsætning for denne modsætning, nemlig at der i begge tilfælde er tale om repetitivt genkommende mønstre, hvoraf det fremgår, at forskellen mellem *redundante* og regelbestemmende mønstre alene ligger i disse mønstres funktion/betydning.

Man kan med udgangspunkt i denne lille semantiske manøvre formulere et begreb om *redundans*systemer, der alene adskiller sig fra begrebet

om regelbestemte systemer i kraft af, at de genkommende mønstre i et redundant system ikke har en generel, forudbestemt funktion eller betydning. Man kan godt forestille sig, at et fuldstændigt regelbestemt system indeholder redundante mønstre, der i så fald må forblive betydningsløse. Man kan ligeledes forestille sig, at der kan indgå regulative mønstre i et redundant system. Så længe, der er tale om helt stabile systemer, er valget af term et ret sofistisk anliggende.⁷ I det øjeblik vi imidlertid betragter systemer, der tillader dannelsen af nye regler, nye funktioner og nye betydninger, giver de to forskellige systembetragtninger helt forskellige forklaringsmuligheder.

Mens et regelbestemt system kun kan tillade nye regler, der kan afledes af de forud eksisterende regler, kan et redundanssystem også tillade dannelse af nye regler gennem – nye – udnyttelser af eksisterende mønstre, der derigennem kan få ændret deres funktion. Redundante systemer har dermed en egenskab, der ikke kan indeholdes i et regelbestemt system. Værdien af denne egenskab er især, at den i kraft af systemets repertoire af ubenyttede mønstre, giver systemet en innovationsmulighed, der ikke kan findes i et fuldstændigt regelbestemt system.

Jeg vil nu, som skridt på vejen til forslaget om, at man antagelig må beskrive bevidstheden som et neurofysiologisk forankret redundanssystem, kort skitsere, hvordan denne forståelse kommer til sin ret, dels som grundlag for beskrivelsen af almensproget, dels som grundlag for beskrivelsen af den form for informationsbehandling, der finder sted i en computer, idet sidstnævnte eksempel samtidig fjerner grundlaget for den – computationelle – bevidsthedsmodel, der er forudsat i teorierne om kunstig intelligens.⁸

Tanken om at betragte almensproget (talt hhv. skrevet) som redundanssystem(er) har rod i flere forhold, heriblandt de teoretiske problemer med at beskrive 1) forholdet mellem diakrone og synkroner relationer, 2) forandringer i sprogstrukturen, der ikke lader sig aflede af strukturernes interne regler 3) almensprogenets plasticitet, 4) forekomsten af såvel over- som underbestemthed og variation af etablerede mønstre hhv af deres regulative funktioner samt – naturligvis – 5) det forhold, at almensproget bygger på en mangfoldighed af figurer og mønstre, der ofte er overflødige eller optræder med skiftende funktioner og betydning, beroende på sprogbrugerens valg i en given kontekst.

De mekanismer, der her sigtes til, kan findes på alle sprogets planer, på det semantiske og syntaktiske niveau og i anvendelsen af det basale – alfabetiske og stavelsesmæssige – inventar.

Som eksempler kan man nævne det velkendte forhold, at vi ofte både bøjer det overordnede og underordnede led, selv om bøjningens betydningsindhold er entydigt angivet, blot vi bøjer det ene led. Det hedder f.eks. *en led* mand, men *den lede* mand. Her har vi – et regelbestemt – eksempel på overbestemthed, hvor selve reglen er overflødig. Omvendt har vi mange andre eksempler på underbestemthed. Det gælder for eksempel

de mange tilfælde, hvor vi bruger samme ord i forskellige – beslægtede eller ubeslægtede – betydninger. En *vis* mand og en *vis* (*viis*) mand skrives i dag på samme måde, selv om der vist helst skulle være en *vis* forskel.

Underbestemthed optræder imidlertid ikke kun i disse mere eller mindre tilfældige eller marginale tilfælde. Den er konstitutiv for almensproget, idet vi her – med de termer, der gælder for skriftsproget – betjener os af et begrænset sæt af betydningstomme figurer, alfabetet, der igen danner grundlag for et yderligere sæt af andre betydningstomme mønstre, stavelsesinventaret, der står til rådighed for såvel ord- som bøjnings- og sætningsdannelse.

Man kan overfor disse velkendte træk spørge sig selv, hvordan det skulle være muligt at beskrive en bestemt regelstruktur for anvendelsen af det sproglige udtryksinventar, når det for en så hyppigt anvendt sekvens som bogstavfølgen /*er*/ f.eks. gælder, at den både optræder som bøjningsendelse for substantiver (pluralis) og verber (præsens) som del af komparativformen (eks: *højere*) som selvstændigt ord (præsensform af at være) og i en meget stor mængde af ordstammer?

Nu har jeg ganske vist her netop anført en række regulative mekanismer for anvendelsen, der for bøjningsformernes vedkommende kan siges at tilhøre sprogets regelstruktur, men det forklarer ikke overforbruget af netop denne – i sig selv underbestemte – sekvens sammenlignet med så godt som alle andre mulige sekvenser.

Eksemplerne giver ikke i sig selv anledning til tvingende konklusioner, men de illustrerer meget godt, at sproget – al lovmæssighed til trods – er overfyldt med forekomster af *overbestemthed* (redundans i udtryksformen) og *underbestemthed* (der tillader forskellige former for indholdsvariation af samme udtryksform) og at begge disse træk danner grundlag for ikke-regelbestemte relationer mellem udtryks- og meningsmønstre.

På udtrykssiden finder vi mange mønstre, der går igen med varierende betydninger, og på indholdssiden tilsvarende, at den samme betydning kan udtrykkes på en flerhed af måder, og at disse relationer er baseret såvel regelbestemte som ikke-regelbestemte anvendelser af det underliggende system af betydningstomme mønstre.

Alt dette handler i sidste ende om det forhold, at vi kun har de samme 28 bogstaver til at udtrykke en hvilken som helt tænkelig mening. Man kan derfor beskrive alfabetet og stavelsesinventaret som en stabiliserende redundansstruktur, der er forudsætning for den sproglige aktivitet, idet vi kun har disse mønstre til at manifestere såvel regler som mening, mens vi omvendt, netop fordi vi har dette underliggende redundanssystem, også kan have et ikke fuldstændigt regelbestemt sprog.

Beskrivelsen af sproget som et redundansbaseret system giver ikke blot mulighed for at forklare, hvordan vi med et variabelt sæt af reguleringer i dette system kan blive med at udtrykke nye meninger. Den giver også mulighed for at forklare, hvordan der kan dannes nye regler, der

ikke kan afledes fra de eksisterende.

Den grundlæggende mekanisme bag disse forhold består i en treleddet figur, der tager sit udgangspunkt i muligheden for at sammensætte alfabetets bogstaver i et – stort – sæt af mulige mønstre (stavelser og endelser). Det første led består her i selve mønstrets dannelse, den første forekomst så at sige.⁹ Andet led udgøres af gentagelsen af forekomsten, der kan være forbundet med (ny) betydning, men ikke behøver at være det, hvorefter det tredje led består i at udnytte et sådant gentageligt mønster til et regulativt formål.

Som eksempel på denne mekanisme for regeldannelse i redundansbaserede systemer kan man nævne en sproglig nydannelse, der opstod for 10-15 år siden, nemlig den i visse miljøer udbredte brug af termerne: børneren, døgneren, fritteren, fjernerer og trykkeren, for de tidligere former: børnehaven, døgnkiosken, fritidshjemmet, fjernsynet og fjernbetjeningen (som man trykker på). Reglen i disse nydannelser er simpel, den består i at erstatte 2. leddet i en række sammensatte ord med en – betydningsvækket – bekendthedsendelse. I det sidste tilfælde erstattes hele termen dog først med en funktion, der er knyttet til termen og dernæst anvendes reglen på funktionsordet. Reglen er klar, og falder i forlængelse af en bredere virkende tendens til at svække mindre væsentlige dele af udtrykkene. Reglen for dens anvendelse – hvilke ord der rammes – er sådan set også klar, men den tilhører på ingen måde sprogets struktur. De ramte ord har derimod det til fælles, at de i perioden udgjorde centrale omdrejningspunkter i mange børns og forældres livsverden. Eksemplet viser, at der her er tale om en ny regel, der ikke kan afledes af eksisterende sprogregler. Dannelsen af den nye regel tager derimod sit udgangspunkt i en første anvendelse, der ikke er regulativ, og hvor svækkelsen antagelig er anvendt med et særligt meningsmarkerende motiv, for derefter at blive akcepteret og udbredt til at gælde som regulativ for en række andre termer.

En lidt anderledes, men beslægtet mekanisme er tilstede i en anden nydannelse fra de senere år, nemlig anvendelsen af ordet rimelig i betydningen temmelig. Det er et rimeligt interessant tilfælde, fordi vi her har et eksempel, hvor et udtryksmønster (ordet rimelig), der allerede er forbundet med en bestemt semantisk værdi (en rimelig pris = acceptabel pris) får udskiftet sin betydning, mens den gamle betydning svækkes eller forsvinder. Om den forsvinder helt, eller består i andre sammenhænge eller blot overvintrer til en senere lejlighed kan næppe afgøres definitivt. Det er derimod klart, at der igen er tale om ikke-regelbestemt forandring i relationen mellem udtryks- og indholdssiden, og at denne forandring har taget sin begyndelse med en første – meningsmotiveret – anvendelse, der siden har vundet en bredere akcept, og er blevet indlejret som en så fast givet betydning, at det i hvert fald i en periode kan anvendes som almindelig hævdunden term. Med disse vilkår ville jeg ikke påtage mig at lave et program for automatisk sprogovættelse.

Så vidt det almensproglige redundanssystem.

Den mere overordnede konklusion på disse indikationer er den, at redundansbaserede systemer – som almensproget – kan beskrives som *regeldannende* systemer, der har egenskaber, (hvoraf den regeldannende er en af de mest centrale) der ikke er indeholdt i *regelbestemte* systemer. Jeg vil nu demonstrere, at gruppen af regeldannende (redundansbaserede) systemer også omfatter det fysisk-mekanisk definerede symbolsystem, der anvendes i computere, og at denne beskrivelse samtidig kan vise, hvordan den computationelle forståelse af bevidstheden som et diskret, finit og regelbestemt symbolsystem kommer til kort.

Jeg vil her tage mit udgangspunkt i det forhold, at en computer kun kan effektuere en regel eller et program eller bare en algoritme under den forudsætning, at reglen udtrykkes i nøjagtig samme form som alle de data, den skal regulere, nemlig som en sekvens af – binære – notationsenheder. Dette er et i sig selv velkendt – nærmest trivielt – forhold, det er implikationerne imidlertid ikke. For mens vi almindeligvis forstår computeren som en maskine, der styres ved hjælp af et program, så foregår denne proces i realiteten på et underliggende fysisk niveau, hvori regelvirkningen fremkommer som resultat af den proces, vi siger, den regulerer. I den daglige anvendelse er det altafgørende, at denne virkning faktisk fremkommer med stor sikkerhed, og vi er derfor tilbøjelige til at forstå det som en selvfølgelig og automatisk – programmeret – procedure.

Kravet om, at også programmet foreligger i den binære repræsentation, indebærer ikke blot, at det, vi opfatter som et forudgivet program, der styrer processen, i realiteten er et produkt af processen, det indebærer også, at en hvilken som helst regel, der fremkommer, kun fremkommer fordi den er blevet behandlet på nøjagtig samme måde – og foreligger i samme alfabet – som alle andre data. I konsekvens heraf kan en hvilken som helst regelstruktur blive genstand for variation i sit indhold og/eller modifikation i sit udtryk, ligesom området for dens anvendelse kan variere i udstrækning til og med fuldstændig suspension.

Dette gælder, uanset om vi ved regelstruktur her forstår et program eller blot en algoritme, en semantisk eller en syntaktisk struktur, og det gælder, fordi det krav, der stilles til en universel regnemaskine, netop er, at den ikke må være begrænset af en enkelt eller et vist defineret sæt af bestemte regler, idet den i så fald netop ikke ville være universel. Kravet om universalitet består således i et krav om, at computeren kan udføre en vilkårlig regneregulering, og dette krav kan kun opfyldes under den forudsætning, at den selv opererer uafhængigt af enhver regneregulering.

De regler, der skal effektueres, kan derfor ikke være en indbygget, invariant del af maskinen, de skal tværtimod specificeres sammen med det materiale, der skal reguleres. For at indfri denne betingelse må maskinen benytte sig af et alfabet, der består af et forud fastlagt, endeligt antal medlemmer. Antallet skal være forud fastlagt, fordi maskinen skal benytte dette alfabet som mekanisk virksomme størrelser, og alfabetet skal bestå

af semantiske tomme figurer, fordi disse figurer undervejs skal kunne varetage betydningsmæssigt vidt forskellige funktioner. I den – nu brugte – binære udgave af dette alfabet, er der som bekendt kun to bogstaver, og disse to bogstaver skal både tjene som udtryksform for regler, for data, for tal, for bogstaver, for elementer af billeder eller lyd, eller hvad vi nu iøvrigt behandler af indhold.

Kravet om at både program og data skal repræsenteres og behandles i et og samme alfabet betyder, der ikke i en computer findes nogen invariant grænse mellem maskinen og materialet, mellem reglen og det regulerede, eller mellem den viden, der er indbygget i maskinens arkitektur og den viden, der behandles.

Selv om der er visse forskelle mellem dette – informationelle – alfabet og det skriftsproglige, har de den samme grundlæggende struktur. Antallet af bogstaver er forskelligt, det sproglige alfabet har visse kvalitetsbestemmelser (som distinktionen vokal-konsonant) og vi bruger også i nogle få tilfælde de enkelte bogstaver som semantiske størrelser (i, ø, å), mens det informationelle alfabet er defineret uden kvalitetsbestemmelser og enhver anvendelse af enkeltbogstaver som selvstændig semantisk værdi er udelukket. Til gengæld besidder det informationelle og helt kvalitetsløse alfabet en endnu mere omfattende kapacitet end det sproglige, idet vi både kan repræsentere almensproglige og formelle udtryk samt talrige andre billedlige og visuelle udtryk i et og samme alfabet, der tilmed har mekaniske egenskaber.

Som kontrast til disse – redundansbaserede – symbolsystemer, der tillader ikke-regelbestemt regeldannelse, kan vi anføre de formelle symbolsystemer. Disse er for det første ikke underlagt noget krav om et begrænset – eller helt fast – antal notationsenheder, vi kan frit indføre nye notationer i en given kontekst, blot vi deklarerer deres semantiske værdi. Dette sidste krav gælder endvidere samtlige notationer i et formelt system, herunder også dem, der defineres som variable, idet de skal defineres som sådanne indenfor et eller andet sæt af værdivariationer. Formelle symbolsystemer forudsætter tilsvarende (også til forskel fra redundansbaserede notationssystemer) en distinktion mellem regel- og data-notationsenheder.

Man kan vel godt sige, der er noget paradoksalt i det forhold, at århundreders bestræbelser på at skabe et formelt, logisk universalsprog for sammenfatningen af den menneskelige viden, fandt sin foreløbige kulmination og afslutning i konstruktionen et nyt, helt asemantisk alfabet, der har mere fælles med det skriftsproglige alfabet end med formelle sprogrincipper, og snarere må opfattes som en perfektionering – og ganske dramatisk udvidelse – af skriftbaserede end formelle repræsentationsformer.

De træk ved det informationelle alfabet, jeg her har beskrevet, skulle være tilstrækkelige til at vise, at den computationelle model af bevidstheden, der danner grundlag for ideen om kunstig intelligens, bygger på

en fejlagtig beskrivelse af det informationelle alfabet, idet man tolker det som et formelt og regelbestemt notationssystem, hvor der er en bundet ækvivalens mellem de notationelle, syntaktiske og semantiske niveauer, mens det informationelle alfabet tværtimod kun kan virke, fordi der ikke er en sådan bundet ækvivalens.

På den anden side synes beskrivelsen af den informationelle repræsentation som redundanssystem, der tillader ikke-regelbestemt regeldannelse at lægge op til en bevidsthedsteoretisk tolkning, der bedre svarer til ønsket om en bevidsthedsteori, der er forenelig med forestillingen om, at vi har en vis frihed til at tænke noget – eventuelt måske endda noget nyt.

Her gælder det imidlertid om ikke blot at erstatte en fejl med en anden. Selv om den informationelle notation har en semantisk plasticitet, har den også to begrænsninger, der begge giver anledning til at betvivle, at der består en direkte analogi mellem dette notationssystem og det neurofysiologiske systems virkemåde. Den ene begrænsning er den, at den informationelle notation er defineret omkring en ret simpel mekanisk struktur, der til gengæld kræver en fuldstændig synkronisering af alle de processer, der foregår. Selv om jeg ikke har nogen dybere indsigt i moderne neurofysiologi, forekommer det mig at være temmelig usandsynligt, at vores samlede sanse- og bevidsthedsmæssige aktiviteter skulle være medieret i et – synkroniseret – fysisk realiseret, binært alfabet. Min tvivl på dette punkt hænger sammen med den anden begrænsning, nemlig, at den fysiske effektivering af de informationelle processer i den computerbaserede model forudsætter en algoritmisk styret afvikling. Det er nu for det første uklart, hvem der inde i vore hoveder kan levere sådanne algoritmer, men endnu vanskeligere at forklare, at vi selv er i stand til at håndtere, det, vi kan kalde uafgjorte situationer, der ville bringe enhver computer til at gå i stå, hvorefter den vil blive stående, indtil en behjertet programmør forsyner den med nye instruktioner.

Det fremgår heraf også, at det ikke kun er de mere rationalistiske kognitionsteoretiske tolkninger, der kommer til kort, men også de neurale netværksteorier og de dermed forbundne begreber om parallelprocesserende systemer.

Når det informationelle alfabet ikke desto mindre kan give anledning til bevidsthedsteoretiske overvejelser beror det – ud over de nævnte »negative« bidrag til at præcisere forskelle – på, at vi her også har fået et alfabet, der gør det muligt at belyse et andet og mere generelt træk ved forholdet mellem det fysiske og et symbolske med en større præcision end tidligere.

Der har i tidens løb været gjort mange forsøg på at beskrive forbindelserne mellem de symbolske udtryks fysiske former og deres mening. Et af motiverne har været ønsket om at kunne forklare, hvordan det går til, at vi uden videre kan opfatte en række fysisk forskellige figurer som en og samme figur, for eksempel hvordan det går til, at vi almindeligvis ikke har svært ved at identificere en lang række forskellige fysiske udgaver af

et bogstav (f.eks. individuelt forskellige håndskrevne varianter eller forskellige sæt af fonts på en computer) som et og samme bogstav?

I det perspektiv drejer det sig om en vis inkongruens mellem det fysiske og mentale. Vi besidder åbenbart en mental kapacitet, der gør os i stand til at kode fysiske variationer i kategoriale grupper, hvori visse fysiske variationer udpeges som mindre betydende til forskel fra andre. Da udgangspunktet imidlertid har været eksisterende symbolske udtryksformer, som f.eks. alfabetiske, har man oftest opfattet de mindre betydende variationer som udtryk for, at disse symbolsystemer var belastet af en manglende præcision og derfor også tolket variationerne som – uhenigtsmæssig og unødigt – støj, der bedre kunne undværes.

Det var også denne problemstilling, der var udgangspunktet for Claude Shannon, da han sidst i 1940-erne formulerede en matematiske kommunikationsteori. Det oprindelige mål var først og fremmest at udvikle matematiske metoder til at komprimere en meddelelse, således at den kunne transmitteres i den mest økonomiske form. Shannon anlagde imidlertid en mere generel betragtning, idet han tog udgangspunkt i et ønske om at finde en metode, der kunne anvendes på et vilkårligt sæt af meddelelser, der forelå i notation, der alene skulle være defineret ved entydige fysiske værdier.

Det i nærværende sammenhæng mest interessante resultat af hans arbejde er, at han – så at sige til trods for den entydige fysiske definition af notationssystemet – også kunne vise, at en sådan entydig fysisk definition aldrig kan være tilstrækkelig til at definere en symbolsk udtrykshed. Forklaringen er den forholdsvis simple, at en hvilken som helst fysisk form, der kan udnyttes som symbolsk udtryksform, også kan fremkomme som fysisk form uden at være intenderet som symbol. Vi har simpelthen ikke andre fysiske former, vi kan bruge til symbolske formål, end dem naturen tillader. Så vidt gik Shannon ikke selv, men han klargjorde problemets kerne, idet han viste, at der til ethvert fysisk nok så veldefineret symbol hører to forskellige støjproblemer. Det ene støjproblem er det almindeligt kendte, at vi skal kunne skelne det enkelte symbol fra den fysiske baggrund og fra andre symboler hhv. fysiske varianter af samme symbol. Det andet støjproblem er, så vidt jeg kan se, mindre kendt, og består i det forhold, at vi også skal kunne skelne en fysisk form, der har symbolsk legitimitet, fra en fysisk identisk form, der ikke har.

I Shannon's sammenhæng bestod problemet i, at det var nødvendigt at finde et middel til at afgøre, om et modtaget 0 var intenderet, dvs afsendt som et 0, eller om det var fremkommet som følge af støj i transmissionsmediet. Det middel, Shannon fandt til at sikre denne distinktion, bestod i at øge meddelelsens redundans.

Det hører her med, at han løb ind i dette problem og fandt svaret, fordi han var startet med den stik modsatte bestræbelse, nemlig med at komprimere meddelelsen ved at reducere eller eliminere den redundans, den indeholdt i forvejen. Når der alligevel er en pointe, beror det på forskel-

len mellem den redundans, han eliminerer, og den, han tilføjer. Mens den første bliver bestemt på notationsniveauet, (ved – f. eks. statistisk – måling af hyppigheden af de enkelte notationers forekomst) bliver den anden bestemt gennem en formel, semantisk kodning, der så enten tilføjes den komprimerede meddelelse, når den sendes, eller indbygges som dekoder i modtageapparatet.

Det interessante heri er bl.a.:

1. at en rent fysisk definition af en notationsenhed ikke er tilstrækkelig, lige meget hvor præcis og entydig den er. Den legitime fysiske form kan nemlig også – altid – eksistere eller opstå i systemet uden at være intenderet – dvs. som følge af støj. Enhver fysisk defineret notation må også forbindes med nogle ikke-fysiske kriterier (ie: semantiske) for legitimitet forekomst. Redundansen i notationssystemet er nødvendig for notationens stabilisering i forhold til det fysiske medie/underlag/udtrykssubstans. Redundansen spiller altså en vigtig rolle for sikringen af den fysiske notations legitimitet som symbol.

2. at redundansfunktionen både spiller en rolle for stabilisering af notationen i forhold til det fysiske medie (legitimiteten) og for stabiliseringen af meningen i notationssystemet. Der er tilmed en indre sammenhæng herimellem, fordi det er muligt at kompensere en reduktion af redundans på det ene niveau ved at øge den på det andet. (f.eks. kompensere reduktion af statistisk bestemt redundans med forøgelse af semantisk bestemt redundans). Redundans i udtrykssystemet kan substitueres med indholdsbestemt redundans.

Det mærkelige resultat af dette er så, at det fører til en konklusion, der strider mod den gængse definition af redundansbegrebet, idet redundansfunktionen er en nødvendig forudsætning for opretholdelsen og overførslen af meddelelsens mening. Den er med andre ord hverken overflødig eller meningsløs. Den er uomgængelig – og den tillader et symbolsk system at variere de mønstre, der tilbydes af det underliggende fysiske eller fysiologiske system.

Jeg kan sammenfatte dette i en kort definition af, hvad der kan forstås ved et redundanssystem, nemlig:

et – selv variabelt – sæt af gentagelige mønstre, strukturer eller systemer, der er karakteriseret ved muligheden for valgbar variation i mønstrenes virknings og/eller betydningsstyrke, og/eller betydningsindhold.

og

som tillader – eller beror på valgbare, meningsbærende anvendelser af mønsterafvigelse og mønstervariation.

Sådanne systemer har fire variationsakser, nemlig

- 1) en akse, der angår betydningsstyrken.
- 2) en akse, der angår mønstervariationen.
- 3) en akse, der spænder fra et mønsters første, eventuelt tilfældige forekomst over gentagelse af denne forekomst til udnyttelse af gentagelsen som regulativ.
- 4) en akse, der angår betydningsindholdet, idet betydningsvariation kan være forbundet med variation på hver enkelt af de andre akser, såvel som med samtidig variation på flere.

Ved betydningsstærke og invariante mønsterforekomster falder begrebet redundans sammen med begrebet lovmæssighed, struktur og/eller ny betydning.

Ved betydningsssvage forekomster af ufuldstændige gentagelser – afvigelser og variation af mønstre, f.eks. individuelle håndskriftsvariationer – nærmer begrebet sig betydningen støj, der evt. i et andet semantisk perspektiv, f.eks. grafologisk, kan tolkes som betydningsbærende.

Herimellem kan der være mange grader, der kombinerer forskellige betydningsstyrker med forskellige grader af mønstervariation. Betydningsstærk mønstervariation vil normalt føre til ny mening, snarere end ny regeldannelse, men en ny regel må altid begynde som ny mening – der enten ligger i selve gentagelsen eller i at give en gentagelse regulativ betydning for noget andet. Enhver regeldannelse begynder med en form, der endnu ikke er regelbærende.

Vi kan endelig her – ved små eller store afvigelser fra et givet mønster – se en forbindelse mellem afvigelse og ny form. Enhver afvigelse fra et givet mønster kan blive ny – legitim – form, der enten straks eller siden forbindes med ny betydning. Hvis et mønster opstår ved gentagelse af noget, der forelå som form før gentagelsen, kan selve gentagelsen være form-innovativ.

Det overordnede sigte med dette redundansbegreb er at beskrive nogle træk ved symbolske systemer, der ikke kan indeholdes i strukturalistiske, formale og regelbestemte symbolteorier, og det overordnede argument for dette begreb om redundanssystemer ligger i, at det tillader:

- Muligheden for at en regel kan dannes ved gentagelsen af en form, der ikke var regel før gentagelsen. Altså dannelse af nye regler, der ikke er afledt af eksisterende regler. (Men evt skabt ved at udnytte udtryks-substansens former på nye måder, som det bl.a. kendes fra æstetisk-kunstnerisk sammenhænge).
- At nye regel- og betydningsniveauer kan opstå ved at benytte etablerede regler og betydninger som redundanspotentiale, idet det nye betydningsniveau hverken er helt uafhængigt eller helt bundet til det underliggende niveaus (og tidligere eksisterende) regelstruktur.

- At en regelstruktur kan suspenderes eller modificeres gennem regelafvigelse til semantiske formål eller i reguleringen af overlappende regelsystemer uden faste vigeregler – som vi kender det i sproget.

Selv om man nu – idet mindste som et muligt tankeeksperiment – godtager denne beskrivelse af almensproget og det informationelle symbolsprog som redundansbaserede systemer, følger det dog ikke uden videre, at vi også kan henføre den menneskelige bevidsthed til denne type. Både det sproglige og informationelle alfabet er således eksternaliserede udtryksformer, der selv forudsætter en mental kodningskompetence. Forståelsen af bevidstheden må i det mindste indbefatte et moment, der ikke kan beskrives på dette grundlag, nemlig selve eksistensen af en symbolgenerativ kompetence, der set fra mit synspunkt formentlig må opfattes om et resultat af en naturhistorisk udvikling, der må være indtruffet et godt stykke tid efter en lang række andre naturhistoriske processer.

Selv om vi kan forklare mange symbolske processer som resultat af andre symbolske processer, kan vi ikke forklare dem alle – og i hvert fald ikke det første symbols dannelse – på den måde, fordi sådanne forklaringer forudsætter en allerede eksisterende symbolsk kompetence, der må opfattes som forudsætning også for evne til at kunne foretage en kodning.

Mens oprindelsen stadig henligger uforklaret, giver den redundansbaserede forståelse af symbolsk aktivitet imidlertid et på nogle punkter mere plausibelt billede af bevidstheden, fordi den ikke stiller krav om en pludselig dannelse af et fuldt udviklet symbolsk regelsystem, hvorudfra den senere symbolske aktivitet kan forklares. Den kritiske forskel mellem en regelbaseret og redundansbaseret forståelse er her igen indeholdt i spørgsmålet om bevidstheden kan danne nye regler, der ikke kan afledes af allerede eksisterende regler?

Risikoen ved at svare ja til dette spørgsmål er først og fremmest, at man kan lade sig stille tilfreds med det, vi allerede tror, vi ved, og ikke søge at finde lovmæssigheder i det, der foreløbig forekommer uforklaret. På den anden side virker det lidt bagvendt, at prøve at forbinde en kontinuitetsfilosofisk forestilling om universet med en beskrivelse af verden som en mængde af hver især regelbestemte, lukkede systemer, der kun har visse – selv regelbestemte – forbindelser, når vi samtidig kan se mange indikationer på strukturel ubestemthed, redundans og støj i forbindelserne og udvekslingen mellem disse systemer – fra den kosmiske baggrundsstøj til det, vi ved introspektion mener at kunne sige om indholdet i vores egen bevidsthedsliv.

Der er adskillige hypotetiske og spekulative momenter i dette ræsonnement, og jeg kan også tilslutte mig Turing's betragtning, man skal aldrig sige aldrig, og derfor ikke definitivt udelukke muligheden for at beskrive bevidstheden som et fuldstændigt og afgrænset, finit system.

Jeg vil dog samtidig foreslå et håndgribeligt testkriterium for, hvordan det kan afgøres, om bevidstheden er et regelbestemt eller redundansba-

seret system ved at formulere en nødvendig, men muligvis ikke tilstrækkelig, betingelse for intelligent adfærd.

Dette forslag kan, også i Turing's ånd, endda udmøntes som et testkriterium for en menneskeligt konstrueret, tænkende maskine, idet man til en sådan maskine må kunne stille et krav, vi ved bevidstheden opfylder, nemlig, *at den kan generere sit eget symbolske udtrykssystem*.

Jeg er overbevist om, at en sådan maskine også vil være i stand til at håndtere uafgjorte tilstande uden at gå i stå, og at den kan det fordi den også – selv – besidder evnen til at arbejde med flere forskellige former for semantisk redundans – uden altid at skulle afvente hjælp fra en ekstern programmør. Jeg tror også, men jeg kan jo ikke vide det, at den vil besidde de biologiske egenskaber, der ligger i dødsvilkåret og driftsorienteringen.

En sådan maskine ligner ikke nogen hidtidig beskrevet maskine, og vi kan derfor roligt fjerne teorien om kunstig intelligens og cognitive science som adækvate modeller for den menneskelige bevidsthed. Men vi kan til gengæld ikke fjerne den underliggende ide om bevidstheden som et fænomen, der udfolder sig i tid og rum, og i samme univers som det, den mekaniske fysik er teori om.

Dermed har jeg ikke udelukket muligheden for at tænke sig en tænkende maskine, og jeg tør godt love, at jeg, hvis eller når jeg møder en maskine, der kan generere sine egne symbolske udtrykssystemer, hvad enten den er regelbestemt eller ej, vil være den første, der viser den vejen til folkeregistret, og den sidste der forlader barrikaden i kampen for denne maskines ret til ytringsfrihed.

NOTER

1. I det øjeblik bevidstheden erkendes som isoleret og adskilt fra andre bevidstheder, opstår spørgsmålet, hvordan bevidstheder kommunikerer indbyrdes/med omverdenen. Moderne bevidsthedsteori implicerer således en særskilt kommunikationsteori, og disse to teoriområder kan samtidig ses som en moderne rekonstruktion af den klassiske dikotomi mellem mellem logik og retorik, men nu på videnskabsinterne præmisser.
2. Jfr. Simo Køppe (1993) p. 454 m. fl. steder. Køppe opererer med fire niveauer, det fysiske, biologiske, psykiske og sociale. De to sidste opfattes som gensidige forudsætninger, der igen forudsætter de to første, ligesom det andet forudsætter det første. Ifølge Køppe indebærer disse forudsætningsrelationer, at de højere niveauer må respektere de love, der gælder de lavere, men de højere niveauer kan ikke reduceres hertil. En sådan model må imidlertid implicere et begreb om det fysiske, der overhovedet tillader, at fysiske processer kan danne biologiske, psykiske og sociale niveauer, og spørgsmålet er så, hvad der kan menes med fysiske love?
3. Finnemann (1994).
4. Turing (1950) og Mazlish (1966).

5. Definitionen er hentet fra Simon, H. & Newell, A. (1976) – og altså noget nyere end den oprindelige teori. Artiklen er udarbejdet som foredrag, hvori de skulle præsentere og sammenfatte grundlaget for forskningen i Kunstig Intelligens. De udpeger her denne symbolteori samt deres begreb om heuristisk søgning som de to – stabile grundelementer.
6. Det ser i hvert fald ud til, at det lettere at aflede et mekanisk system af et ikke-mekanisk end omvendt. Vi er således i stand til at konstruere nye maskiner uden at have en mekanisk beskrivelse af, hvordan vi gør det, mens vi ikke har noget eksempel på en maskine, der har konstrueret et ikke-mekanisk system.
7. Det er i denne forbindelse interessant at notere sig, at Claude Shannon i nogle sammenhænge identificerer redundansen med de systembestemte dele (altså regelstrukturen) i en meddelelse.
8. Det skal bemærkes, at jeg går ud fra en mere omfattende redegørelse i Finnemann (1994).
9. Her kan man igen skelne mellem forekomst af et mønster, der blot optræder som støj, og den første udnyttelse af mønstret som symbolsk legitim form.

UDVALGT LITTERATUR

For mere omfattende litteraturliste henvises til Finnemann (1994).

- COURTES, J. & GREIMAS, J. A. (1987): *Semiotik. Sprogteoretisk Ordbog* Aarhus: Aarhus Universitetsforlag. (Fransk orig: (1979) *Semiotique. Dictionnaire raisonne' de la theorie*. Paris: Hachette).
- FINNEMANN, N. O. (1994): *Tanke, Sprog og Maskine*. København: Akademisk Forlag.
- KØPPE, S. (1993, 2. udg.): *Virkelighedens Niveauer*. København: Gyldendal.
- MAZLISH, B. (1966): The Fourth Discontinuity. Reprint in Bannon, L. & Pylyshyn, Z. (1989) *Perspectives on the Computer Revolution*. Norwood: Ablex.
- SHANNON, C. & WEAVER, W. (1949): *The Mathematical Theory of Communication*. Reprint (1969). New York. Raven Press.
- SIMON, H. & NEWELL, A. (1976): Computer Science as Empirical Inquiry: Symbols and Search. Reprint in Bannon, L. & Pylyshyn, Z. (1989) *Perspectives on the Computer Revolution*. Norwood: Ablex.
- TURING, A. (1950): Computing Machinery and Intelligence. *Mind* vol. 59, No. 236, Oct. Reprint in Bannon, L. & Pylyshyn, Z. (1989) *Perspectives on the Computer Revolution*. Norwood: Ablex.
- WITT HANSEN, J. (1985): *Filosofi*. København: Gyldendal.