Intelligens og bevidsthed i digitale kredsløb

SKREVET AF CAND MAG. OLE FREHR

INDHOLDSFORTEGNELSE

AFSNIT **SIDE**

Indledning

Computerudviklingen raser der ud af. Konkurrencen i branchen er hård, og det nyeste udstyr er forældet i morgen. Som tiden går bliver maskinerne hele tiden hurtigere og bedre, og derfor er der et spørgsmål, der i den forbindelse bliver mere og mere interessant: Er computere intelligente og besidder de bevidsthed?
Dette essay vil undersøge disse to aspekter, ved at fremdrage forskellige filosoffer der har udtalt sig om samme, blandt andet ved forskellige tankeeksperimenter. Jeg vil forsøge at forholde mig til det principielle i synspunkterne - netop fordi udviklingen går så hurtigt - og håber derfor, at essayet også kan læses i fremtiden, uden at være synderlig forældet.

I det følgende vil betegnelsen 'bevidste systemer' optræde. I mangel af bedre bruger jeg dette patologiske udtryk, om alt og alle der kan have eller som tilskrives bevidsthed: mennesker, dyr, maskiner og computere.

Computere og intelligens

EDB-maskinernes evne til hurtigt at bearbejde store mængder tal, har hjulpet menneskerne på mange områder: på arbejdet, i hjemmet og i fritiden. I de senere år er der sket stor fremgang inden for forskningen i kunstig intelligens, hvilket ofte forkortes AI (artificial intelligence). Her forsøger man blandt andet at fremstille såkaldte 'ekspert systemer', der efterligner menneskelig arbejds- og tankegang indenfor et givent område. Kompleksiteten af disse systemer kan variere meget, og skellet mellem hvad der virkeligt definerer et ekspert system, og hvad der bare er en række regler, er uklart.
Inden for AI er der to hovedstrømninger: der er den 'klassiske skole' og den 'moderne skole'. Disse skoler prøver hver for sig at konstruere intelligente systemer, men de metoder de anvender til at nå dette mål, er forskellige.
Den klassiske skole forsøger at konstruere intelligente maskiner ved en regel-metode: alle regler, love og sammenhænge der er relevante for opgaven der skal udføres, indkodes i maskinen. Vores computere virker i dag på denne måde: reglerne er programmet til maskinen, der således danner opskriften på, hvordan det givne arbejde skal udføres. At afvige fra reglerne er ikke muligt, og derfor giver denne måde at opbygge systemer på ufleksible systemer.
Neurale netværk er en anden metode, som den moderne skole benytter sig af, til at opbygge intelligente systemer. Ved at aflure den menneskelige hjernes karakter og faconen den opererer på - med sine neuroner og sammenkoblingerne mellem disse -

og bagefter efterligne dette elektronisk, prøver man at komme tættere på menneskelig intelligens. Neurale netværk har den fordel at de i en vis forstand kan 'lære', da de kan bearbejde data som de aldrig har 'set' før. Til trods for at det er en ganske anden tankegang der ligger bag dette paradigme, knytter den sig stadigvæk meget til den klassiske skole. Blandt andet bruges computere fra den traditionelle skole, til at danne neurale netværk.

Kan disse to metoder frembringe systemer der kan betegnes intelligente? Dette spørgsmål afhænger af hvilken form for intelligens der tales om. Tvivlen består ikke i hvorvidt disse systemer er smarte. Det er de afgjort. Vi lader computere arbejde som 'talknusere' for os, eller vi kan få den til hurtigt at gennemsøge store mængder data, hvilket letter vores arbejde.

Intelligens kan opdeles i forskellige kategorier[1]; som intelligens for det logisk-matematiske, intelligens for sprog, intelligens for det kropslige-kinæstetiske, o.a. Når computeren omtales som intelligent uden at dette er specificeret, gælder dette naturligvis kun indenfor områder det er muligt for den at besidde. Computeren kan (endnu) ikke siges at have kinæstetisk intelligens (fornemmelser for muskelbevægelser).

Der er mange der mener, at computere ikke er intelligente. Alle der anvender computere ved hvor omhyggelig og overbærende det er nødvendigt at være, for at kommunikationen mellem maskine og menneske skal foregå smertefrit. Selv det mindste fejltryk, kan foresage gravende ulykker. Personer der mener, at maskiner ikke er intelligente, havner ofte i en modsigelse: de vil gerne vedgå at mennesker der kan gange to fircifrede tal sammen i hovedet på kort tid, er intelligente. Men hvorfor kan maskiner der hurtigere kan beregne det samme, eller mere komplicerede regnestykker, da ikke kaldes intelligente? 'Fordi de er maskiner' vil nogle svare, men dette er cirkelargumentation; det var netop dette, der var antagelsen. Hvorfor kan arbejde udført af mennesker kaldes intelligent, mens det samme arbejde udført af en maskine, ikke kaldes intelligent ?

Svaret hænger sammen med den ramme computeren arbejder indenfor: en lukket verden. En lukket verden er kendetegnet ved følgende faktorer:

> "a. der er mulighed for fuld information om sættet af aktuelle og potentielle fænomener og relationer.
> b. der er mulighed for kendskab til alle de regler og lovmæssigheder, der behersker denne verden.
> c. den anvendte formalisme (logik eller regler) passer perfekt til alle relationer og griber enhver mulig nuance.
> d. der er åbenlys og entydig sammenhæng mellem mål og midler;

[1] Se Mogens Hansen: "Intelligens - om hjernen, tænkningen og erkendelsen", pp.55-81

denne verden er hundrede procent rationel."[2]

Muligheden i at beregne alle relationer og begivenheder i et system, er kun muligt hvis området der arbejdes indenfor er meget afgrænset. Så længe kommunikationen foregår inden for dette snævre felt, kan det fra brugerens synspunkt se ud, som computeren er intelligent. For eksempel en maskine med hukommelse og en algoritme til at genkende mønstre i spørgsmål, kan givetvis kommunikere i dagligsproget om bevægelsen og placeringen af farvede kasser i et virtuelt rum[3]. Kaldes en sådan maskine intelligent er konsekvensen, at maskiner af enhver art, der udfører arbejde efter en givet forskrift, kan kaldes intelligente: d.v.s. alt fra dampmaskiner til pengeautomater.
Hvad der i det hele taget skal forstås ved intelligens i computersammenhæng, er et spørgsmål der i tidens løb er blevet besvaret på mange forskellige måder. En af de kendteste forsøg, stammer fra Turing.

Turingtesten for intelligens

Turings artikel fra 1950 giver en testopstilling på, hvordan det er muligt at afgøre om computere er intelligente. Turing opdeler ikke intelligens i kategorier som ovenfor, men anvender begrebet i en mindre formel betydning.
Det er muligt at lave en Gallup-undersøgelse, for at få et statistisk materiale der kan belyse, hvad der menes med intelligens i computersammenhænge. Men dette mener Turing er absurd, og omformulerer spørgsmålet 'Kan computere tænke ?' til 'Kan computere simulere menneskelige svar på alle tænkelige spørgsmål ?'[4] Hvis man svarer bekræftende på det sidste spørgsmål, vil man samtidig svare bekræftende på det første spørgsmål.
Ved at lade en udspørger anvende tastatur og skærm til at kommunikere med, skal hun finde ud af, om det er en computer eller et menneske, der giver svarene på hendes spørgsmål, fra et andet lokale. De to personer i denne 'quiz', spiller faktisk sammen mod computeren, så udspørgerens gode spørgsmål og svarerens gode svar, skal få computeren til at fremstå dum, så den let kan identificeres. Computeren skal også yde sit yderste for at dens svar ligner menneskelige svar, men i nogle tilfælde, for eksempel

---

[2]Fra Ole Fogh Kirkeby: "Ægte intelligens - om bevidsthedens program", p. 97

[3]Se Niels Ole Bernsen & Ib Ulbæk:"Naturlig og kunstig intelligens", p. 131-133

[4]Det sidste spørgsmål er min transskription af formuleringen i Turings artikel.

spørgsmål hvor der indgår komplicerede regnestykker, skal den gøre sig dummere end den er, og ikke give svaret med det samme selvom den kan, da den ellers hurtigt afslører sig selv.
Turings tese er altså, at hvis udspørgeren ikke kan afgøre om det er computeren eller mennesket der svarer, må computeren siges at være intelligent.
Her mener jeg at Turing begår en fejlslutning. I hvert fald er det usikkert, hvilken form for intelligens han er ude på at blotlægge - det kan ikke være menneskelig intelligens. Dette ses hvis forsøget udføres igen, men denne gang erstattes computeren i forsøget med Poul Thomsen (fra TV-udsendelsen 'Dus med dyrene') og mennesket med en abe. Nogle aber kan lære at 'tale' via staveplade, og kan forstå ca. 1000 ord. Lad nu aben gøre sit ypperste for at svare og lad Poul Thomsen gøre sig dummere end han er og forsøge at tale 'abesprog'. Hvis udspørgeren ikke kan skelne imellem de svar der kommer retur, kan vi da konkludere at Poul Thomsen har samme intelligens som en abe? Nej, selvfølgelig ikke, og på samme måde er det heller ikke menneskelig intelligens computeren har.
Er computeren da mere eller mindre intelligent end mennesket? Svaret på dette spørgsmål afhænger af computerens forståelse, og dette aspekt formår Turingtesten ikke at vise - hvilket jeg vil forklare i det følgende.

Intelligens og forståelse

Maskiner der udelukkende følger regler slavisk, som i det klassiske paradigme, vil under alle omstændigheder være mennesket inferiør. Disse maskiner bliver programmeret til at arbejde ud fra en given plan eller program med bestemt mål, uden at betvivle de informationer den modtager. Maskinen har/kan ikke have forståelse af de data den beregner, da arbejdet den udfører udelukkende er symbolmanipulation i en formel konstruktion. At vedgå at disse maskiner har 'en form for intelligens' er ikke selvmodsigende: de udfører effektivt arbejde i et lukket system; altså systemer kendetegnet ved, at de eneste kendsgerninger der karakteriserer systemet, findes i dets database. Så længe der kun arbejdes indenfor denne ofte snævre, men veldefinerede ramme som lukkede systemer giver, kan computeren effektfuldt beregne de konsekvenser der med deterministisk sikkerhed indtræffer, og fremdrage karakteristiske mønstre og symmetrier på samme måde som mennesker ville gøre det.
Men systemet mangler forståelse. Forståelse fremkommer først, når det er muligt at forklare, tydeliggøre, eksemplificere og belyse begreber i sammenhænge hvori de

indgår, og når der foreligger en mulighed for at forklare hvordan og hvorfor, disse sammenhænge er interessante. Dette formår computeren ikke. Når symbolmanipulationen i disse systemer går så hurtigt, er det fordi maskinen ikke behøver at have forstået meningen med de symboler den behandler. Reglerne som systemet anvender, angiver hvad der er muligt, hvad der skal gøres, hvordan og hvornår det skal gøres. På intet tidspunkt optræder der et 'hvorfor ?'. Systemet indeholder ingen nysgerrighed om, hvorfor dette eller hint skal beregnes, da der ikke er forståelse for indholdet i symbolerne. Den interesse og det engagement maskinen synes at imødegå enhver opgave med, er kunstig frembragt af mennesker. Alt arbejdet er ren formel symbolmanipulation udført af rå regnekraft. Maskiner der følger det klassiske paradigme arbejder udelukkende efter formelle regler i lukkede systemer, besidder derfor en form for intelligens (ville nogle hævde), men de mangler i hvert fald forståelse.

Har computeren til gengæld fuld forståelse af alle spørgsmålene den modtager i Turingtesten, forstår betydningen af sætningen både semantisk og syntaktisk, ved at ordene optræder med forskellige konnotationer der afhænger af situationen, kan skelne mellem ironi og oprigtighed når dette er tydeligt, med andre ord: formår at indgå i dialog i dagligsproget, er computeren intelligent. For behersker man dagligsproget, kan man også digte, rime, lyve, lovprise, hovere, smigre og benytte alle dagligsprogets facetter til at udtrykke håb, frygt, begær; og med et minimum af hukommelse, er der ingen der kan skelne mellem et menneske og en computer i en dialog. Kan computeren endog genkalde viden og drage konklusioner hurtigere end mennesket, vil sådan en computer være mere intelligent end mennesket.

Computere med en sådan evne, er i dag meget langt fra at kunne virkeliggøres, og de vil stadigvæk være science-fiction emner langt ud i fremtiden. Selv om andre dyr kan kommunikere sammen på højt plan, og nogle er i stand til at lære en del af dagligsproget, er det kun mennesket der har den fulde forståelse af dagligsproget.

Turingtesten viser ikke om computeren har denne forståelse for dagligsproget, da udspørgeren ud fra retursvarene ikke kan bedømme om svareren har en forståelse af spørgsmålet eller ej. Computeren kunne være en af den klassiske skoles nyeste påfund: den kunne udelukkende følge regler der var passende komplicerede og derved fingere et intelligent svar uden at have forståelse, men dette gør den ikke intelligent. Derfor viser Turingtesten ikke om computeren er intelligent, og fejlen består kort sagt i, at det spørgsmål Turing i virkeligheden ønsker at besvare omskrives. Men det omskrevne spørgsmål er ikke ækvivalent med det første. Bare fordi computere simulerer menneskelige svar, kan de ikke nødvendigvis tænke selv.

Searles kinesiske værelse

Filosoffen John Searle har også beskæftiget sig med emner der vedrører maskiner, intelligens og bevidsthed. Han stiller sig skeptisk overfor synspunktet, at computere skulle kunne tænke selv. Med udgangspunkt i Turings test, når han til denne konklusion.[5]

Antag at der eksisterer en computer, der kan kommunikere på det kinesiske sprog. Stilles maskinen spørgsmål på kinesisk, er den i stand til svare som personer, der kan det kinesiske sprog flydende. Ifølge Turingtesten udfører denne maskine intelligent adfærd. Den må derfor have forståelse for det kinesiske sprog og være intelligent. Kan denne afgørelse være korrekt, spørger Searle? Svaret er nej, og han argumenterer på denne måde:

Betragt følgende påstande:

1) Syntaks er ikke tilstrækkelig for semantik.

Påstanden er et resultat af vores opfattelse af de to begreber. Den beskriver den distinktionen imellem dem: det vi fortolker som udelukkende rent formelt, og det der har indhold.

2) Computerprogrammer er udelukkende defineret ved deres formelle, eller syntaktiske, struktur.

Dette er essentielt for vores opfattelse af computerens måde at fungere på, og, ifølge Searle, sandt per definition. Operationerne der udføres, frembringes med abstrakte symboler, der manipuleres efter nærmere bestemte regler. Unødvendigheden i at betragte det semantiske aspekt, gør netop computerne så effektive.

3) Bevidstheden har et mentalt indhold; specielt har den et semantisk indhold.

Når jeg tænker på noget, har det, jeg tænker på et indhold. Vores tanker er ikke udelukkende symboler defineret formelt eller syntaktisk, det jeg tænker på har en mening.

D.v.s mine tanker har visse mentale indhold, udover de formelle egenskaber de måtte have.

Konklusionen af disse tre påstande er, at intet computerprogram i sig selv, er tilstrækkelig til at systemet kan få bevidsthed. Searle understreger at der ikke er tale om at senere udvikling inden for computerteknologi skulle kunne ændre denne konklusion. Den er principiel, og kan ikke omstødes på grund at større eller hurtigere regnekapaci-

[5]Argumentationen er hentet fra John Searle: "Minds, brains & science", pp. 28-41

tet.

Pointen, at computeren ikke har bevidsthed, understreges ved hjælp af et tankeeksperiment, Searle har konstrueret: det kinesiske værelse. Lad os gentage forsøget med maskinen der kan kommunikere på det kinesiske skriftsprog, men denne gang udskiftes computeren med Searle selv. Searle, der hverken kan læse eller skrive med kinesiske tegn, lukkes inde i et rum fyldt med skuffer der indeholder kinesiske tegn. Rummet er helt lukket og den eneste måde at kommunikere med andre udenfor rummet på, er ved at skubbe små sedler med kinesiske tegn ind og ud under døren. Når der glider en seddel ind til Searle, består hans arbejde i at slå op i en stor 'ordbog' der også er i lokalet. Ved - et for et - at 'læse' tegnene på sedlen, beskriver ordbogen hvilket tegn han skal tage fra skuffen. 'Står det sådan-et-krussedulletegn, tag da sådan-et-krussedulletegn.' De tegn han har samlet sætter han i rækkefølge på en ny seddel som han skubber ud under døren og ud af rummet.

For personer udenfor rummet kan det virkelig se ud som om, at 'rummet kan forstå kinesisk', da det kan kommunikere via små sedler. Men det paradoksale er, at Searle inde i værelset stadigvæk ikke forstår kinesisk. For ham har det hele tiden været ren manipulation med symboler, der ikke har betydning for ham.

> "But this feature of programs, that they are defined purely formally or syntactically, is fatal to the view that mental processes and program processes are identical. And the reason can be stated quite simply. There is more to having a mind than having formal or syntactical processes. ... The reason that no computer program can ever be a mind is simply that a computer program is only syntactical, and minds are more than syntactical. Minds are semantical, in the sense that they have more than a formal structure, they have a content."[6]

Ligesom Searle der på intet tidspunkt har forstået kinesisk, har computeren heller ikke forståelse, og at tilskrive den bevidsthed, er forfejlet. Bare fordi maskinen kan sammensætte symbolerne, så det ser ud som om den er bevidst, behøver den ikke at være det. For at være bevidst, er der brug for en form for interpretation, så der kan tilskrives en mening til symbolerne der indgår i systemet.

Spøgelset i det kinesiske værelse

Er Searles argumentation helt holdbar? Har han virkelig fundet et principielt

[6] John Searle: "Minds, brains & science", p. 31

resultat, der ikke kan ophæves af videre forskning? Lad os betragte argumentet nærmere.

De tre påstande - og derved konklusionen - synes at have en vis gyldighed. Semantik er afgjort mere end syntaks: der er med andre ord indholdet af begreberne til forskel. Så længe computere udelukkende arbejder på det syntaktiske niveau, vil den ikke være i stand til at nå det semantiske niveau. Men på hvilket niveau befinder vi os, i tankeeksperimentet om det kinesiske værelse?

Påstanden er, at både computeren og det kinesiske værelse er i stand til at kommunikere på flydende kinesisk. Dog kun via skrift, hvilket kan være en begrænsning i forhold til det talte sprog, men må alligevel siges at være tegn på stor intelligens. Evnen til at kommunikere på planer hvor der ikke er grænser for hvad der kan diskuteres, er der endnu ikke andre end mennesker, hverken dyr eller maskiner, der besidder. I tankeeksperimentet tillader vi os at antage dette; måske kan det lade sig gøre i fremtiden. Men sprog er ikke bare strenge af symboler, der overføres fra én bærer af sproget til en anden bærer. Begreberne har et meningsmæssigt indhold. Syntaks er ikke fyldestgørende i en kommunikation i sproget: kommunikation i eksperimentet foregår derfor på et semantisk niveau.

Betragt nu følgende ræsonnement:

At have ben er en nødvendig betingelse for at kunne gå. Man kan muligvis tale om personer eller dyr der 'går på hænder', men det betyder blot, at personen anvender nogle andre af sine lemmer som ben. For at kunne gå, er det nødvendigt med nogle lemmer af en art, der kan bevæge sig på en bestemt måde, således at der fremkommer bevægelse fra position A til position B, og disse lemmer kaldes ben.

Min computer har ingen ben. Den er en stationær maskine, der kun bevæger sig i det tilfælde nogen tager den op og flytter den et andet sted hen. Nu laver jeg et tankeeksperiment, en tænkt situation, hvor computeren går, men uden at den har nogen ben. Denne situation synes meget mere end akavet. Hvordan skulle den overhovedet kunne lade sig gøre, at computeren bevæger sig ved at gå uden at have nogen ben? Det er i modstrid med, at ben er en nødvendig betingelse for at kunne gå.

Dette må betegnes som en paradoksal argumentation, når et system sættes til at arbejde i en situation, hvor betingelserne for situationen ikke er til stede. Uheldigvis begår Searle den samme fejl:

Semantik er, som ovenfor beskrevet, nødvendig for sprog. Searles påstand er, at computere ikke kan have forståelse af et udtryks semantiske indhold. Som computere i dag er opbygget, accepteres påstanden praktisk talt overalt. Hvad sker der nu, hvis vi

i et tankeeksperiment anbringer computeren i en situation, hvor den anvender sproget, men ikke benytter nogen form for semantik? Dette er også paradoksalt, og argumentationen er ugyldig.

Siden værelset kan kommunikere, må der være noget i værelset, eller værelset som helhed, der forstår kinesisk, og kan anvende sproget *med den tilhørende semantik*.

> "They argue that it is the whole system, ... taken as a totality, that understands Chinese. But this is subject to exaxtly the same objection I made before. There is no way that the system can get from syntax to the semantics. I, as the central processing unit have no way of figuring out what any of these symbols means; but then neither does the whole system."[7]

Searle begår altså den fejl, at han lader computeren - eller det kinesiske værelse - kommunikere i sprog, hvilket er umuligt for den, og derfor er resultatet en antinomi. Der er i tankeeksperimentet sket en utilladelig sammenblanding mellem det semantiske og det syntaktiske niveau.

Fastholdes synspunktet, at værelset virkeligt kan kommunikere på kinesisk, hvad er det da, der forstår kinesisk? For mig er der en ting der springer i øjnene: ordbogen. Der knytter sig flere uløste spørgsmål til denne bog: Hvem har lavet den? Hvordan er den opbygget? Hvordan slår man op i den? Kan den ændre sig selv når begreberne ændrer betydning? Searle har givetvis valgt det kinesiske sprog, fordi tidligere henviste hvert tegn i det kinesiske skriftsprog til et bestemt begreb på en af de kinesiske dialekter. Tankeeksperimentets umulighed vil være tydeligere, hvis hvert tegn ikke var et begreb, men en stavelse, som det er tilfældet i dag. Antallet af kombinationer som stavelser kan indgå i, er uendelig stort. Som et mindstekrav er det nødvendigt med en uendelig tyk bog, der skal kompensere for dette. Men de konsekvenser der da drages af tankeeksperimentet, kan ikke overføres til vores situation.

Selvom der anvendes tegn som hver især har en bestemt betydning, synes tankeeksperimentet ikke at kunne lade sig gøre. Hvordan skal man kunne finde det tilhørende 'svar-tegn', når man ser et tegn? Tegnene, eller begreberne for så vidt at dette er det samme, ændrer betydning afhængig af konteksten hvori den indgår. Hvordan ordbogen afgør betydningen af disse polysemer, må anses som meget kontroversielt.

Hvis det er muligt med en sådan ordbog, hvordan slår man så op i den? Tegnene kan muligvis klassificeres efter udseende eller andet, men der kan gå uendelig lang tid før det rigtige tegn er fundet, da bogen skal indeholde alle tegn for alle begreber sproget

[7] John Searle: "Minds, brains & science", p. 34

indeholder, i alle de forskellige former de kan optræde.
Kan ordbogen ikke ændre sig selv (hvordan skulle den kunne gøre det?), kan den ikke forstå begreber der er opstået senere end bogen er skrevet. Den udvikling som sproget hele tiden er i, hvor nye begreber dannes, og begreberne ændrer betydning, som også sker kontinuerligt, kan der ikke tages højde for. Sproget bliver fastlåst i en gammel kancellistil eller lign, og vil udelukkende være interessant at studere for lingvister, etnografer og andre videnskabsfolk, som et kuriosum fra fortiden
Sidst men ikke mindst: hvem har skrevet bogen? Den eller de personer der har tabelleret sproget på denne måde, kan ikke unddrage sig, at være underlagt en vis form for historicitet. Ellers skal ordbogen være guddommelig inspireret, men dette skridt er for drastisk for at redde tankeeksperimentets holdbarhed. Ordbogen optræder som et spøgelse, der ikke er til at få hold på. I eksperimentet udfører den mirakuløst meget intelligent arbejde, men ved nærmere studering synes den at være subtil, udefinerlig og unddrager sig enhver form for undersøgelse.
Lige meget om tegnene der anvendes i det kinesiske værelse, står for et begreb eller en stavelse i et ord, er tankeeksperimentet ikke godt til at understrege den pointe Searle har analyseret sig frem til i den 'udvidede syllogisme': computere der ikke kan benytte sig af det semantiske niveau, kan ikke have bevidsthed. Konklusionen er sand, men har ikke den principielle karakter, som Searle gerne vil tillægge den. Fordi computere i dag udelukkende benytter sig af syntaks, kan der ikke sluttes, at dette vil være tilfældet altid.

Klassisk AI forskning

Den klassiske AI forskning er et projekt, der på forhånd er født til at mislykkes. Problemet ligger blandt andet i umuligheden i at nedskrive en regl for enhver tænkelig situation. Selv ukontroversielle og (for mennesker) letoverskuelige omstændigheder, vil uundgåeligt skulle implementeres. Et ekspertsystem der skulle kunne gå i banken for mig og indløse en check, vil bryde sammen den dag det regner, med mindre programmøren på forhånd eksplicit har påpeget nødvendigheden af paraply i regnvejr. Man har tidligere forsøgt at udfærdige beskrivelser af hvordan forskellige situationer i hverdagen opstår og forløber. Til dette har R. Schank introduceret begrebet 'script'[8]. Scriptet kan opfattes som en opskrift eller et manuskript. Der er et script for at gå på restaurant, et for at købe ind, o.s.v. Restaurant-scriptet beskriver hvilke personer der er

[8]Se for eks. Ole Fogh Kirkeby: "Ekspertsystemer og kunstig intelligens", p. 49-50

involveret (kokken, tjeneren, kunde), årsagen til scriptet (få stillet sin sult og forkæle kæresten, for eks.) og handlingen i forskellige scener (går ind, får tildelt et bord, går til bordet, sætter sig, beder om vinkortet, o.s.v.). Argumentet for denne måde at opdele verden på er, at tænkning ikke kun består af fri association. Intelligent adfærd kræver indlærte regler og struktur.
Beskrivelsen giver fornemmelsen af hvad det vil sige at 'gå på restaurant', og nogle typiske kendetegn. Men langt fra altid følges sådan en plan: måske ønsker gæsten at vaske hænder inden han sætter sig, vinkortet er forsvundet, alle borde er optagede, eller andet. På vej hen til det anviste bord møder man måske en gammel bekendt. I 'restaurant scriptet' udspiller sig så et 'møde-med-gammel- ven script'. Beslutter alle at gå hjem til vennen og spise i stedet for, opstår et nyt script. Hvad skal dette script hedde? Er tjeneren stadig involveret i dette sidste script? Hvilke scripts er afsluttede og hvilke er i gang? Alle disse spørgsmål er uløste.
Det er sikkert muligt at elaborere på synspunktet, men spørgsmålet er, hvor givtigt det er at opdele verden i 'kasser' på denne måde. Programmeres maskinen til at ræsonnere i 'kassetænkning' resulterer det i en klodset og akavet 'kasseadfærd'.

Dele af menneskelige handlinger skyldes en art refleks: ringer min telefon tager jeg røret og svarer. Jeg undrer mig ikke først hvorfra ringetonen kommer eller om jeg skulle have fået tinnitus, tænker ikke på om telefonen uden grund pludselig er væk, at den forsvinder hvis jeg rører ved den, eller lignende. D.v.s. jeg tager meget for givet, selvom tingene kan have ændret sig med større eller mindre sandsynlighed. Muligheden for at min kæreste har taget telefonen med hen i telebutikken fordi den var gået i stykker, men havde glemt at oplyse mig om dette, eksisterer. I den forbindelse står AI forskere overfor en vanskelighed, kaldet rammeproblemet[9]: Det er nødvendigt at fortælle maskinen hvilke parametre der er konstante og hvilke der er ændret, så computeren forstår, at telefoner for eks. ikke bare forsvinder ud i det blå af sig selv, i modsætning til regnbuer. Refleksen jeg udfører når jeg handler delvist ubevidst, kan sikkert omskrives til en regl, men pointen er, at antallet af variable vokser hurtigt. Det er muligt at beskrive et rum der indeholder et bord, en stol og en blomst ret præcist, men for systemer med et minimum af kompleksitet, er antallet at variable uoverskueligt stort. Selv med en stor stab af programmører med tid og penge nok til at implementere alle egenskaber ved alle ting, foreligger en principiel og indlysende umulighed: det er ikke muligt at forudse uforudsete hændelser. Med mindre der er tale om et lukket system,

[9]Se Ole Fogh Kirkeby og Torben Tambo: "Guds ur", p. 69

er man forhindret i at programmere sig ud af problemet. Hver gang en fejl findes, kan en regl implementeres og afskære fejlen i at forekomme igen og således *bagefter* efterrationalisere, men udefra kommende faktorer kan uventet spille en afgørende rolle, uden at vi har mulighed for *på forhånd* at tage højde for dette.
Dertil kommer hvad der matematisk er kendt som 'trelegeme-problemet': vi er endnu ikke er i stand til matematisk at beskrive tre eller flere fænomeners indbyrdes påvirkning på hinanden. Så længe der er tale om borde og stole i et stillestående rum, er problemet overskueligt. Oftest er det dog mere interessant, at simulere omgivelser af noget større kompleksitet.
Derfor er klassisk AI forskning uden fremtid. I hvert fald som et projekt der skal kunne stå alene, som basis for en maskine der skal kunne gebærde sig konstruktivt i relevante omgivelser, som vi mennesker gør.

At computeren engang kommer til at kunne indgå i dialog i dagligsproget, mener jeg er en utopi. (Det ville naturligvis kræve at samtalen foregår via tale og ikke skrift, da betoningen af ord ofte har afgørende betydning i kommunikationen, men det er et teknisk problem, der muligvis kan løses allerede i dag.) Trods min skepsis angående fremtidige computeres formåen, er jeg fortrøstningsfuld. Hvis computeren nogensinde får evnen til at kommunikere på lige fod med mennesket, mener jeg det er en fantastisk bedrift, og hvis ikke, kan vi alligevel prise den højt, betragte den som et fantastisk værktøj, og anvende den hvor det er muligt.

Menneskets mystiske genialitet

Mennesket er som helhed unikt. Måden vi behandler information på er enestående, og i dag ikke forstået til bunds. Der er mindst to områder af informationsbehandling, hvor AI forskere gerne vil aflure os kunsten, da en sådan opdagelse vil føre til store landvindinger indenfor AI teknologi.
Kort fortalt drejer det sig om, for det første, hvordan mennesket er i stand til at glemme eller frasortere allerede eksisterende information. For det andet hvordan vi får eller slutter os til information der ikke er direkte givet.
Mennesker besidder den færdighed, at vi, afhængig af situationen, let sorterer informationen vi modtager. Vi har en evne som AI forskning har svært ved at tage højde for: evnen til at skelne det væsentlige fra det uvæsentlige. Fortalte nogen os, at de i går aftes gik ind i en bygning, fik tildelt et bord, satte sig, bestilte noget mad og vin, ville vi undre os, og spørge dem om hvorfor de ikke bare fortalte os, at de gik på restaurant.

Det er ikke nødvendigt for os at vide, at det var i en bygning restauranten var placeret: informationen var redundant. Man kan sige at vi sier eller destillerer det centrale fra, uden at vide hvordan vi gør det. Hvilken viden i en bestemt situation der anses for vigtig og hvilken der ikke gør, er ofte umulig at afgøre på forhånd.
Sker der ikke en sortering i mængden af information, bliver resultatet en videnseksplosion. Hele tiden sorterer vi i informationsmassen vi modtager, og derved undgår vi, at bunker af ubrugelig information ikke hober sig op. Måden vi sorterer på, ligner en variant af rammeproblemet: maskinen skal gerne have et eller flere kriterier for, hvilke dele af den eksisterende information den kan se bort fra. Kan vi ikke beskrive disse kriterier eller formulere en proces der kan efterligne denne handlemåde, er vi nød til at lade maskinen lagre al den information den modtager. Informationen kan indeholde vigtige oplysninger, som mennesker i den givne situation ville tage for givet (common sense viden), men som computeren ikke har mulighed for at gennemskue.

> "En stor del af common sense viden og ræsonneren drejer sig om kausalrelationer og rumlige og tidslige relationer mellem ting, processer og begivenheder. ... Problemet med at skabe kunstig intelligens er, at næsten vilkårlige dele af common sense viden kan være relevant og nødvendig næsten overalt under problemløsning"[10]

Videnskaben står overfor en stor opgave, når de skal forsøge at finde en teori der skal forklare, hvordan vi frasorterer information.

Som ovenfor beskrevet er der altid noget vi tager for givet; både når vi selv ræsonnerer og i kommunikation med andre. Begreberne vi anvender kan altid tydeliggøres, vi kan forklare videre omstændigheder, sætte begreberne i relation til andre begreber eller situationer. Men i vores omgang med andre mennesker, kommunikerer vi på et 'overfladisk plan'. Tag udtalelsen: 'i aftes vi gik på restaurant'. Sætningen er fuldt forståelig og nødvendiggør ingen yderligere eksplicitering: det er ikke nødvendigt at definere, om 'i aftes' var klokken 18 eller 20.30, om restauranten var en restaurant eller en bistro, om vi virkelig gik derhen eller tog bussen. Man behøver ikke for sit 'indre øje' at gennemgå 'Restaurant scriptet' for at se om der hersker overensstemmelse med det sagte og scriptet. Sætningen er også fuldt forståelig, selvom der er flere trin i scriptet der er sprunget over: gå ind af døren, finde et bord, bestille. Dette gør os i stand til at kommunikere sammen, uden at skulle definere i dybden, selv med ord der er meget kontekstafhængige.

[10]Niels Ole Bernsen & Ib Ulbæk: "Naturlig og kunstig intelligens", p. 160

Muligvis er den viden vi afleder fra informationerne ineksakt. Tales der om en restaurant, slutter vi automatisk, at restauranten også har borde og stole, selvom dette ikke behøvede at være tilfældet, i en alternativ asiatisk og eksotisk restauration. Begivenheder som disse, har fået filosoffer til at tale om en typeteori.[11] Teorien går ud på, at vi organiserer og klassificerer ved hjælp af prototypiske fænomener, det vil sige begreber rubriceres ved at anvende de mest karakteristiske træk. For eksempel er en prototypisk fisk et forholdsvist lille dyr med gæller, finner, hale, skæl i stedet for hud, og ingen ben eller arme. Herved har man givet en forklaring på hvorfor vi bruger begreberne på det abstraktionsniveau som vi gør: det giver os en metode til at klassificere begreber i verden.
Teorien forklarer dog ikke hvordan de prototypiske begreber opstår. Er det en form for eidetisk variation, eller anamnese? Så længe dette spørgsmål ikke er besvaret, er teorien, i kognitionsforskningens øjemed, ligegyldig, da det ikke giver noget svar på, hvordan teorien skal implementeres på en computer.

Konklusionen er, at det er et uomgængeligt punkt at fremsætte nogle teorier, der kan forklare vores tilsyneladende lemfældige omgang med informationsmassen, og samtidig kan implementeres på computer. Er det muligt i fremtiden at løse dette problem, tror jeg at videnskaben vil være nået et langt stykke af vejen mod målet: den kunstige intelligens.

Chalmers argument for bevidsthed

Matematikeren og filosoffen David Chalmers har skrevet om bevidsthed. Han har bl. a indført en distinktion imellem 'de lette problemer' og 'det svære problem' i forbindelse med bevidsthedsmæssige fænomener[12]. De lette problemer er alle de vanskeligheder, som vi sandsynligvis kan løse ved hjælp af metoder fra den kognitive videnskab. Området dækker overvejelser som evnen til at kategorisere omgivelsernes stimuli og informationens integration. Chalmers indrømmer selv at 'lette problemer' i den sammenhæng er et relativt begreb, og at der kan gå lang tid, før alle detaljer er beskrevet.
Det svære problem er det problem som endnu ingen har løst. Hver gang nogen har

[11] Se Niels Ole Bernsen & Ib Ulbæk: "Naturlig og kunstig intelligens", pp. 64-68

[12] Denne skelnen er beskrevet i D Chalmers "Facing up to the problem of conciousness" pp. 1-2

påpeget visse egenskaber ved hjernen, og angiver dem som årsagen til bevidstheden, kan Chalmers stille spørgsmålet: 'Hvorfor er det netop disse egenskaber der konstituerer bevidstheden, og hvordan gør de det ?' Spørgsmålet er oplagt, for eksempel når Crick og Koch måler oscillerende svingninger på 35-75 Hz i den del af hjernen der hedder cerebral cortex, og hævder at de er bevidsthedens grundlag. Det svære problem består med andre ord i, hvordan bevidsthed overhovedet kan opstå af fysiske kvaliteter.

Chalmers har også argumenteret for, at komplicerede computere vil have bevidsthed[13]. Argumentet udmunder i det han kalder organisatorisk invarians, hvilket vil sige, at systemer der er organiseret på samme måde som den menneskelige hjerne, vil have de samme egenskaber som hjernen, d.v.s. også besidde bevidsthed. Synspunktet har den konsekvens, at hvis man opbygger et tilstrækkelig kompliceret system der efterligner hjernens måde at arbejde på, uanset hvilket materiale der anvendes til dette - man kunne lade Indiens befolkning 'kopiere' hjernen, sådan at enhver neuron blev erstattet med en inder, og kommunikationen imellem dem kunne foregå med signalflag - vil der opstå bevidsthed. I dette tilfælde vil Indiens befolkning da have en 'fælles' bevidsthed. Som følge heraf, vil Indiens befolkning i dette store samarbejde som helhed, få kvaliteter som bevidstheden kan have - som had, glæde og smerte - til at opstå.

Argumentet for at computere besidder bevidsthed består af et tankeeksperiment, der forløber på følgende måde: Det er muligt, ved hjælp af kunstigt integrerede kredsløb, at efterligne måden hjernen er organiseret på, og få kredsløbene til at fungere som neuroner arbejder i hjernen, og kredsløb kan derfor erstatte neuroner. Tag en forsøgsperson og udskift kontinuerligt flere og flere af personens neuroner med digitale kredse. For eksempel er det muligt at erstatte en lille lokal del af hjernen med chips; lad denne del være synscenteret. I forsøget tilsluttes nu både det kunstige og det naturlige synscenter til hjernen via en kontakt. Kontakten gør det muligt at skifte mellem to tilstande: i første tilstand virker hele hjernen normalt (kunstige synscenter er slået fra), og i anden tilstand (kunstige synscenter er slået til) virker hjernen også normalt, bortset fra at chip arbejder i stedet for neuroner i synscenteret, så hjernen stadig virker på samme måde som før. Organisationen af hjernen er ikke ændret, da chips og neuroner har samme funktion, og der vil ikke opstå nogen ændring i personens adfærd.

Antag (reductio ad absurdum) at computere organiseret på samme måde som den menneskelige hjerne, ikke har bevidsthed, og at der derfor heller ikke følger nogen

[13]Se for eksempel D. Chalmers "The puzzle of concious experience" p. 68

bevidsthedsmæssig oplevelse af 'at se'. Nu begynder vi at 'zappe', så forsøgspersonens hjerne hele tiden skifter mellem de to føromtalte tilstande. Ser personen på noget rødt, vil der i første tilstand indtræffe en oplevelse af rødhed, mens der i den anden tilstand (per antagelse) ikke er en oplevelse af noget. Personen oplever derfor at verden 'blinker', når der skiftes mellem de to tilstande - 'dansende kvalia' som Chalmers kalder det. Muligheden for at sige dette er ikke tilstede, da der ikke må ske en ændring i personens adfærd. Absurditeten ligger altså i, at personen ikke må sige at oplevelsen ændrer sig, til trods for at den gør det. Derfor har computere og alle andre systemer der er organiseret på samme måde som menneskets hjerne bevidsthed.

Chalmers fejlslutning

Lige meget om man er enig i Chalmers ræsonnement eller ej, må man medgive at argumentet er svær at angribe. Jeg vil dog forsøge dette alligevel, hvis konsekvensen for argumentet er, at for eksempel en stor elektrisk togbane opbygget på samme måde som hjernen, vil kunne opleve følelser som smerte og glæde.
Chalmers ønsker at argumentere for, at når en neuron udskiftes med en digitalt kreds, sker der ikke en ændring i bevidstheden, fordi de virker på samme måde.

> "Because chips and neurons have the same function, they are interchangeable, with the proper interfacing. Chips therefore can replace neurons, producing a continuum of cases in which a successively larger proportion of neurons are replaced by chips"[14]

Reductio ad absurdum argumentet der anvendes for at konkludere dette, mener jeg ikke er gyldigt. Hvad er det der antages?
For det første er der selve reductio ad absurdum argumentet: at alt der er organiseret på samme måde som den menneskelige hjerne, ikke vil have bevidsthed. Men dette er ikke det eneste der antages. Chalmers påstår, at fordi organisationen af hjernen ikke er blevet ændret, vil der ikke ske nogen ændring i forsøgspersonens ydre opførsel, når der skiftes mellem de to tilstande. Udskiftningen af en enkelt neuron med en chip vil ikke ændre noget; den elektriske kreds arbejder ligesom neuronen, så hvorfor skulle udskiftning af flere neuroner ændre noget? Dette sorites-lignende argument er ikke holdbart. Jeg kunne let forestille mig noget andet ville ske i dette kontinuum af udskiftninger.

[14] D. Chalmers "The puzzle of conscious experience", p. 68

Mennesker tænker i helheder. Når vi i eksperimentet ser på den tomat, der foresager oplevelsen af rødhed, ser jeg ikke udelukkende en rød plamage. Hele situationen medtænkes, og der associeres let og hurtigt i en lang række af tanker, hvor den ene association foresager den næste. Ser jeg en tomat, tænker jeg straks på om den er sprøjtet med gift, måske er den lidt større end min nabos plejer at være (er det mon fordi den har fået kemikalier?), jeg skal huske at spørge om han kan passe katten når vi skal besøge mine forældre i ferien, a propos forældre skal jeg have købt den gave til min mor (gad vide om butikkerne har længe åbent i morgen?), o.s.v. Det kræver ikke kun stor koncentration, men også mange års træning ved hjælp af meditation at stoppe denne 'indre monolog' - et stadie som kun de færreste mennesker når.

Ændres flere neuroner til chips, kan det være, at bevidstheden også vil ændre sig. Måske vil bevidstheden ændre sig men kun ganske lidt, i den kontinuerlige udskiftning af neuroner: associationen kan ændres lidt, den generelle holdning overfor genstanden kan ændres lidt, en følelse der plejer at optræde i sådanne forbindelser er forandret lidt, e.t.c. Måske vil hele den situation hvori man opfatter tingene ændres, men kun meget lidt. Dette vil være tilfældet, såfremt bevidsthed er mere end hvad Chalmers materialistiske synspunkt vil gøre den til. Muligheden for at Chalmers har ret, eksisterer stadigvæk, men jeg påpeger, at vi ikke kan vide dette på forhånd. Da der aldrig kan opstilles de fuldstændige identiske forsøgsopstillinger i psykologiske test som denne - den nøjagtige samme associationsrække vil ikke forekomme to gange i træk når en tomat betragtes - er det principielt umuligt at afgøre hvem der har ret. Derfor bliver Chalmers nød til at antage, at den ydre opførsel ikke ændrer sig overhovedet, når neuronerne udskiftes. Bemærk at der nu er tale om en antagelse, og ikke en kendsgerning eller et faktum.

I Chalmers argumentation indgår derfor to antagelser: for det første at computere ikke har bevidsthed og den anden, før skjulte men nu blotlagte antagelse, er, at testpersonens adfærd ikke ændres ved udskiftning fra neuron til chip,. Når han i reductio ad absurdum argumentet når til en absurditet, er det ikke til at afgøre hvilken af de to antagelser der er forkert. Det kan være den første der er falsk, den anden eller dem begge to!

Bevidsthedens uforudsigelighed

Bevidsthed er et begreb, jeg forbinder med uforudsigelighed. Intelligens er egenskaber ved bevidste systemer. Såfremt computere kan kaldes intelligente, er det altså ikke a priori udelukket, at de kan have bevidsthed, men intelligens er nødvendig

for at systemet kan være bevidst.
Min tese er, at bevidste systemer er uforudsigelige. Beder jeg en person om at beregne en Fouriertransformation for mig, kan personens svar ikke forudsiges. Jeg kan have en anelse om hvorvidt personen har tid, er kvalificeret, er villig eller forpligtet til at hjælpe, men reaktionen kender jeg ikke (Er det ikke det der gør livet interessant ?). Et svar kunne være, at personen ikke gad på grund af træthed, men at vi kunne regne opgaven i morgen. Naturligvis er der mange situationer, hvor vi kan have ret stor vished om den respons vi vil få, i vores omgang med andre uforudsigelige systemer. Autistiske personer er et eksempel, hvor der hyppigt optræder forudsigelighed, til trods for at autister er bevidste. Dette tilbageviser dog ikke min tese om, *at alle bevidste systemers vil have reaktioner der ikke er mulige at forudsige*. Selv personer der har været involveret i en ulykke, og er blevet lam i hele kroppen, kan både være bevidste og uforudsigelig. Selv om personen ikke kan kommunikere, kan man mærke at han den ene dag kan være glad, for dagen efter at være irriteret.
Uforudsigeligheden er med andre ord umuligheden i, at i alle henseender at få den respons man forventer. Jeg har således argumenteret for, at bevidste systemer både er intelligente og uforudsigelige.

Maskiner bygget efter klassisk AI, er meget lidt intelligente og helt forudsigelige. Det er på samme tid de regelbaserede systemers force og begrænsning, at være underlagt konformitet, og hvad angår forudsigelighed, er det klart en begrænsning. Man har forsøgt at argumentere for, at computere er uforudsigelige, men denne argumentation består i at påpege, at der opstår programmeringsfejl, virus og brist p.g.a. slitage på de mest uventede tidspunkter[15]. Denne begrundelse holder ikke, fordi uforudsigelighed er blevet forvekslet med upålidelighed. Uforudsigeligheden skal ikke forstås som en overraskelse i form af en kortslutning i et kredsløb, men som en ikke-programmeret ytring eller handling.

Kunne jeg forestille mig menneskeskabte maskiner - der ikke bygger på nutidens AI - der var bevidste? Er det muligt i fremtiden at lave 'menneskelige' computere? Den logiske konsekvens af ovenstående argumentation er, at systemer der er forudsigelige, ikke kan have bevidsthed. For at en computer kan siges at besidde bevidsthed, er det ikke tilstrækkeligt, at den altid bare giver svaret, når jeg beder den beregne en Fouriertransformation, men for eksempel svarede, at nu var den træt, fordi den havde været tændt i mange timer, for derefter at lukke sig selv ned.

[15]Se F. J. Crosson: "Human and artificial intelligence", p. 121

Skulle sådan en situation opstå i dag, ville alle straks tro, at der var tale om en gimmick fra programmørens side eller en computervirus. Hvad fremtiden bringer kan vi ikke afgøre, men let fastslå, at ingen af de maskiner vi anvender i dag ville kunne være spontane på denne måde, de er ikke uforudsigelige. Når maskiner genererer tilfældige tal, er tallene uforudsigelige, men pointen er, at når vi beder maskinen om at gennemløbe programmet der finder tallene, vil den per automatik udføre ordren, og ikke overveje konsekvenser (mængden af arbejde opgaven medfører), moralske aspekter (om hasardspil er lovligt) eller andet. Det er bl. a. sådanne overvejelser, der medfører uforudsigelighed.

Ingen bevidsthed uden liv

I Sørlanders bog "Det uomgængelige", undersøges betingelserne for bevidsthed. Her fremkommer synspunktet, at simpel fysisk kompleksitet ikke er nok for at bevidsthed opstår, men der må indgå en speciel struktur: nemlig de enheder bevidsthedsmæssige fænomener består af. Men da det er "logisk umuligt"[16] at få fysiske love til at forsage ikke-fysiske processer, sluttes følgende:

> "Det kan altså konkluderes, at bevidsthed forudsætter liv. En fysisk materie, som ikke opfylder betingelserne for liv, kan heller ikke opfylde betingelserne for bevidsthed. Dette er en rent begrebslogisk konklusion - på et rent begrebslogisk problem..."[17]

Hvordan 'liv' i denne sammenhæng skal opfattes - om det også dækker amøber, insekter og planter - er svært af afgøre. Sørlander vil sandsynligvis benægte at bevidsthed kan eksistere på disse niveauer, idet han ofte taler om sanseorganer, nervesystemer og hjerner.

Han forklarer altså umuligheden i, at computere skal kunne få bevidsthed, ved at påvise problemet i at konstruere bevidsthed udelukkende ved hjælp af fysisk materie. Desuden påpeger han, at hvis vi opbygger et system, bliver det således et rent fysiske system, og derfor giver vi ikke plads til systemets eget formål. Det antages nemlig, at en

> "... betingelse for, at et fysisk objekt kan have bevidsthed er, at det umiddelbart kan *forandre sig selv* med formål. ... Og da et fysisk objekt er karakteriseret ved at være rumligt, så må den fundamentale betingelse for, at et objekt kan have bevidsthed,

[16] Se Kai Sørlander: "Det uomgængelige", p. 130

[17] do, p. 131

være, at objektet kan *bevæge sig selv* (eller dele af sig selv) med formål. Bevidsthedsvæsner er altså karakteriseret ved ... at de kan bevæge sig selv formålsrettet."[18]

Når man læser i bogen kan man få den opfattelse, at Sørlander har taget udgangspunkt i bevidsthedsvæsner (fortrinsvis mennesker) og set at de har mobilitet tilfælles, og derudfra sluttet, at bevægelse er nødvendig for bevidsthed. Dermed har han på forhånd ikke kun afskåret stationære computere fra at kunne være bevidste, men, hvad værre er, også personer der er blevet lam i hele kroppen.
Argumentation går med andre ord på, at vores fysik bygger fysiske systemer, der ikke kan bevæge sig med formålsrettet adfærd, da en betingelse for denne adfærd er, at der må eksistere visse mentale oplevelser, som sanseindtryk og tilbøjeligheder - som Sørlander forklarer ud fra videre argumentation. Bevidsthedsvæsner er derfor ikke blot konstrueret af det fysiske, men også af noget som adskiller sig principielt derfra: de må være levende. At bygge maskiner der har bevidsthed, er derfor udelukket.

Sørlander under lup

'Første trin' i argumentationen, (at betingelsen for bevidsthed er at systemet kan forandre sig selv med formål) er et princip der svarer meget til min opfattelse af 'uforudsigelighed' som tidligere beskrevet, men jeg mener ikke, at bevægeakten er afgørende. Uforudsigelige systemer, må have et formål med at være uforudsigelige, da der ellers bare er tale om en refleks - forstået som en respons der ikke først er gennemtænkt eller planlagt. De må have en motivation til at udføre en givet respons. Består systemer udelukkende af reflekser er de ikke bare uforudsigelige, så der eksisterer reaktioner der ikke er mulige at forudsige, men også uberegnelige, så alle reaktionerne systemet udfører ikke kan forudsiges; som hop fra springbønner eller henfald fra radioaktivt materiale.
Opfattes det 'at forandre sig selv med formål', som at være i stand til at 'bryde ud' og standse arbejdet i en vilkårlig proces på et vilkårligt tidspunkt, selv inden processen er gået i gang, vil dette svare til at være uforudsigelig. Jeg mener derfor, at jeg i det følgende kan benytte begge udtryk, således at de dækker det samme, uden at dette ændrer ved Sørlanders anvendelse og opfattelse af udtrykket.
Sørlander mener, at uforudsigelighed er et nødvendigt kriterium for bevidsthed, og deri

[18] do, p. 116

vil jeg give ham ret. Men er det sikkert at mentale oplevelser er et nødvendigt kriterie for uforudsigelighed, og at det derfor er liv, der afgør om et system besidder bevidsthed eller ej? Det kan tænkes at computere i fremtiden kan lære at 'tænke' på flere planer via 'metatænkning', hvilket der er intet selvmodsigende i. Stillet overfor en opgave, præsterer den både at beregne selve opgaven og ræsonnere over opgaven. For eksempel kunne den give instruktioner og/eller gode råd, angående opgavens udformning eller andet. Kvintessensen i uforudsigelighed består måske i, at kunne ræsonnere over sit ræsonnement: 'Hvorfor valgte jeg netop denne løsning ?', 'Stemmer dette resultat overens med tidligere resultat ?', o.s.v. Liv er ikke nødvendigvis en forudsætning for tanker på et sådant niveau. Spørgsmålet er om Sørlanders konklusion - at liv betinger uforudsigelighed - er realistisk, udelukkende fordi dette har været normen indtil i dag, og sikkert vil være det i lang tid endnu. Computere der formår at ræsonnere på metaniveau, er også i stand til at ræsonnere over deres egen situation. Herfra er der ikke langt til at maskinen kan fremsætte den uforudsigelige, ikke-programmerede sætning: 'Sluk ikke for mig igen - jeg har ret til at være tændt, hvis det er det jeg ønsker !'.

Personligt er jeg skeptisk, overfor muligheden for konstruktion af bevidste computere, men principielt kan vi ikke udelukke dette. Hvad fremtiden bringer, vil jeg end ikke gætte på, men påpege, at det fysiske paradigme videnskaben i dag anvender, nok ikke er fyldestgørende til at opfylde betingelsen for uforudsigelighed.

Bevidsthed som fysisk konstruktion

Sørlander har med ovenstående argument indirekte givet en kritik af et af tankeeksperimenterne i Erich Klawonns "Jeg'ets ontologi". I Klawonns egne ord lyder eksperimentet således:

> "'Teletransport' er et af standard-eksemplerne i den moderne diskussion om personlig identitet. Det er en (formodet) fremtidig rejsemetode, der involverer to avancerede computere - én ved rejsens begyndelsespunkt, en anden ved bestemmelsesstedet. Den første underkaster den rejsendes fysiske struktur en gennemgribende undersøgelse (helt ned til individuelle atomers placering); den rejsendes krop nedbrydes, og en fuldstændig beskrivelse af kroppens fysiske struktur afsendes til computer nummer to, der rekonstruerer ham på bestemmelsesstedet ved hjælp af lokal materie."[19]

[19] E. Klawonn: "Jeg'ets ontologi", p. 18

Muligvis er tankeeksperimentet logisk muligt, d.v.s. at det ikke a priori kan forkastes, men selve computerens udformning i eksperimentet er problematisk. Vanskeligheden ligger naturligvis i, at personen der teletransporteres, gerne skulle forblive et helt menneske af kød og blod - og med bevidsthed. Computeren på bestemmelsesstedet skal derfor være i stand til at fremstille bevidsthed, så det er ikke tilstrækkeligt, at det uhyre store kompleks af atomer udelukkende bliver sat rigtigt sammen igen. Ifølge Sørlanders analyse må computeren betragtes som en miniatureudgave af Gud: den må være i stand til at frembringe liv, som forudsætning for bevidsthed. Sandsynligheden for at mennesker i fremtiden kan konstruere maskiner, der kan frembringe liv, er formodentlig mindre, end sandsynligheden for at konstruere uforudsigelige, 'metatænkende' maskiner.

Kan computeren i eksperimentet frembringe bevidsthed, uden selv at have bevidsthed ? Mennesker kan godt fabrikere lysende lygter, uden at mennesker selv kan lyse. Men mindst to antagelser er benyttet i konklusion: For det første har vi alt materiel som skal anvendes til at konstruere lygterne med, og for det andet ved vi hvordan lys frembringes. Denne observation er der intet odiøst i, men beskriver tværtimod almindelig praksis i vilkårlige grene indenfor videnskaben, som fysik og kemi. Skelnen mellem de to former for viden kaldes også deklarativ-procedural sondringen, da deklarativ viden er viden om hvordan noget forholder sig, og procedural viden er viden om hvordan noget skal gøres. Skal computeren frembringe bevidsthed må alt materialet som anvendes til at konstruere bevidsthed forefindes, og viden der fortæller hvordan bevidsthed opstår, skal være mulig at frembringe.

Men Sørlander påpeger en vigtig beskaffenhed ved fysikken: den behersker per definition ikke det mentale niveau, og kan ikke slå bro over det skisma dualismen indeholder. Bevidstheden er privat, urumlig, indeholder en intentionalitet og er flygtig da den er af kvalitativ natur. Fysikken derimod beskæftiger sig med det offentligt tilgængelige, rumlige, der ikke har intentionalitet men kan kvantiseres. Det afgørende er derfor ikke om maskinen der skal konstruere bevidsthed selv har bevidsthed, men hvorvidt det materiale der skal anvendes kan fremskaffes. Jeg er enig med Sørlander i, at med det paradigme fysikken betjener sig af i dag, er det principielt ikke er muligt at forme bevidsthed med de midler vi har til rådighed.

Igen mener jeg, at det ikke nødvendigvis altid må vedblive med at være sådan. Emner der for bare få år tilbage blev anset som eller sidestillet med sort magi, fordi det ikke var muligt at kvantisere den effekt fænomenet udførte, er i dag selvstændige forskningsområder. Dette gælder blandt andet for emner som akupunktur og placeboeffekt.

Måske er det med bevidstheden som med placeboeffekt: fænomenet er svært at beskrive fysisk, til trods for at der synes at være stor enighed i fænomenets realitet og gyldighed, og derfor kan det muligvis forklares i fremtiden. Men for at komme bevidstheden meget nærmere, er det ikke nok med videre forskning, men brug for en videnskabelig revolution, så det nuværende paradigmes problemer kan overvindes.
Klawonns tankeeksperimentet kan derfor ikke anvendes før efter en sådan revolution har indtruffet. Indtil da kan situationen reddes, hvis man er villig til at vedkende sig en materialistisk teori, så bevidstheden udelukkende skal beskrives som fysisk-kemiske processer.

> "Problemet [med materialismen] er ikke blot at forklare 'awareness'. Den, der vil reducere al tænkning og sprog til fysiske processer, eliminerer begreberne 'mening' og 'sandhed' og fratager derved sig muligheden for at hævde *sandheden* af sit synspunkt. Hvis tænkning blot er hjerneprocesser, som igen skal opfattes som fysiske processer på linie med vandfald og tordenvejr, så har hverken tanker eller sætninger mening - så lidt som vandfald og tordenvejr har mening eller udsiger noget."[20]

Læses 'Jeg'ets ontologi' indser man let, at den materialistiske udvej ikke er mulig for Klawonn; og vælges denne løsning alligevel, er spørgsmålet om ikke medicinen er værre end sygdommen.

Afsluttende bemærkninger

Jeg har i dette essay beskæftiget mig med intelligens og bevidsthed i computersammenhænge. Alle tankeeksperimenterne der indgår er blevet kritiseret, enten på grund af fejl i argumentationen eller skjulte antagelser. Mit eget bud på, hvad det er der karakteriserer bevidsthed har jeg også forklaret: uforudsigeligheden i systemet. Hele tiden har jeg forsøgt at holde argumentationen på et principielt plan, ved at beskrive umuligheden i at konstruere bevidste systemer ved hjælp af det fysiske paradigme der anvendes i dag, men samtidig holdt muligheden åben for, at fremtiden måske vil kunne overvinde dette problem.
Det er bevidst, at jeg ikke selv har taget store udflugter ud i tankeeksperimenternes verden. En sådan argumentation kommer hurtigt til at indeholde mange 'hvis'er', og det var ikke meningen, at dette essay blot skulle være en samling luftkasteller, men et

[20] D. Favrholdt: "Den cartesiske fejl". Philosophia Årgang 21 nr. 1-2, 1992

forsøg på at beskrive og kritisere synspunkter angående bevidsthed og intelligens i digitale kredsløb. Hvorvidt det er lykkedes, overlades til læseren at bedømme.
Jeg vil slutte med en morsom og paradoksal graffitti, jeg læste for længe siden:

Computere er ikke intelligente; de tror bare selv at de er det!

Litteraturliste

Bøger:

Bernsen, Niels Ole & Ulbæk, Ib: *Naturlig og kunstig intelligens*. Nyt Nordisk forlag, 1993

Crosson, F J: *Human and artificial intelligence*. Meredith Corporation, 1970

Kirkeby, Ole Fogh: *Ekspertsystemer og kunstig intelligens*. Borgen 1987

Hansen, Kirsten: *Erkendelse og bevidsthed*. Gyldendalske boghandel, 1987

Wahlgren, P: *Automation of legal reasoning - a study of artificial intelligence and law*. Kluwer and taxation Publishers 1992

Michie, Donald & Johnston, Rory: *The creative computer mashine and human knowledge*. Viking 1984

Searle, John: *Minds, brains and science*. Penguin Books 1989

Kirkeby, Ole Fogh & Tambo, Torben: *Guds ur - Om den store videnskab og dens kulmination i kunstig intelligens, computere og neurale netværk*. Gyldendal 1992

Artikler:

Turing, A M: *Computer machinery and intelligence*. Mind Design, John Haugeland (Ed.) MIT 1997, pp. 29-56

Chalmers, David: *The puzzle of concious experience*. Scientific American, December 1995, pp. 62-68

Chalmers David: *Facing up to the problem of conciousness*. Fra internettet URL: http://ling.ucsc.edu/~chalmers/papers/consciousness.html pp. 1-16

Spelling, Kai: *Skabende tænkning*. Intelligens og tænkning, Berlingske Forlag 1972, pp. 88-112

Alchourrón, C & Bulygin, E: *Limits of logic and legal reasoning*. Expertsystems in law. Antonio Marino (Ed.) Elsevier Science Publishers B.V. 1992, pp. 9-29

Mulder, R V & Noortwijk, C van & Keerkmeester, H O: *Knowledge systems and law - the Juricas project*. Expertsystems in law. Antonio Marino (Ed.) Elsevier Science

Publishers B.V. 1992, pp. 163-171

Derudover uddrag fra

Sørlander, Kai: *Det uomgængelige*. Munksgaard 1994

Klawonn, Erich: *Jeg'ets ontologi*. Odense Universitetsforlag, 1991

Favrholdt, David: *Den cartesiske fejl*. Philosophia Årgang 21 nr. 1-2, Filosofisk forening i Århus, 1992

Kirkeby, Ole Fogh: *Ægte intelligens - om bevidsthedens program*. Munksgaard 1989

Hansen, Mogens: *Intelligens - om hjemen, tænkningen og erkendelsen*. Forlaget Ålykke 1989