

Hvad blev der af HAL?

Af Simon Laub - en stor HAL fan, i hvert fald når HAL lytter med.

HAL er som bekendt hovedpersonen i Stanley Kubrick og Arthur C. Clarkes science fiction film "2001 - a Space Odyssey" fra 1968. En rigtig drømme computer, der kan tænke, tale, se, føle, mund aflæse og engang imellem gå rigtig amok.

Nu er det jo sådan med science fiction film, at de ofte er mere fiktion end videnskab. I den obligatoriske science fiction rumkrig affyrer helten sit laser våben mod en fjendens rumskib. Hvorfor ser vi nu et laserstråle? Der er jo ingen luftmolekyler eller andet til at afbøje fotoner over mod tilskuerne. Fjendens rumskib bliver ramt og springer i luften med et stort brag. Men hvordan? Der er jo ingen lyde i det ydre rum?

Filmen "2001" er ganske anderledes. Kubrick og Clarke skabte her et univers, hvor de gjorde sig stor umage med at få science delen så rigtig som muligt. Deres forestillinger, om hvor langt man ville være nået i 2001, viste sig dog at være noget vel optimistiske. Nu her i 2002 kan vi se at det ikke var i 2001, at det blev helt hverdagsagtigt trivielt at tage på bemandede rejser til Månen. Og da slet ikke det år, hvor bemandede missioner til Jupiter var på programmet. Computere med HAL's egenskaber så vi heller ikke så voldsomt meget til.

For hvad var det egentlig for en computer vi blev præsenteret for i filmen "2001"? Hvad mangler

ens årgang 2001-02 computer egentlig for at blive som HAL? Siden filmens præmiere er disse og lignende spørgsmål blevet behandlet i et utal af bøger og artikler. Hvor HAL jubelåret 2001 var kulminationen.

Sikkert ude af jubelåret synes jeg, at det er tid at gøre status over nogle af de ting man har sagt om HAL i de senere år. Så i det følgende vil jeg kigge lidt nærmere på HAL's evner, og hvad man i de sidste år har ment om dem. Dernæst vil jeg gå lidt nærmere ind på det, der for mig er det centrale spørgsmål - hvem er HAL egentlig? Og hvordan har man dog bildt HAL ind, at han har et rigtigt "jeg", ligesom en rigtig Homo Sap. ?

Til det formål har jeg fundet inspiration i bøgerne "HAL's Legacy" af David Stork og "The feeling of what happens" af Antonio Damasio.

30 år ude i fremtiden.

Ifølge Moores lov forbedres computere til ca. dobbelt ydeevne hver gang der er gået ca. 1 ½ år. Og 30 år ude i fremtiden er nu engang lang tid. Så i 1968 var Kubrick og Clarke i deres gode ret til at mene, at HAL i 2001 ville bestå Turing testen i overlegen stil (hvor en computer består Turing testen, hvis en tester, der føre samtale med den igennem keyboard og skærm, ikke kan afgøre om det er en person eller en computer der tales med). Den kedelige sideeffekt (med at HAL bagefter

testen sikkert ville slå testeren ihjel) var sikkert heller ikke sådan at udelukke.

Men helt præcist at regne ud hvordan verden ser ud 30 år ude i fremtiden kræver en god krystalkugle. Bemærk f.eks. at ombord på rumskibet Discovery bruger astronauterne Bowman og Poole stadig papir og blyant. Laptops og PDA'er var ikke lette at forudsige i 1968. Mange andre ting var Kubrick og Clarke dog ganske tæt på at få forudsagt korrekt.

"HAL's Legacy" blev udgivet i forbindelse med HAL's fødselsdag d. 12 Januar 1997. Heri giver Marvin Minsky, Roger C. Schank, Douglas B. Lenat m.fl. et billede af hvor langt menneskeheden egentlig er fra at skabe en computer a la HAL. Ikke overraskende viser det sig at nogle af HAL's talenter er lige om hjørnet, andre er langt væk.

HAL's syn og tale.

HAL har et skarpt syn. HAL genkender ikke bare ansigter. Da astronaut Bowman viser ham nogle tegninger, er HAL istand til både at genkende personerne på Bowman's tegning og bemærke at Bowman er blevet bedre til at tegne. HAL har altså (selvfølgelig) også kunstnerisk sans. Nu er det jo så man ikke forstår, hvorfor man i 2002 stadig går rundt med nøgler, for at komme ind i ens hjem? Hvorfor er der ikke et videokamera, der genkender een og byder velkommen?

Var Kubrick og Clarke langt forud for deres tid vedr. HAL's syns evne, så var de endnu længere fremme med HAL's talegaver:

HAL: "I enjoy working with people."

Indholdet i sig selv er et mysterium, når man tænker på hvad HAL senere gjorde ved astronauterne. HAL's udtale af ordene gør det endnu bedre. Hvis man fremhæver "I" så er det til forskel fra "you". Hvis man fremhæver "working" så er det til forskel fra "playing". Hvor HAL naturligvis betoner perfekt, som om han forstod indholdet – end dog meget godt.

Og HAL kan naturligvis tale med om alt. En mindre imponerende HAL ville kun have lokal intelligens, og ville kun kunne tale om nogle få emner. F.eks. styring og vedligeholdelse af rumskibet Discovery. Dvs. en sådan computer ville være klog på de få områder den ville have behov for at vide noget om - og holde mund resten af tiden. Hvis HAL endelig skulle kunne tale om alt, skulle man tro, at han ville bruge den "nemme" taktik. Dvs. den "taktik" som mange mennesker også benytter sig af :-). Disse mennesker har basalt set ca. 10 yndlings historier (om sig selv), de godt kan lide at fortælle og som de ønsker og forventer at personen i den anden ende også skal påskønne. Øvelsen går så ud på at finde et ord, begreb, ide eller pointe i det den anden fortæller, som så kan udløse en af ens egne 10 historier.

Men sådan er HAL jo ikke - han samtaler rigtigt med folk.

HAL's almene viden.

For sådan at kunne samtale med folk om alt må alt HAL jo have et stort forråd af almen viden. Det kan virke lidt mystisk, hvis ide'en bare var at HAL skulle flyve rumskibet - og dog!

Sammenligner man med et computer, der sættes til at styre en bil ude i den almindelige trafik, er det vel godt nok at en sådan robotchauffør har lidt ide om værdien af et menneskebarn, sammenlignet med f.eks. et egern. At hunde oftere vil løbe ud foran en bil end gamle damer i rullestol osv.

Men gad vide hvor HAL's common knowledge kommer fra? Har vi allesammen sendt de UML diagrammer vi sidder med i det daglige ind til HAL? For sådan at kick starte hans viden om verden? Og lyder det ikke lidt mærkeligt, at HAL er født og blevet "operational" på en bestemt dato? Skulle man ikke forvente, at han som et barn skulle have en lang indlæringsperiode før han kunne blive "operational" og kaptajn på et rumskib?

HAL's følelser.

Måske er HAL blevet så stopfodret med facts, at det er gået ud over hans følelsesliv?

Hvis følelseskolde mennesker skal kategoriseres som robotter, må man finde på en ny betegnelse for HAL! Som bekendt dræber HAL 4 ud af 5 besætningsmedlemmer, og synes ærlig talt noget af en kold fisk efter mordene. Det er sikkert derfor man i "the Operating Manual for the HAL 9000 computer" anbefaler, at der placeres små sprængladninger rundt omkring i Mainramen? Sådan

"just in case". Måske vidste BBC interviewerens noget da han talte med HAL?

BBC interviewer: "HAL, despite your enormous intellect, are you ever frustrated by your dependence on people to carry out actions?"

HAL: "Not in the slightest bit. I enjoy working with people. I have a stimulating relationship with Dr. Poole and Dr. Bowman."

Hvordan ved HAL hvad det vil sige at være frustreret? Mennesker bliver frustrerede, når de ikke får hvad de ønsker. Når andre menneskers mål og handlinger lægger dem hindringer i vejen. Pointen er at det vel altid er i hvert fald lidt frustrerende at arbejde sammen med andre mennesker. Andre mennesker vil altid have lidt anderledes mål. Så enten aner HAL ikke hvad "frustrerende" betyder eller også lyver han. Men den politiske korrekte besvarelse er jo selvfølgelig af den type, som folk bliver forfremmet på. Så måske er HAL bare meget klog?

Under alle omstændigheder er følelser jo vigtige. De hjælper os mennesker med at skabe en reaktion på en (ofte kompliceret) situation. Så selvfølgelig skal et væsen som HAL også have følelser. Fornemmelsen af at tænde for en morgensur computer skal vi nok lære at blive glade for her i de kommende år.

HAL som skakspiller.

Selvom HAL's følelsesliv ikke virker godt. Og det derved ikke kan støtte ham til at træffe kloge beslutninger i denne komplekse verden. Så er han alligevel ret god til at begå sig! Hans planlægningsevner er fænomenale.

F.eks. spiller HAL skak ligesom et menneske. Istedet for at spille det optimale træk, ser man i filmen HAL spille et "fælde-træk". Det skaber en situation, hvor Franks følgende mest oplagte træk giver en umiddelbar fordel, men på lang sigt ødelægger hans position og gør ham skak mat. Hvis Frank ikke var faldet i fælden ville han have fået en bedre position end HAL. Men HAL ved hvor dygtig - eller rettere dårlig - Frank er. Og ved at han vil falde i fælden. Denne spille stil er jo selvfølgelig helt modsat almindelige skakcomputere, som ikke på den måde lægger fælder ud. De forventer istedet at deres modstander altid foretager det bedst mulige træk, og tilrettelægger deres spil ud fra det.

At lave et sådan fælde træk forudsætter kendskab til modstanderens styrker og svagheder. Og at man kan lægge planer ud fra det.

HAL's mordplaner.

HAL kan altså skabe sine egne planer. Som publikum til filmen "2001" er det da ihvertfald svært tro, at HAL ligefrem skulle være hardcoded til at slå besætningen ihjel.

Planlægning kræver at man ser på risici vs. gevinstchancer. Kigger på usikre informationer og

andre uklarheder. Ved at lægge en plan hierarkisk op kan man forhåbentlig dele en stor plan op i små delmål. Som man så kan udarbejde mindre planer for at opnå. Men let er det ikke.

Når man tænker på en år 2002 computers planlægningsevner kommer i hvert fald jeg let til at tænke på en LEGO MindStorm robot, der stabler byggeklodser oven på hinanden efter størrelse, hvilket så iøvrigt ikke nødvendigvis går godt. Skal vi ikke bare nøjes med at konstatere at HAL er anderledes?

Da astronauterne Dave Bowman og Frank Poole, mener at der er noget i vejen med HAL, beslutter de at slukke for ham. Som enhver fornuftig tænkende skabning er HAL ikke meget for sådan at blive "slukket" for. Så i stedet gør HAL det af med Poole under en rumvandring. Bowman bliver sendt på evig rumvandring og fornuftigt nok nægtet adgang til rumskibet igen ("I'm sorry Dave, I'm afraid I can't do that"). Og de øvrige 3 astronauter bliver termineret, mens de sover - som et fornuftigt forebyggende træk.

Dave Bowman: "Open the pod bay doors, HAL."

HAL: "I'm sorry Dave, I'm afraid I can't do that."

Dave Bowman: "What's the problem?"

HAL: "I think you know what the problem is just as well as I do."

Dave Bowman: "What are you talking about, HAL?"

HAL: "This mission is too important for me to allow you to jeopardize it." [On Dave's return to the ship]

HAL: "Look Dave, I can see you're really upset about this."

HAL: "I know I've made some very poor decisions recently, but I can give you my complete assurance that my work will be back to normal."

"Look Dave, I can see you're really upset about this...I honestly think you ought to sit down calmly, take a stress pill and think things over."

HAL er altså en maskine, der siger "I'm afraid I can't do that". Men i virkeligheden mener "det vil jeg ikke". Første gang HAL's skabere så den slags adfærd må de i sandhed have været stolte !?

Ikke nok med at HAL ved alt om hvad en interesse-konflikt er. Så er han også istand til at håndtere den høfligt. Virkelig cool.

Lyve, det er jo næsten højdepunktet af menneskelig (social) intelligens. Det kræver at løgneren ved hvad den anden person vil være tilbøjelig til at tro på. Skabe en historie som passer til situationen og modificere den løbende som situationen kræver det. Noget de fleste mennesker i virkeligheden ikke kan.

Men HAL gør det strålende.

HAL's: "Look Dave, I can see you're really upset about this...I honestly think you ought to sit down calmly, take a stress pill and think things over."

Er jeg dog ikke så imponeret over. HAL har tydeligvis aflæst Dave følelsesmæssige situation korrekt. Men hvis meningen er at få Dave til at blive mindre ophidset er en sådan kommentar næppe vejen.

Men det er da et imponerende studie i maskin intelligens vi her er vidne til: "Jeg har slået Dave's kollegaer ihjel, og forsøgt at lukke ham ude i det ydre rum nær Jupiter ved at forhindre ham adgang til rumskibet. Jeg, HAL, deducerer hermed at Dave vil være ophidset".

Sherlock Holmes kunne ikke have gjort det bedre.

ErrorCorrection og HAL.

Med disse fænomenale planlægningsevner er det ikke underligt at HAL er stolt af sig selv:

HAL: "Let me put it this way, Mr. Amer. The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error".

Igen, hvis HAL var et menneske ville han selvfølgelig straks blive forfremmet efter denne svada.

Selvtillid er jo altid en god ting. Men udtalelsen rummer nu også elementer af rent vanvid! Og man burde nu have overvejet om det var en god ide at lade HAL lede missionen.

Sammenlignes Kubrick og Clarkes Discovery-rumskib med rumfærgen, så styres rumfærgen ikke af en, men af flere computere. Ved majority voting blandt flere HAL lignende computere (hvoraf nogle kunne stå på Jorden), skulle man have råd til at en enkelt gik amok. Men selvfølgelig, at koble flere HAL computere sammen er måske heller ikke helt ufarligt?

At sætte HAL sammen med andre computere er så iøvrigt nok en naturlig tanke i 2002. Men 1968 var et år før internettets (ARPANET) fødsel. Og manuskriptet til filmen var ihvertfald ikke skrevet på en computer med Windows. Måske var fremtidstroen bare større dengang?

HAL's "jeg"

Det helt store spørgsmål for mig er imidlertid, hvordan man har bildt HAL ind, at han har et "jeg", som han kan lægge alle sine andre fantastiske egenskaber op ovenpå.

Det får vi nok aldrig at vide, men Antonio Damasio har gjort sig overvejelser om "jeg'et", som HAL's konstruktører også må have været omkring.

Bevidsthed og "jeg"

Antonio Damasio's overvejelser findes i bogen "The feeling of what happens". Her fortæller han

om hvordan krop, følelser og bevidsthed hænger sammen. Og hvordan et menneskes "jeg" skabes.

Damasios baggrund er, som andre hjerneforskeres, et stort materiale af patienter, som har haft læsioner forskellige steder i hjernen. Dette har man så kunnet studere effekten af. En patient har f.eks. skadet hjernecentret amygdala - og kun det - vedkommende fungerer så egentligt ganske normalt bortset fra "små" sideeffekter, såsom en manglende evne til at forstå frygt. En skade på Broca centret giver alvorlige taleforstyrrelse osv.

En organisme skal kunne reagere på sine omgivelser.

At reagere på omgivelsernes ændringer er selvsagt vigtigt for en organisme. Så en mekanisme der kan hjælpe med at behandle de billeder, som en organisme danner af sig selv og sine omgivelser vil givetvis være en fornuftig ting.

Bevidsthed er netop en sådan mekanisme. Herunder også dannelsen af et "jeg". Men også følelser kommer med i værktøjskassen, der skal bruges i overlevelseskampen.

Vel er samspillet i en hjerne mellem krop, følelser, bevidsthed, "jeg" mm. stadig i stor udstrækning ukendt land, men Damasio tør alligevel godt give sit bud på hvordan det måske hænger sammen.

Ved at kigge lidt nærmere på en følelse kan man få et overblik:

Følelser (glæde, sorg, frygt, vrede, overraskelse, væmmelse mm.) handler om en organismes liv, mere præcist dets krop, hvor meningen som sagt er at opretholde livet.

Følelserne skal skabe (i krop og hjerne) en reaktion til en bestemt situation. Reaktionen kan så være at flygte, slås, spise, parre sig osv. Eller måske skal der først skabes en regulering af den indre tilstand, så der gives mulighed for en sådan senere reaktion.

En fugleunge i en rede kan se et stort objekt flyve over reden. Ungen ved ikke at dette kunne være en fjendtlig ørn, og er ikke bevidst om at situationen er farlig.

Alligevel kan der formes et visuelt billede i fugleungens hjerne. Andre dele af dens hjerne kan lave en respons på disse indre billeder, neuralt og motorisk, nemlig at krybe sammen i reden.

Fugleungen behøver ikke at være bevidst om dens følelse af frygt. Ungen behøver sågar ikke at have en fornemmelse for et "jeg" der gør disse ting. I.e. man kan selvfølgelig godt forestille sig en organisme med en krop og en hjerne, hvor hjernen indeholder en repræsentation af kroppen - uden at der er et "jeg" nogen steder. Det giver nok liv, og en repræsentation af liv. En organisme der kan reagere på sine omgivelser. Men (selvfølgelig), uden et "jeg" vil man mangle en ejer af dette liv.

Bevidsthed som højdepunktet af evolutionen?

Ved mere avancerede adfærdsmønstre er det dog nok en god ide at have både bevidsthed og et "jeg".

Og her er der en del konkurrerende teorier om hvordan vi danner bevidsthed og et "jeg". Damasio forkaster de fleste inden han når frem til sin egen.

En af de forkastede ideer går ud at det er sproget der giver os bevidsthed. Ikke en god ide ifølge Damasio. Ifølge Damasio refererer sprog til eksisterende ting. Hvis "jeg" og bevidsthed var frembragt af sproget ville det være et temmeligt enestående tilfælde af ord uden bagvedliggende begreb? Derudover har han studeret patienter med alvorlige hjernelæsioner i sprogcentre- uden at han har kunnet konstatere at de skulle være mindre bevidste end alle os andre.

Hos Damasio er bevidsthed heller ikke det samme som vågenhed, hukommelse, sprog eller ræsoneringsevne. Bevidsthed er heller ikke højdepunktet på den biologiske evolution. For ham er højdepunktet snarere det samlede menneskelige samfund med dets sociale-, religiøse- og politiske organisationer. Og oven på bevidsthed har han mange andre ligeså vigtige byggesten: Arbejdshukommelse, hukommelse, sprog, kreativitet, retfærdighedssans, sociale færdigheder mm.

Men bevidsthed, en organismes bevidsthed om dets eget selv og dets omgivelser, er selvsagt en

vigtig ingrediens i hvad det vil sige at være menneske. Den biologiske funktion der kan få os til at kende glæde, lidelse mm.

Uden fornemmelsen af et "jeg" fanget i at være bevidst om noget, ville de tanker en hjerne genererer være uden ejer. Organismen ville ikke vide hvem disse tanker tilhørte.

”Core consciousness” og ”Extended consciousness”.

Som udgangspunkt mener Damasio at menneskelig bevidsthed ikke er en monolit. Istedet kan bevidsthed, ifølge Damasio, indeles i "core consciousness" (som giver en organisme en fornemmelse af et "jeg" på et givet tidspunkt på et bestemt sted) og "Extended consciousness" (som giver et mere udvidet "jeg". En identitet og en person. Og placerer "jeg'et" i en individuel historie, bevidst om liv levet og med forventninger om en fremtid - opmærksom på "jeg'ets" kontekst).

De to typer bevidsthed skaber så to slags "jeg". "Jeg'et" der skabes af "core consciousness" er en transient størrelse, som genskabes hele tiden for hvert objekt som hjernen interagerer med (et "core self"). Mens vores sædvanlige forståelse af et "jeg" hænger sammen med "extended consciousness" og er "core consciousness" udvidet med de mange facetter, som karakteriserer en person. Damasio kalder dette udvidede "jeg" det autobiografiske selv.

I begge tilfælde skal "jeg-et" have et uforanderligt referencepunkt. Og her kommer organismens le-

geme så i betragtning - da en modstrid med denne antagelse jo nok betyder at organismen er død. Et "proto-jeg" består så af de neurale mønstre, der repræsenterer kroppens status i et givet øjeblik. Men "core self" er ikke identisk med dette "proto-jeg".

I "core consciousness" er alt en relation – mellem enhver form for ydre objekt, og så et "jeg". Dette er en konstant opdagelse. En ting og lige ved siden af en fornemmelse af et "jeg", der observerer denne ting.

Der er to spillere, en organisme og et objekt, hvor "core-consciousness" dannes igennem forholdet mellem disse to. Når man observerer en ting, får man lavet en neural repræsentation af det. Men ikke nok med det - en sådan observation vil ændre organismens egen repræsentation (proto-jeg'et).

Et nyt map repræsenterer så denne ændring. Relationen mellem objekt og organisme. Altså, først mappes ændringerne i protoselvet (forudsaget af et ydre objekt). Dernæst mappes relationen mellem ændringer i protoselv og ydre objekt.

Med andre ord - DENNE MAPNING ER SÅ EN ORGANISME, DER ER VED AT REPRÆSENTERE SIN EGEN ÆNDRERE STATUS.

Dette pågår hele tiden, hvorved vi konstant har et vidende "jeg". Dette "jeg" er vi altså bevidste om, hvorimod vi ikke var bevidste om proto-jeg'et.

Ud af den tomme luft er der skabt noget nyt - et "jeg" !!!

En organisme, der er i gang med at repræsentere sine egne skiftende states, skaber altså på forbløffende vis en vidende enhed inde i denne fremad løbende historie.

Man ved at man eksisterer fordi man er midt i en historie med een selv som hovedperson placeret midt i handlingen at vide.

Den fortløbende historie (der fortælles om et protoselv, der ændres af ydre objekter) er det samme, som at føle, at man er ved at vide. Hvilket igen er starten på bevidsthed. Når man så har fået genereret et "jeg" kan alle de tanker, billeder mm., der ellers er i hjernen, blive "mine" tanker og billeder.

Ifølge Damasio er der sikkert ikke een, men flere sådanne bevidstheds-generatore i hjernen. At processen opleves glat, med en ejer og et objekt skyldes blanding og udglatning - men hvordan denne "udglatning" af disse processer så laves udtaler Damasio sig ikke om.

Et "jeg" er et svar på et spørgsmål, der aldrig blev stillet.

Organismen spurgte aldrig "hvad sker der", "hvad er relationen mellem disse billeder og denne krop" - Men følelsen af et "jeg" der overvåger, er begyndelsen til svaret på disse spørgsmål. Hos mennesker kan denne historie så hurtigt omsættes til ord.

Damasios vigtigste pointe er nok, at "jeg'et" hos ham ikke forklares ved et lille super kontrol center et eller sted inde i hjernen, som så styrer resten af hjernen. Men at "jeg'et" frembringes via mekanismen beskrevet ovenfor.

Psykologer mener så i øvrigt, at mennesker danner det autobiografiske "jeg", omkring ved 18 måneders alderen. Før må verden så unægtelig være noget forvirrende. Ethvert objekt holdes op mod proto-jeg'et og man har et "core self", men der er ingen relation mellem de forskellige objekter, der er ingen fortid og der er ingen fremtid.

Nogle vil vel så sige at det lyder dejligt. Det ligner da den totale uskyld, når man således ikke er tynget af fortiden eller af forventninger om fremtiden. Under alle omstændigheder er "core consciousness" noget ganske andet end det "autobiographical self", hvori der indgår træk der bl.a. er resultatet af opvækst, kultur og et helt livs oplevelser. Noget der er ganske uden for "core consciousness" rækkevidde.

"Jeg'et", følelser og bevidsthed behøver en krop.

Hvis bevidstheden begynder at lyde noget mekanisk, så får Damasio det hurtigt til at lyde biologisk igen, ved at klarlægge hvor tæt et menneskes følelser, bevidsthed og "jeg" er forbundet til kroppen.

Det sker ved igen at kigge nærmere på følelser. Hvordan er det lige vi bliver bevidste om en f.eks. en følelse? Ifølge Damasio kan man kigge på et

hændelsesforløb, hvor et objekt kommer indenfor f.eks. synsfeltet. Det kan så give anledning til en følelse, som så kan gøres bevidst på samme måde som man kan blive bevidst om ydre objekter - her er der bare tale om et indre objekt. Men i princippet er følelsen bare et objekt som alle de andre objekter, i hvert fald ifølge Damasio.

Mere præcist - et objekt processeres visuelt, og der dannes en repræsentation for dette objekt. Denne repræsentation kan så måske aktiverer en bestemt følelse. Følelsen udløser så forskellige aktiviteter i kroppen og i andre hjernecentre.

Som man så kan blive bevidst om ved at protoselvrepræsentationen ændres p.g.a. disse følelser (objekter). På sammen måde som der kan skabes bevidsthed om andre objekter. Her er der så skabt bevidsthed om en følelse.

Det bemærkelsesværdige i historien her er at både følelser og bevidsthed er relateret til en krop. Kroppen repræsenteres i proto selvet. Følelser fortæller en historie relateret til en krop. Ved hjælp af Damasio's "bevidsthedsgenerator" kan protoselvet og følelsen tilsammen give os bevidsthed om følelsen.

Bare ord?

Men bevidsthed er stadig et uhyre subjektivt begreb. Man kan ikke opleve en andens bevidsthed. Hvordan? Skulle man flytte sin egen "core consciousness" ind i en andens hjerne? det ville jo ikke gøre nogen forskel, så vil bevidstheden jo stadig bare være andens. Men tager man noget af sit eget "autobiographical self" med ind i en an-

dens bevidsthed er det jo ikke længere den andens bevidsthed man oplever.

Til trods for dette mener Damasio nu godt at man kan tale om bevidsthed uden at blive indfanget i bevidsthedens subjektive natur.

Men der er selvfølgelig andre opfattelser af hvad bevidsthed er end Damasio's. Mange vil sikkert være stødt på f.eks. Daniel Dennetts "multiple draft model" (beskrevet i hans bog "Consciousness Explained") - her danner de forskellige hjernecentre parallelt deres egen forståelse af virkeligheden - og ovenpå det har man så noget, som Dennett sikkert ville kalde, software (kørende i hjernen). Denne "software" kører så serielt og giver derved et centrum og et fokus for disse mange forskellige parallelle historier.

Det endelige billede af hvad der mere præcist sker, må vi jo så nok vente med lidt endnu. De forskellige udlægninger nu konkurrerer snarere om at være den bedste metafor for hvad der virkelig sker.

Men HAL er sikker i sin sag:

HAL: "I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do."

Hvad han så end mener med det!

Tror vi på HAL?

Arthur C. Clarke er på det seneste citeret for at sige "det altid vil være lettere at lave intelligente organiske menneske-lignende hjerner ved hjælp af ufaglært arbejdskraft, end ved at skabe computer baseret kunstig intelligens".

Det er langt fra 1958, hvor Herbert Simon og Allen Newell skrev ("Heuristic Problem Solving"), at i 1985 vil maskiner være istand til at udføre alle de funktioner et menneske kan. Nu er det god latin at være lidt mere opmærksom på de problemer man kan støde ind i på vejen mod maskin intelligens.

For at finde nogen der tror på HAL i dag, må man gå til bøger a la Ray Kurzweils populære bestsellers "the age of spiritual machines". Ray Kurzweils budskab er her at maskiner faktisk vil blive ligeså intelligente som mennesker. Og mere til.

Længe inden computere når fuldt på højde med menneskelig intelligens vil kunstige dele til mennesker være vidt udbredte. Kunstige arme, fødder, hofter osv. er allerede trivielt. Derefter følger computere til behandling af f.eks. lyd data og plugget direkte ind til de rigtige neuroner i hjernen - så døve kan høre. Ud fra en sådan "Plug and Play" tankegang kan alle centre i den menneskelige hjerne efterhånden erstattes. Først vil man scanne de enkelte dele og så forsøge at "reverse engineer" dem. Derefter følger forbedringer til den menneskelige hjerne.

Og i den modsatte lejr har man lyseslukkerne, der med stor selvfølgelighed kan fremmane alverdens grunde til at HAL bare er fiktion. Jeg kommer i tanken om neurologen Susan Greenfield, der er kommet med mange indvendinger mod denne "det er let nok" attitude.

Ifølge hende vil problemet bl.a. være denne een til een mapning mellem nogle neuroner og en bestemt bevidsthedstilstand eller hjerne funktion. Hjernen er mere holistisk, hvor mange forskellige centre tilsammen hjælper til med at skabe en bestemt funktion. Som forskellige instrumenter spiller sammen i en symfoni. Der er ifølge hende ikke noget "neuralt center", "gen" eller "kemisk stof", der alene styre eller bestemmer en bestemt adfærd. Snarere kan man sige at disse ting kan bidrage i samspil med andet til en højere funktionalitet.

At samspillet er vigtigt kan man se på hvad diverse stoffer kan gøre ved hjernen. Man hører ofte at nogle får "bevidsthedsudvidende" oplevelser af at tage stoffer. Men stoffer er nu "bevidsthedsindskrænkende", ikke "bevidstheds-udvidende". Et rusmiddel går ud på at hæmme forbindelsen til personlige værdier, historie og sædvanlige måde at opfatte tingene på.

Kort sagt forståelse. Forståelse bliver hæmmet og ved indtagelse glider man ned mod den tilstand, der kan efterlignes af computere - at genkende ting, men uden at kunne sætte dem i en forståelses ramme. Det bliver sværere efter indtagelse at finde relevante associationer til en given situation. Hvilket igen gør det sværere at kunne tilføje mening til en given situation. Man bliver i stedet di-

rigeret af ydre oplevelser. Praktisk hvis man ikke ønsker at integrere oplevelser i fortid og fremtid! Endnu flere drugs begynder så at slå endnu flere ting fra. Så man f.eks. ikke kan holde balancen mere, så talen bliver sløret osv. Til sidst slås bevidstheden fra.

Svagere neural kommunikation. Som resulterer i restriktioner på de individuelle personlige associationer. I fjernelse af følelsen af ens eget selv. I fjernelse af abstrakt logisk tænkning. At der fokuseres på her og nu af ydre input osv. Fodboldfolket ved hvad jeg taler om.

At spille et enkelt instrument er svært. At få tingene til at gå op i en højere enhed, som i en symfoni, er rigtig svært.

Spørgsmål og svar.

HAL jubelåret havde selvfølgelig sin del af katastrofer, selvom det helt store nedbrud, vi har ventet på siden år 2000, stadig lader vente på sig.

D. 11 September var en sådan katastrofedag. D. 11 September var der iøvrigt også møde i datalogforeningen, hvor en af datalogiens grand old men, Kristen Nygaard, overbragte historier og visdom til os fremmødte. En af de historier jeg bed mærke i var historien om nogle norske studerende, der ikke skulle nyde noget af (tilsyneladende uakademiske) praktiske projekter. I hvert fald ikke noget der involverede noget så jordnært som bankkonti og den slags. Dvs. Altså, indtil de blev talt til rette, og fik forklaret at verden består af netop praktiske ting såsom bankkonti og lign.

Min fortolkning af historien var, at hvis man tror at man kan snyde eller springe over et eller sted i konstruktionen, så begynder tingene selvfølgelig at gå galt på en eller anden måde.

Arthur C. Clarke sammenfattede historien om HAL således: "If you understand 2001 completely, we failed. We wanted to raise far more questions than we answered."

Inspiration.

HAL's Legacy. - 2001's computer as dream and reality.

Edited by David Stork.

MIT Press. 1997. ISBN 0-262-19378-7

Software Development. January 2001.

"2001 - how close are we to creating HAL?"

The age of spiritual machines. - When computers exceed human intelligence.

Ray Kurzweil.

Penguin books 2000. ISBN 0 14 02.8202 5

The feeling of what happens. -Body, emotion and the making of consciousness.

Antonio Damasio.

Vintage. 2000. ISBN 0 09 928876 1