

Hjernen i computeren – Computeren i hjernen

Af Benny Lautrup og Lars Kai Hansen

Hjernens årti

I 1989 udnævnte USA's daværende præsident George Bush og den amerikanske kongres halvfemserne til hjernens årti, og den amerikanske beslutning om at satse store midler på hjerneforskning blev hurtigt fulgt op af tilsvarende beslutninger i Europa. Disse initiativer har medvirket til at hjerneforskningen i disse år producerer stadigt voksende mængder af nye data om hjernens funktion. Den amerikanske neuropsykolog Antonio Damasio påstår ligefrem i bladet *Scientific American* fra december 1999, at vi har lært mere om hjernens funktion i halvfemserne end i hele den forudgående periode op til 1989. Damasios egne resultater inden for bevidsthedsforskning er et godt eksempel på det nye billede af hjernen, der tegner sig. Ved detaljeret analyse af ældre og nye sygehistorier og ikke mindst ved brug af hjerneskaning har han påvist, at følelseslivet spiller en afgørende rolle for vores evne til at tage beslutninger. Patienter med skader i frontal-lappen, den store foldede del af hjernen bag pandebenet, får karakteristiske personlighedsforstyrrelser, der bedst kan beskrives som mangel på følelsesliv. Man skulle måske så tro, at de ville blive 'kolde' og beregnende individer, men det mest karakteristiske ved deres adfærdsændringer er faktisk, at de har meget svært ved at træffe beslutninger, og simpelthen bliver forvirrede. Uden følelser ser man måske verden i et for klart lys. Patienterne fornemmer, at de ikke har tilstrækkeligt grundlag for at træffe en 'rigtig' beslutning, og de syge hjerner kaster sig ud i endeløse diskussioner med sig selv.

Billeder af den arbejdende hjerne

Hjerneforskningen har tidligere været fokuseret på hjernens funktion på det cellulære niveau, de såkaldte neurale netværk. I halvfemserne har man for eksempel fuldstændig kortlagt visse rundormes og iglers nervesystem. Dele af de langt mere komplicerede nervesystemer i højerestående dyr, som for eksempel primaternes synsapparat, er efterhånden også forstået, men der er langt til at opnå et cellulært matrikelkort over menneskets hjerne. Ved at lytte med på grupper af nerveceller er der opnået viden om deres signalering, informationsbehandling og arbejdsdeling. Ligeledes er der opbygget omfattende modeller til at simulere nervesystemer i computere. I sådanne modeller, kunstige neurale netværk, efterlignes nervesystemets evne til indlæring, men modellerne er også blevet brugt til i helt andre sammenhænge at løse komplicerede tekniske problemer. Den danske forskning i kunstige neurale netværk har især fundet anvendelser inden for det biomedicinske område. Den hastige ingeniørmæssige udvikling af nye målemetoder og analyseteknikker er en afgørende faktor i de mange nye gennembrud i forståelsen af hjernen. Hjerneskaningen fik således et kraftigt skub fremad i 1991-92 da tre tæt konkurrerende amerikanske grupper opfandt den revolutionerende teknik, som kort betegnes fMRI (funktionel-magnetisk-resonans-billeddannelse). Med fMRI kan man bogstaveligt talt se ind i den levende,

arbejdende hjerne. Det er dog ikke nervecellernes informations-behandling, der synliggøres direkte, men derimod den øgede blodgennemstrømning, som er en følgevirkning af informationsbehandlingen.

Med hjerneskaning undersøger man områder ned til ca. en mikroliter - et område der er en millimeter på hver led. I et sådant volumen kan der være helt op til en million nerveceller, og informationsbehandlingen er ganske kompleks. Med fMRI kan sådanne områder studeres med en tidslig nøjagtighed på cirka ét sekund. For at opnå mere detaljeret viden om det tidslige forløb, må man supplere med måling af elektriske hjernebølger (EEG). Desværre er det vanskeligt at afgøre præcist, hvor i hjernen et givet EEG signal stammer fra, og det forsøges i disse år at samkøre fMRI og EEG for at opnå den kombinerede effekt af både rumlig og tidslig nøjagtighed. Ved brug af hjerneskaning er det blevet opdaget, at blodgennemstrømningen ved forestillingsbilleder med lukkede øjne, er næsten identisk med den, der opstår, når man har åbne øjne. Det betyder, at hjernen så at sige kan fremvise erindrede billeder i sit eget synsapparat, en slags indre biografforestilling. Ved måling på sovende forsøgspersoner har man fundet, at hjernens tilstand under drømmesøvn faktisk ikke adskiller sig fra, hvad man måler, når hjernen er vågen. Under drømmesøvn fortsætter hjernens informationsbehandling rent internt og man kan undre sig over, at hjernen spilder energi på informationsbehandling, som ikke ser ud til at være direkte relevant.

Hjernens funktion er at behandle information

Det interessante ved hjerneceller - frem for andre celletyper i kroppen – er, at de i deres netværk kan repræsentere information om omverdenen, herunder kroppen. De lysfølsomme celler i øjet og de lydfølsomme i øret er måleinstrumenter, som sender deres målinger videre til behandling i hjernebarken. Disse signaler repræsenterer verden direkte på analog form, i modsætning til digital eller symbolsk form. En mere interessant informationsbehandling finder sted i de højere, abstrakte, niveauer i hjernebarken, som er et tyndt lag af nerveceller i de yderste millimeter af hjernens overflade. I de visuelle områder bag til i hovedet dannes der en række afbildninger af den scene, som hjernen er i gang med at analysere. I et område lægges der for eksempel vægt på dybdeafstand mellem objekter, i et andet lægges der vægt på farveforskelle, i et tredje er det kanter og hjørner. Opdeling af billedet og genkendelse af objekter finder ligeledes sted i hjernebarken.

Hjernebarken har to funktioner, analyse og syntese. Hjernen analyserer data gennem opsplitting i komponenter, for eksempel dybde og farveforskelle, og disse komponenter fortolkes ved sammenligning med data fra tidligere oplevelser. Efterhånden som flere og flere neuroner drages ind i databehandlingen bliver repræsentationerne mere og mere ”abstrakte”. Desuden bliver komponenterne mere og mere ”digitale”, forstået på den måde at informationen bliver opfattet som bestående af skarpt adskilte kategorier: mand eller kvinde, indenfor eller udenfor, mængde eller individ, og så videre. I computersprog kaldes dette mønstergenkendelse eller klassificering. I sidste ende er en beslutning altid digital, for enten handler man, eller også lader man være. I disse højere områder af hjernen findes der for eksempel netværk, som er specialiseret til at afgøre, om et objekt er naturligt eller menneskeskabt eller til at genkende ansigter. Det forunderlige er, at hjernen ikke bare kan genkende og reagere på ydre påvirkninger, den kan som nævnt også lave ”snyde-

data” og danne sine egne billeder, lyde og fornemmelser. I moderne ingeniørpraksis anvendes ofte simuleringer, som prøver at efterligne nogle egenskaber ved virkeligheden.

På samme måde kan vores hjerne simulere data i den indre ”biograf”. Dette kunne være en mulig forklaring på spørgsmålet om, hvorfor hjernen spilder energi på at drømme. Det kan være, at vi i drømmesøvnen simulerer mulige fremtidige scenarier og danner strategier, som skal hjælpe os til at begå sig i dem. I Damasio's bog om bevidsthed ”Descartes Error” påstår han, at det er bevidsthedens rolle at skabe fremtiden. I vores indre film forestiller eller drømmer vi os mange mulige fremtidige scenarier, og det er disse simuleringer, som giver os evnen til at kortlægge, forstå og begå os i verden.

Damasio taler om, at Selvet simpelthen er forestillingen om ejerskabet af filmen. ”Dette er min film, og min film kan indeholde en scene af dig og den scene, som jeg tror, du spiller i din biograf, herunder hvad min film tror, at din film tror, der foregår i min film “. I en given ”film” er der bogstaveligt talt en hovedrolle, nemlig den som spilles af den krop og hjerne, som skaber filmen. Asymmetrien mellem denne rolle og alle de andre roller i filmen er årsag til vores opfattelse af individualitet og personlige egenskaber. Hvis asymmetrien bliver for stor bliver den til ensomhed og i patologiske tilfælde til mentale sygdomme som autisme.

Et lille tankeeksperimentet kan illustrere hvordan filmen spiller: Forestil dig at du spiller kort med dine venner. Hvor mange ”lag” af film har du egentlig kørende i den indre biograf, når du tager pokerfjæset på.

Computere med personlighed

Alletiders film-computer, HAL fra Kubricks film 2001, blev i filmen ”født” i 1997. I anledning af fødselsdagen udgav den amerikanske fysiker David Stork en bog, hvori han lod en række af USA's førende computereksperter beskrive, hvor langt de er kommet med at realisere en computer med ”personlighed”. Der er faktisk flere områder, som computeren allerede har indtaget. HAL kan forstå tale og selv udtrykke sig i følelsesfuld tale. HAL kan se, og som bekendt er det en vigtig pointe i filmen, at HAL selv lærer sig mundaflæsning. Men talegenkendelse, billedbehandling og mundaflæsning er faktisk blevet implementeret på PC'er i dag! Næsten alle bogens eksperter er optimister og gør detaljeret rede for de mange fremskridt i deres respektive områder. Men de vender tilbage til et fælles problem. Computeren kan genkende ordene og se objekterne, men den kan ikke forstå dem. Ordet ”forstå” skal her forstås som det at danne en forbindelse mellem det genkendte ord og en indre model af verden, altså den sammenhæng ordet optræder i.

Hjernen kan ikke bare genkende og relatere indtryk til en model af verden (den indre film). Der er også et afgørende feedback, så den basale måling og databehandling indstilles efter, hvad hjernen for øjeblikket ”tror”, der foregår. Vi ser det, vi tror vi skal se. Dette situationsafhængige feedback gør hjernens databehandling enormt følsom og dynamisk, og kan slet ikke realiseres i computere endnu. Måske kræver det, at fremtidens computere er drømmere, der som os blandt andet bruger natten til at simulere og planlægge.

Den 11. maj 1997 blev verdensmesteren i skak, Gary Kasparov, for første gang besejret af IBM's computer, Deep Blue, i en regulær turnering. Hans umiddelbare reaktion var vrede og frustration og en kort overgang en påstand om, at det ikke var maskinen, men mennesker bag den, der havde besejret ham. Det blev dog hurtigt klart, at denne påstand var uholdbar, fordi han var jo faktisk verdens bedste menneskelige skakspiller, så hvem skulle IBM dog have benyttet? IBM vandt reelt verdensmesterskabet i skak, men forstår på grund af programmets kompleksitet egentlig ikke hvorfor! Spørgsmålet er, om Kasparov forstår sine netværk af nerveceller bedre end IBM forstår sit program. Er han ikke lige så uvidende om, hvilke komponenter af hans hjerne, der førte til, at han tabte?

Hvordan kan man i det hele taget skelne mellem en computer og et menneske? Alle har oplevet skuffelsen over at blive mødt med en indtagende kvindestemme i telefonen, for derefter at konstatere, at det må være en maskine, fordi den ikke kan svare på selv de mest elementære spørgsmål. Dette er essensen af den test, som den britiske matematiker Alan Turing opstillede i 1950. En maskine må betragtes som svarende til en person, hvis det gennem verbal kommunikation på ingen måde er muligt at afgøre, at den ikke er en person. Turing's test er ikke ideel. Først og fremmest fordi man sagtens kan forestille sig personer fra andre kulturer eller andre planeter, som nemt kan skelnes fra det sædvanlige vesteuropæiske menneskebillede. Til gengæld har testen den fordel, at man ikke behøver at vove sig ud i fritsvævende definitioner af, hvad for eksempel intelligens er. Godt nok findes der fysiologiske størrelser, der har sammenhæng med intelligens, men de fleste egentlige intelligens-tests er dog temmelig kulturafhængige.

Hjernen er både analog og digital

Man siger ofte, at den væsentligste forskel på et menneske og en moderne computer er, at computeren er digital, mens mennesket er analogt. En digital maskine kan kun optræde i et endeligt antal tilstande, lige som et datostempel, der kun kan indstilles på endeligt mange datoer. Biler, cykler og de fleste af de maskiner, vi ellers benytter os af, er derimod analoge, hvilket vil sige, at deres bevægelige dele kan indstilles på uendeligt mange måder. Fordelen ved at benytte en digital maskine er, at dens tilstand kan kopieres fejlfrit. Dette gælder også for en digital meddelelse. Digitalt er simpelthen ensbetydende med stabilitet, og det har lige siden computerens barndom været en afgørende faktor, netop fordi vi ikke er forvænt med denne præcision hos os selv. Set fra et fysisk synspunkt er mennesket nemlig et analogt system med uendelig mange indre tilstande. Men vi betjener os på den anden side af sproget med et endeligt antal ord, og vi kan kun nå at ytre et endeligt antal ord i løbet af vores liv. Så fra dette synspunkt kan vi på et givet tidspunkt kun være i et endeligt antal sproglige tilstande, karakteriseret ved rækkefølgen af alle de ord, vi har ytret overhovedet indtil dato.

Sprogligt set er vi derfor digitale. Vi kan ikke sprogligt udtrykke noget, som en digital computer ikke kan bringes til at sige, fordi digitale computere er universelle, hvilket blandt andet vil sige, at de kan frembringe enhver rækkefølge af ord. En tilstrækkelig stor digital computer vil altid kunne bestå den rene Turing-test. Den skal bare have en tabel, der indeholder svaret på ethvert tænkeligt spørgsmål, eller bedre: på enhver rækkefølge af spørgsmål og svar. Det er dog kun overfladisk, at der er en stor afstand

mellem det analoge menneske og den digitale computer. På samme måde som visse sider af mennesket er digitale, så har den digitale computer altid en underliggende, fysisk, analog side. De mønstre af nuller og ettaller, som en computertilstand består af, opnås fysisk oftest i form af elektriske spændinger, der i princippet kan have uendelig mange indstillinger. Men computeren er konstrueret således, at den inden for et ret snævert område af temperatur, fugtighed og magnetfelt er i stand til at opretholde en skarp adskillelse - for eksempel et par Volt - mellem nuller og ettaller. Hvis denne adskillelse svigter, giver det anledning til fejl, som afslører det underliggende analoge niveau. Det samme kan såmænd også ske for mennesket, hvis alkoholmængden ("fugtigheden") overstiger et par promille, så sproget bliver til snøvlende fuldmandssnak.

Den digitale computer er deterministisk, hvilket betyder, at når den befinder sig i en bestemt tilstand af nuller og ettaller, vil den efterfølgende tilstand være entydigt bestemt. Ethvert tilløb til indeterminisme undertrykkes med hård hånd af elektronikken. Viser indeterminismen virkelig sit ansigt, bliver maskinen straks bragt til reparation. Når processoren brænder af eller harddisken ryger, er der meget lidt tilbage af den pålidelige determinisme, sælgeren pushede. En død maskine er selvfølgelig vældig pålidelig, men det var ikke lige det, den blev anskaffet til. Den digitale side af mennesket, sproget, er formodentlig ikke deterministisk, men afhænger af kroppens underliggende analoge fysiske tilstand. Undertrykkelsen af kroppens analoge tilstand er bestemt ikke fuldstændig og kendskab til alt, hvad et menneske har sagt og hørt af ord, er givetvis ikke tilstrækkeligt til at forudsige det næste ord. Dette strider ikke mod, at en digital computer kan bestå Turing-testen. Hvis indeterminismen er vigtig for det menneskelige særpræg, kan man meget let indbygge et vist mål af indeterminisme i en digital computer.

Disse betragtninger har måske noget at gøre med den fri vilje, som vi mennesker mener at have. Der er intet i fysikken, som giver plads til en bevidst frihed til at vælge, så hvor kommer den følelse fra, hvis det ikke er en ren illusion? Vi kommer tilbage til dette spørgsmål til slut i artiklen.

Neurale netværk

En menneskehjerne består af omkring 100 milliarder nerveceller. Hver af disse er forbundet med omkring ti tusinde andre nerveceller, hvilket giver en million milliarder forbindelser i hjernens neurale netværk, og en total ledningslængde på over en million kilometer. Gennem disse forbindelser, kaldet synapser, kommunikerer de adskilte nerveceller hen over et lille gab. I synapserne er der en kemisk aktivitet, som kan ændres af de signaler, der passerer igennem, og dette danner grundlaget for vores permanente hukommelse. En nervecelle består i almindelighed af et stort antal inputkanaler og en enkelt outputkanal. Groft set kan en nervecelles funktion forklares ved, at de indkommende signaler samles sammen over et tidsrum på nogle millisekunder, og hvis det samlede signal overstiger en vis tærskelværdi, så udsender nervecellen selv et signal, ellers gør den ikke. Det er denne proces, der er fundamentet for hjernens informationsbehandling, og som simuleres i kunstige neurale netværk.

Den typiske tidscyklus for en nervecelle ligger på lidt over et millisekund, altså en tusindedel sekund. Ingen nervecelle kan derfor udsende mere end tusind signaler per sekund. Hvis man ønsker at simulere hjernen i realtid i en digital computer, så skal hver simuleret synapse "besøges" af programmet mindst tusind gange i sekundet. Det betyder, at programmet skal udføre en milliard milliard "besøg" per sekund. Til sammenligning kan den hurtigste moderne PC maksimalt udføre en milliard operationer per sekund, så hjernen er mindst en milliard gange hurtigere end en PC, og måske endda meget mere, fordi en synapse er et kompliceret biokemisk system, som kan kræve et længerevarende "besøg".

Mange vil påstå, at en maskine aldrig vil kunne simulere en person. Men den konklusion kan vi faktisk ikke drage på grundlag af kendskab til nutidens computere. Selv om en PC virker aldeles uintelligent, så er det svært at forudsige, hvordan en milliard samarbejdende PC'er vil virke. At generalisere med en faktor en milliard er faktisk ganske utilladeligt. Det ville svare til at påstå, at når man ved, hvad man kan få for en krone, så ved man også, hvad man kan få for en milliard! Eller til at forudsige, hvad menneskeheden kan finde på, ud fra kendskabet til sine seks nærmeste venner. Ydermere er der et samspilsproblem. En milliard millimeterstore rundorme har det samme antal nerveceller som Mozart's hjerne, men kan sikkert ikke skrive en symfoni.

Hvis den nuværende vækst i computerhastighed fortsætter cirka 45 år endnu, vil hjemmecomputeren have en beregningskapacitet, der tillader realtidssimulering af menneskehjernen i software. Det kan godt være, at det noget tidligere bliver muligt at bygge speciel hardware, eller at udnytte internettets samlede computerkapacitet, til at simulere hjerneagtige systemer. Hjernen vil imidlertid aldrig kunne simuleres i realtid i en seriel computer af samme type som vores nuværende PC. Dette kan vises ved et simpelt og fundamentalt argument, der benytter både kvantemekanik og relativitetsteori. Argumentet går på følgende måde. En seriel simulator skal, som vi har argumenteret for, mindst udføre en milliard milliard operationer per sekund. En operation foregår derfor i så kort et tidsrum, at lyset kun når at bevæge sig et stykke på størrelse med et enkelt atom. Det betyder, at computeren må være mindre end et atom for at kunne gennemføre en fuldstændig tilstandsændring. Både højre og venstre side skal helst være enige om, at operationen er udført, og det tager ifølge Einstein mindst lyspassagetiden. Men atomer er på grund af kvantemekanikkens begrænsninger forholdsvis simple systemer, der ikke kan foretage vilkårlige digitale beregninger. Konklusionen er derfor uundgåeligt, at enhver computer, der kan simulere hjernen i realtid, nødvendigvis må benytte sig af parallel processing, hvor mange computere samarbejder.

Fri vilje?

De største computere i dag er også parallel-computere, med flere tusinde processorer, og deres kollektive hastighed når op på omkring tusind milliarder operationer per sekund. Dette er stadig en million gange mindre end det, der kræves for at simulere hjernen. I parallelt virkende computersystemer er kommunikation mellem processorerne det allervigtigste problem. Det er faktisk ganske umuligt i dag at nå til en forbindelsesgrad som i hjernen, hvor hver nervecelle som nævnt har ti tusind forbindelser. Et andet alvorligt problem er synkroniseringen af de mange processorer.

Opgives synkroniseringen i parallelsystemer, vil der kunne optræde tilfældige variationer i beregningerne, fordi processorerne kommer lidt ud af takt. Det kan måske til gengæld give anledning til, hvad vi vil kalde fri vilje.

Skakcomputeren, Deep Blue, havde en sådan fri vilje. Op til 32 hurtige specialprocessorer samarbejdede om at udføre beregningerne, og hver processor blev tildelt sin del af beregningsarbejdet. Den processor, der havde det bedste forslag til næste træk, fik lov til at flytte brikken. I en skakturnering spilles der på tid, så når et træk nødvendigvis skulle udføres, var man nødsaget til at afkræve processorerne et svar. Disse svar kunne variere en smule afhængigt af processorernes og deres omgivelers fysiske tilstand, især temperaturen. Selv om hver af processorerne til slut ville komme til et entydigt resultat, tog det ikke nødvendigvis den samme tid. Så det var ikke givet, at den samme processor ville vinde konkurrencen om at få lov til at flytte brikken, selv om udgangspositionen var den samme. Når Deep Blue bagefter analyserer partiet under mindre tidspres, vil den formodentlig kunne afgøre, om den valgte rigtigt. Den vil lige som et menneske hævde, at den kunne have valgt anderledes, og måske også hævde, at det trukne træk var det bedste på det tidspunkt. Vi ved imidlertid, at dens "valg" var betinget af tilfældige variationer i de fysiske omgivelser ganske uden for dens kontrol. Er vores situation i grunden så forskellig? Der er som sagt intet i fysikken, som giver plads til menneskets bevidste frihed til at vælge.

Nu har menneskets fri vilje faktisk to sider. For det første kan vi ikke gøre rede for den fysiologiske proces, der fører til en beslutning. Hjernens dynamik er som de fleste ikke-lineære dynamiske systemer kaotisk og ikke til at forudsige ret langt frem - i lighed med vejret. Dette giver en oplevelse af uforudsigelighed, både når vi vekselvirker med andre og med os selv, selv om den underliggende kaotiske dynamik strengt taget er deterministisk. Hertil kommer den indeterminisme, som stammer fra for eksempel varmebevægelser i hjernen og som i Deep Blue's tilfælde førte til "valget". Måske allervigtigst er der en uundgåelig og principiel kvantemekanisk indeterminisme i de basale atomare processer. Hjernen både forekommer og er indeterministisk, og vi har som Deep Blue efterfølgende en klar oplevelse af, at "kunne have gjort noget andet", når vi handler. For det andet har vores hjerne en ubændig trang til at skabe forklaringer på ethvert forløb. Vi anerkender simpelthen ikke tilfældighedernes spil, men forsøger at finde en underliggende forklaring på, hvad der end sker. Gennem evolutionen har det været en kæmpefordel at kunne aflure årsagssammenhænge, at tænke "baglæns". Men det har også den ulempe, at vi ofte finder forklaringer, hvor der ikke er nogen.

Fra et naturvidenskabeligt synspunkt må oplevelsen af fri vilje altså ses som et samspil mellem tilfældige hændelser og søforklaringer. Der er ikke plads til, at ånd kan påvirke stof gennem viljen, som ved telekinese, der på trods af mange forfinede eksperimenter aldrig er blevet påvist. Mennesket er så fuldt af søforklaringer, at samfundet må ty til sindrigt udtænkte retssystemer, hvor kun faktuelle beviser duer, for at skille fup fra fakta. Men hvis frihed til at handle er en illusion, hvem er det så, vi straffer i retssystemet? Straffer vi i virkeligheden mennesker for deres gener og omgivelser og de tilfældigheder, der har afgjort deres såkaldte "frie valg". Noget kunne tyde på det. Samfundet undskylder tit kriminalitet med dårlig opvækst eller sindssyge i gerningsøjeblikket. Men dette synspunkt efterlader os med et problem: Hvem har egentlig besluttet at skrive disse ord?