

Människan har alltid drömt om att skapa en robot som skulle kunna tänka rationellt och logiskt, gråta, skratta, älska och vara medveten om sig själv.

AV HANS LILJENSTRÖM

Jag förlorar medvetandet. Jag är rädd, säger datorn HAL oroligt i filmen 2001 när besättningen på rymdskeppet Discovery långsamt kopplar ner den (bild 1). Det är snart bara fyra år kvar till år 2001, men hur sannolikt är det att vi kommer att ha tänkande datorer likt HAL inom så kort tid – om ens någonsin? Hur långt har forskningen och tekniken kommit i dag?

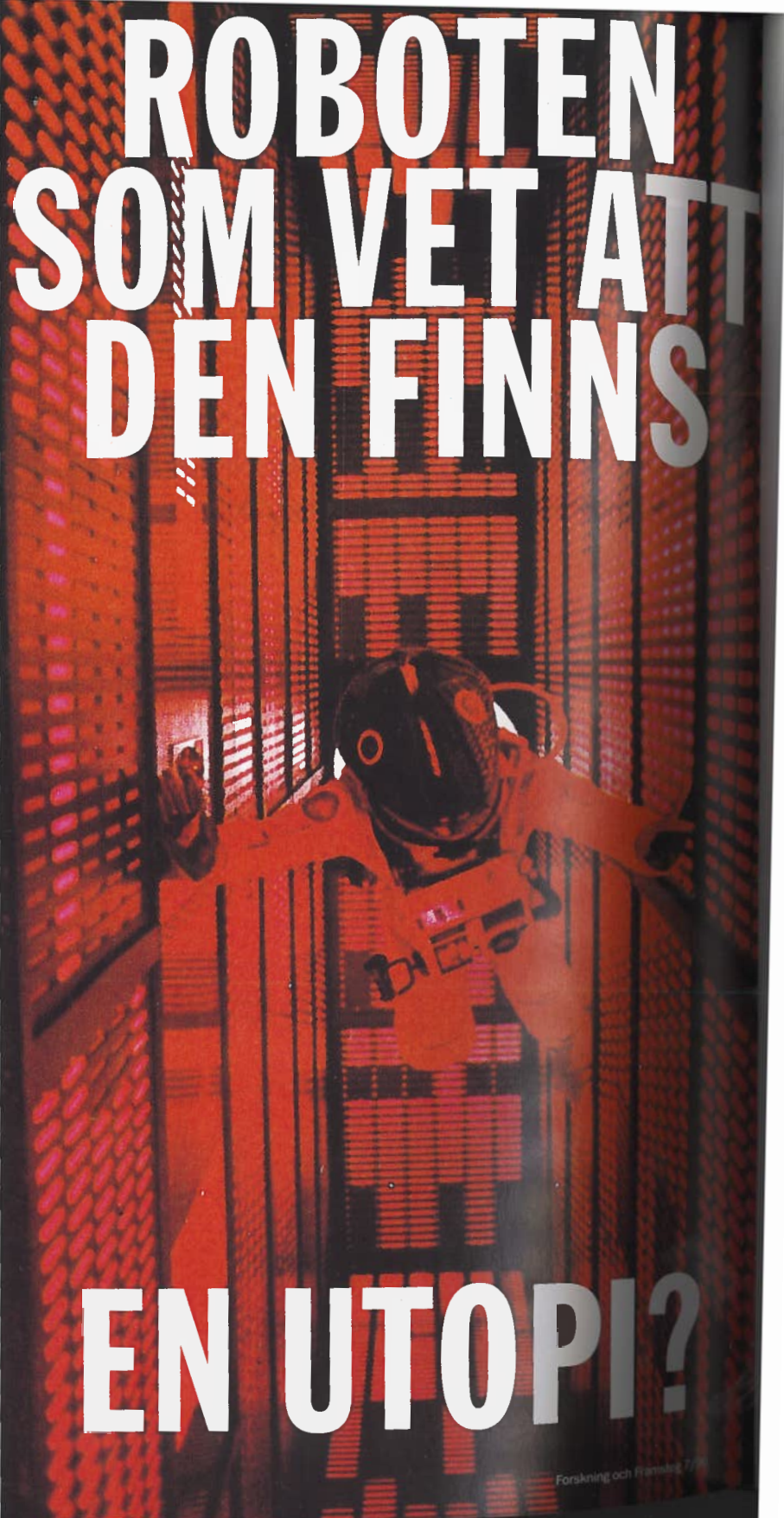
Robotar som "tänker"

Det närmaste man kan komma i den vägen i dag är förmodligen roboten Cog (bild 2) som utvecklas av dataforskaren Rodney Brooks grupp vid ett av USAs mest framgångsrika forskningscentrum, Massachusetts Institute of Technology. Den är tänkt att bli mer och mer människolik, genom att ha "mänskliga sinnen" och kunna reagera på ett "mänskligt" sätt för olika sinnesintryck. Denna robot lär sig genom växelverkan med människor och föremål i dess omgivning, på liknande sätt som ett litet barn. Brooks har tidigare varit med om att utveckla insektsliknande robotar – "bugbots" – med flera "sinnesorgan" och med förmåga att förflytta sig med ben.

En annan typ av tänkande datorsystem är den amerikanske forskaren Douglas Lenats CYC (som i ency-

ROBOTEN SOM VET ATT DEN FINNS

EN UTOPI?



clopedi) och schackdatorn Deep Blue, som närapå slår världens främsta schackmästare (bild 7). Här försöker man gå förbi alla utvecklingssteg i evolutionen och i ett barns uppväxt (eller i en schackspelares mångåriga erfarenhet). I stället utnyttjar man redan känd kunskap som programmeras in i datorn.

I Sverige pågår också forskning för att på ett eller annat sätt efterlikna hjärnans funktioner. På Kungliga Tekniska Högskolan (KTH) i Stockholm finns tex prototypen till en talande turistguide som ska kunna ge upplysningar och svara på muntligt ställda frågor om tidtabeller och annat i anslutning till skärgårdstrafiken. Där finns även seende robotar och datorsystem som kan minnas och urskilja mönster, på ett sätt som liknar det man tror att människo-hjärnan använder. Den mest imponerande minnes- och beräkningskapaciteten har dock paralleldatorn Stindberg vid KTH, den största datorn i sitt slag utanför Sverige.

Men hur är det egentligen med tänkandet? Kan man säga att dessa datoriserade system kan tänka och, i så fall, vilket slags tankar skulle de kunna ha? Är det över huvud taget möjligt att beskriva tänkande i matematiska formler, som datorer kan använda?

Bickemin är grunden

Alla försök hittills har gällt att kopiera eller simulera en viss aspekt av hjärnans funktioner, som tex att känna igen ett ansikte eller talade ord, att kunna spela spel, syntetisera tal eller styra en robotarm. Men man har ännu inte lyckats föra ihop ett flertal sådana funktioner till en självständig samverkande enhet, även om roboten Cog ingår i ett sådant försök.

Kanske är det just förmågan att integrera många olika funktioner och uppleva det som en enhet, ett subjekt eller jag, som är det mest typis-

PETER MENZEL/INA



[2] Amerikanen Rodney Brooks, en av världens främsta forskare i tänkande datorer, är i full fart att skapa Cog – en tänkande dator med mänskliga sinnen (upptill). På bilden nedtill ser vi Brooks med sin förra skapelse, en insektsliknande robot.

ka för mänskligt tänkande. Men skulle Cog, eller något annat konstgjort system, kunna uppleva sig som en enhet? Vad innebär det egentligen att tänka, att äga mentala funktioner och intelligens – och att vara medveten om det?

Det är svårt att definiera vad tänkande är, för att inte tala om intelligens och medvetande. Det som vi normalt menar med tänkande är i alla fall inte enbart en logisk serie av steg enligt enkla regler, såsom i schackspel, eller en kombination av minnen och föreställningar i givna sekvenser. Mänskliga mentala funktioner inbegriper också nyfikenhet, humor och kreativitet. Ofta räknas även subjektivitet, dvs att vi har medvetna upplevelser, som en del av vårt tänkande.

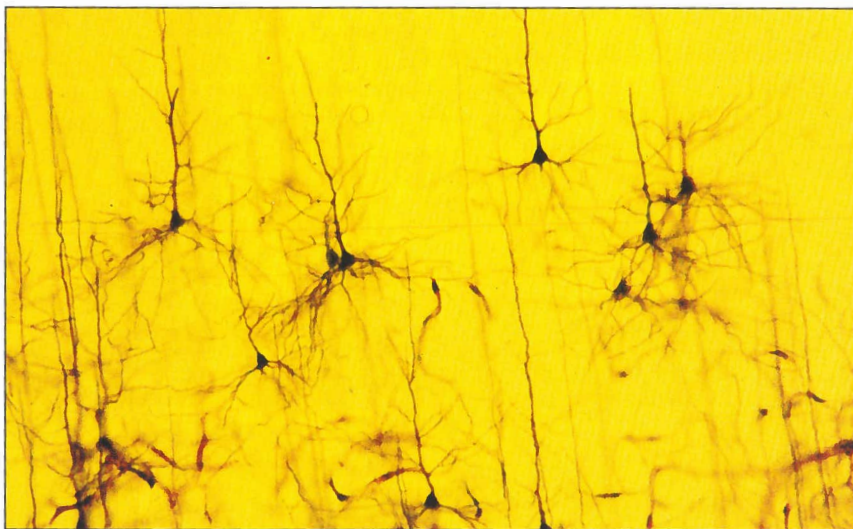
Skulle det räcka med att kopiera vad som sker i och mellan nervcel-

lerna i hjärnan, eller är hjärnans biologiska materia också viktig för att åstadkomma mentala processer? Går det att ersätta de kemiska och biologiska byggstenarna i hjärnan med kisel, dvs med elektroniska kretsar och komponenter, för att åstadkomma tänkande?

Många hävdar att det är just bickemin i levande celler som är nödvändig för tänkande – åtminstone ett tänkande som liknar människans. Därför måste ett konstgjort tänkande system också bygga på biomolekyler, menar man.

Andra anser att om man bara kan efterlikna hjärnans arbetssätt, främst genom att följa principen för hur nervceller kommunicerar med varandra, så bör också de mentala egenskaperna följa med. Tänkande ses då som ett dataflöde som lika gärna kan ske i datorkretsar som i

Konstgjorda neuronnät lär sig att hitta



HARVARD MEDICAL SCHOOL

3 Hjärnans väv av "grå celler" ska efterliknas med de artificiella neuronnäten. Redan i dag används neurala nätverk bl a till att finna mönster i stora mängder data som t ex fysiker eller astronomer samlat på sig och till att ställa feldiagnoser i komplicerade tekniska system och i medicinska sammanhang.

nervcellskretsar. Kanske det viktigaste ändå är hur det tänkande systemets materia, vare sig det gäller biomolekyler eller kisel, är organiserad och strukturerad. Många menar att det är strukturen på systemnivå, dvs hur nervcellerna kommunicerar med varandra, som är avgörande för de mentala processerna. Om det gick att härma hjärnans nervnätstruktur med kiselkomponenter kanske man också skulle kunna få konstgjorda mentala processer (och medvetande?) – men som eventuellt är helt annorlunda än människans.

Allvetande datorer

Tanken att kunna skapa konstgjort tänkande, eller artificiell intelligens, AI, är minst lika gammal som den första datorn, Eniac, vilken i år fyller 50 år (bild 6). De första försöken att mekanisera mentala processer gjordes dock i samband med de mekaniska räknemaskiner som Blaise Pascal, Gottfried Wilhelm von Leibniz och andra konstruerade redan på 1600-talet. Den moderna

datorn bygger dock på matematiska teorier som utvecklades av bl a John von Neumann i USA och Alan Turing i Storbritannien på 1930- och 40-talen. Härav har vi fått begreppen von Neumann-maskiner och Turing-maskiner. Dessa utgör grunden för dagens datorer där logiska processer utförs i sekventiella steg och där alla tal uttrycks i nollor och ettor.

Enligt Alan Turing kan en maskin sägas tänka om den klarar följande test: En (mänsklig) utfrågare sätts i ett rum där hon kan kommunicera via ett tangentbord med en dator eller en människa i ett annat rum. Om utfrågaren genom sina frågor inte kan avgöra om det är en dator eller en människa hon kommunicerar med, kan datorn sägas tänka (som en människa).

Detta är förstås ett mycket begränsat sätt att se på tänkande och på kommunikation, och mycket kan anmärkas på detta test. Även Turing själv var skeptisk till frågan "Kan maskiner tänka?". Likväl har det fungerat vägledande för dem som

velat utveckla "tänkande" datorer. Hittills har dock ingen dator klarat testet.

I datorernas ungdom fascinerades många av deras stora minnes- och beräkningskapacitet. Det fanns de som trodde att man skulle kunna få vanliga datorer att tänka som människor. Man liknade också hjärnan vid en stor dator. Så kallade expertsystem utvecklades, där en utfrågare kunde vägledas genom databaser för att t ex ställa en diagnos på en sjukhuspatient. Det "allvetande", tidigare nämnda programsystemet CYC utvecklas enligt denna klassiska AI-tradition.

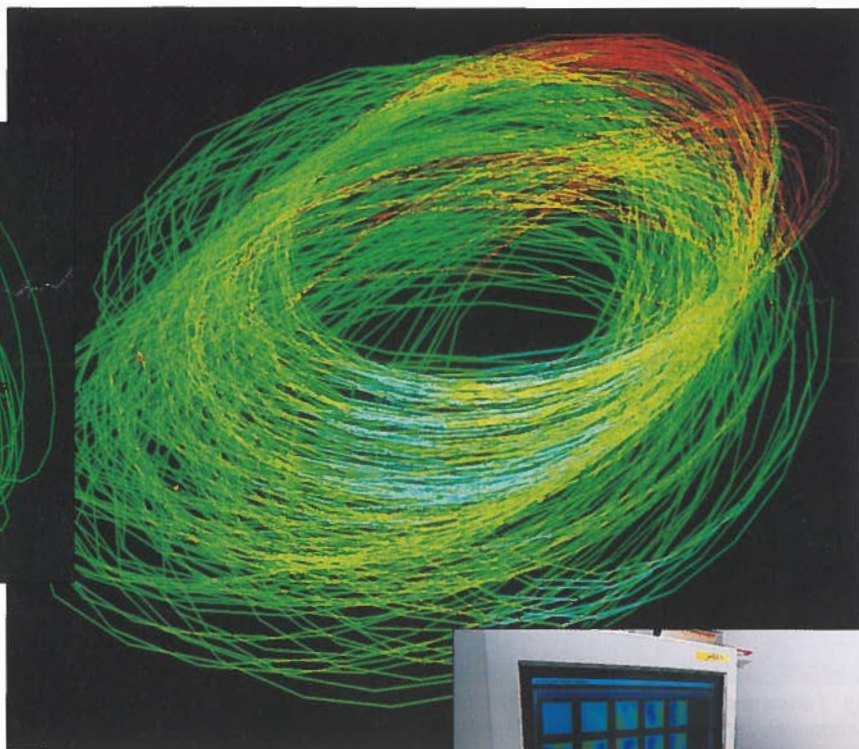
Det har emellertid visat sig att den vanliga datorns seriella arbetssätt, där varje steg i beräkningsprocessen är beroende av föregående steg, inte lämpar sig för alla typer av informationsbehandling. Speciellt inte för den typ av mönsterigenkänning som kännetecknar människors och djurs bearbetning av sinnesintryck. På motsvarande sätt har man också insett att minnen inte lagras i hjärnan på samma sätt som i en dator, dvs på bestämda adresser i ett register.

"Biologisk" dator

En annorlunda utveckling inom AI bygger mer på biologiska fakta om hjärnans struktur och arbetssätt. Det som framför allt kännetecknar hjärnans struktur är dess nätverk av sammankopplade nervceller, neuroner, vilka kommunicerar med varandra via långa utskott. Miljontals celler arbetar samtidigt, parallellt, med att bearbeta en viss information, och fastän det hela tiden dör tusentals hjärnceller märks ingen nämnvärd försämring av funktionen. Hjärnan är tydligen inte så beroende av de enskilda nervcellerna eller av att informationsbehandlingen sker i en enda serie steg. Detta skiljer sig markant från den konventionella datorns sätt att arbeta. Om



4 **Luktsystemet** i hjärnan uppvisar både vila (till vänster) och vid förnimmelsen av en bekant doft (till höger) ett kaotiskt beteende, dvs oordnat men med en viss bakomliggande struktur (dock mer regelbunden vid vittring än i vila).



WALTER FREEMAN, UNIVERSITY OF CALIFORNIA

en enda komponent, t ex en transistor eller en processor, skulle gå sönder så mister datorn hela sin funktion och en given beräkning kan inte utföras.

De första försöken att härma både struktur och funktion hos hjärnan kan man spåra till 1940-talet, och det första fungerande systemet, perceptronen, såg dagens ljus i början av 1960-talet. Det var dock först i början av 1980-talet som utvecklingen av de sk artificiella neuronnäten, ANN, tog riktig fart. Då hade John Hopfield vid California Institute of Technology utnyttjat egenskaper hos andra dynamiska system, vilka tidigare hade studerats inom fysiken.

Hopfield kunde visa att nätverk av enkla element, kopplade till varandra via konstgjorda synapser som varierar i styrka, kan fungera som associativa minnen. Kopplingsstyrkorna kan vara positiva eller negativa. Nätverkelementen kan i det enklaste fallet anta två tillstånd, att jämföra med nervcellernas aktiva och inakti-

va tillstånd. Varje element i nätet lägger ihop signalerna från andra enheter, multiplicerade med motsvarande vikter. Om summan överstiger ett visst tröskelvärde skickas en signal ut till alla andra nätverkelement.

Mönster, vilka kan representera t ex en bild eller en doft, kan lagras i dessa nätverk genom att vikterna ändras mellan de olika elementen. Om två element samtidigt är aktiva förstärks kopplingen mellan dem. Ett givet minnesmönster kan återkallas när ett nytt mönster, som kan vara en ofullständig och störd version av det lagrade minnet, presenteras för nätet. Det aktuella indatamönstret har då kopplats samman med det lagrade mönstret. Att denna teori verkligen fungerade kunde lätt visas med datorsimuleringar och med att de sk Hopfield-näten kunde förverkligas i elektroniska kretsar.

Förutom dessa återkopplade nätverk finns andra typer av ANN, t ex framåttmatade nät. Där går informa-

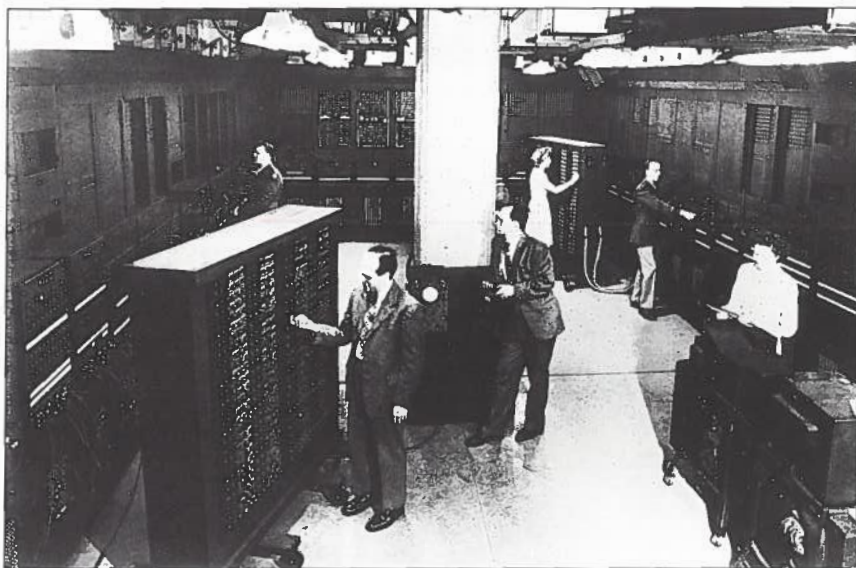


5 - I min egen forskning undersöker jag vilken roll svängningar, kaos och brus kan ha för neuronnätets funktion, säger artikelförfattaren Hans Liljenström.

tionsflödet i en bestämd riktning, från ett lager av noder i nätverket till ett annat, ofta i ett slags trelagersstruktur. Under inläringen ändras kopplingsstyrkorna när ett önskat svar och det svar som nätverket ger är olika.

Denna typ är inte särskilt biologiskt motiverad, men används i många tillämpningar. Nätverksstrukturen och sättet att förändra kopplingsvikterna vid inläring är det som främst skiljer de olika ANN-modellerna. Gemensamt är förmågan att hitta mönster och sammanhang i ofullständig och ostrukturerad

Svängningar, kaos och brus tycks vara gy



6 I datorernas ungdom fascinerades många av deras stora minnes- och beräkningskapacitet. Den första datorn, amerikanska Eniac, fyller 50 år i år.

rad information. Roboten Cog använder sig av denna teknik.

Kaos i hjärnan

De flesta ANN-modellerna arbetar med en relativt enkel dynamik, dvs aktiviteten går mot stabila stationära tillstånd, där ingen fortsatt förändring sker när ett visst minnestillstånd väl har nåtts. Den mänskliga hjärnan, däremot, kännetecknas av mycket komplex dynamik. Denna återspeglas bl a i den elektriska aktiviteten, vilken kan mätas med EEG, elektroencefalogram, som uppvisar både svängningar och ett kaotiskt beteende.

Det har länge varit känt att svängningar inom vissa frekvensband är kopplade till olika medvetandetilstånd, t ex vakenhet och olika grader av sömn. De tycks vara relaterade även till vissa sinnesförnimmelser och inlärning, kanske också till uppmärksamhet och medvetande.

Hur är det då med kaos? Om det existerar i hjärnan, vilken betydelse har det i så fall? Kaos i matematisk

mening är ett slags oordning men med viss struktur (bild 4). Ibland kan man finna en ganska enkel matematisk formel som ger upphov till det kaotiska beteendet. Genom att variera en given parameter kan ett system övergå från ett ordnat till ett till synes helt oordnat beteende. En kemisk eller elektrisk signal i hjärnan skulle på så sätt kunna reglera t ex dess känslighet och flexibilitet. Kaos skulle alltså kunna ha en roll på makroskopisk, storskalig nivå i hjärnan.

På en mikroskopisk nivå tycks det också finnas ett visst mått av oordning, men här mer strukturlöst, i form av brus. Man har t ex funnit att vissa nervceller avfyra sk aktionspotentialer spontant, dvs skickar ut signaler oberoende av yttre signaler. Dessutom är utsöndringen av signalsubstanser i synapskopplingarna mellan olika nervceller mer eller mindre opålitlig. Allt detta yttrar sig som brus och osäkerhet i nervcellernas alstrande och överföring av signaler.

I min egen forskning undersöker jag vilken roll svängningar, kaos och brus kan ha för neuronaltens funktion. Med en datormodell av lukt-hjärnan studerar jag dessa fenomen i samband med inlärning och minne. Stora nätverk med många tusen element och miljontals kopplingar simuleras på en sk Connection Machine, en stor paralleldator vid KTH. Resultaten tyder på att svängningar, kaos och brus, vilka normalt är oönskade i artificiella system, kan ha en gynnsam effekt för de neuronalas systemens funktioner.

Denna typ av modell, liksom andra som försöker fånga någon viktig del av hjärnans neuronalas strukturer och processer, kan göra de konstgjorda systemen mer lämpade för växelverkan med naturliga miljöer. Kanske kan de också föra oss närmare en förståelse av våra mentala funktioner och leda till att vi på konstgjord väg kan efterlikna vissa delar av vårt tänkande.

Något fattas ändå

Men har vi då nått vårt mål, har vi lyckats åstadkomma verklig intelligens? Är t ex kaos och brus nog för att ge systemet den flexibilitet och "spontanitet" som kännetecknar mänsklig intelligens?

Många menar att det krävs något mer. Vad är det då som fattas? Även om våra tankar ofta tycks följa bestämda banor och mönster, sker också mycket oförutsägbart och oväntat. Kreativitet, fantasi, intuition, "sunt förnuft", humor och känslor förefaller vara sådana aspekter av vårt mentala liv som det verkar svårt (eller omöjligt) att överföra till konstgjorda system. Det har också påpekats att datorer aldrig kan förstå meningen i den information som behandlas. Allt detta tycks vara förknippat med ett subjektivt medvetande.

Det är svårt att tänka sig att ett medvetande skulle kunna uppstå

samma för tänkande

spontant i en vanlig dator utan en avancerad växelverkan med omgivningen. För det krävs både någon form av sinnesorgan som kan "avläsa" omgivningen, och dessutom också ett motoriskt system som kan användas för aktivt utforskande av världen. Konstgjorda system kan, som vi har sett, ges motsvarande sinnessystem och motorik som liknar djurs, men det är inte detsamma som att vi lyckas ersätta den biologiska evolutionen eller en individs uppväxt i en omväxlande miljö.

Vissa, som den amerikanske filosofen Daniel Dennett, anser att det bara är en tidsfråga tills vi har artificiella system som äger samma mentala egenskaper som människan. Dennett, som också är inblandad i utvecklingen av roboten Cog, menar att alla stora problem redan är lösta. Enligt honom har vi nu förstått grunderna för alla medvetandeprocesser, och i princip skulle det vara möjligt att härma dessa processer med dagens datorer.

Datorn känner inte

En annan amerikansk filosof, David Chalmers, hävdar dock att vi måste skilja mellan lätta och svåra problem. De lätta problemen, som t ex inlärning och minne eller att känna igen ansikten, kan vi antagligen snart förstå i termer av neurobiologiska processer och även simulera på datorer. Vissa typer av logiskt tänkande skulle också vara möjligt att efterlikna. De svåra problemen däremot, de som rör förståelse, intuition och subjektiva upplevelser, kommer vi kanske aldrig åt - åtminstone inte med datorer och matematiska formler.

Den brittiske matematikern Roger Penrose anser att det aldrig kommer att gå att överföra den typen av medvetandeprocesser till datorernas språk. Skälet är att dessa processer inte är beräkningsbara, enligt Penrose. Endast i system där det



7 Boris Kasparov behöll sin världsmästartitel när han den 17 februari i år vann mot schackdatorn Deep Blue, som dock besegrade världsmästaren i ett och spelade oavgjort i två av sex partier.

existerar icke-algoritmiska processer kan medvetande uppstå, och hittills har vi endast funnit detta hos biologiska system, menar Penrose.

Vi är kanske på väg att förstå och efterlikna rationellt, logiskt tänkande, men det är långt kvar till en förståelse av medvetande och drömmar. Rationellt tänkande har en tydlig funktion för individens överlevnad, men vi vet inte vilken nytta subjektivt medvetande och drömmar kan ha för människan, än mindre skulle kunna ha för en dator. Kanske uppstår det subjektiva medvetandet naturligt vid en viss grad av komplexitet och organisation hos ett system. Så länge vi inte har löst dessa problem kan vi inte heller svara på om datorer skulle kunna tänka, drömma eller ha medvetande. Frågan är om vi ens skulle vilja ha sådana konstgjorda system.

Förhoppningen är att vi en dag ska få, om inte medvetet tänkande system, så i alla fall konstgjorda system som utför "logiskt tänkande", kan göra "smarta" förutsägelser och

kan fatta "kloka" beslut i människans tjänst. Dessutom kan forskningen om artificiell intelligens med neuronätsmodeller leda till en större förståelse av vad biologisk intelligens är - och inte är. Medvetandet, som vi känner det, kommer dock kanske aldrig att helt kunna förklaras. ■



HANS LJLJENSTRÖM

ÄR DOCENT I TEORETISK BIOFYSIK VID KUNGL TEKNISKA HÖGSKOLAN I STOCKHOLM. HAN UTVECKLAR DATORMODELLER AV HJÄRNAN OCH ANDRA BIOLOGISKA SYSTEM. FORSKNINGEN FINANSIERAS BL A AV FORSKNINGSRÅDSNÄMNDEN.