

hela taget, kan ha fört med sig mindre bra saker. Författarna har vissa poänger som lånar sig till den typen av analyser. Det finns möteskulturer motiverade av "god stämning" som innebär ett slöseri med arbetstid. Det finns rundgång i policydokument som uppehåller sig kring signalord men lämnar sakfrågorna därhän. Det finns ängslighet som står i vägen för resoluta beslut, överdrivna hänsyn som leder till handlingsförklaring. Men det är märkligt att knyta detta till "feminisering" och att inte försöka bryta upp förklaringen i mer potenta förklarande kategorier. Träffsäkerheten hade vunnit dels på ett snävare och mer noggrant beskrivet material, dels på större ödmjukhet. Som diagnos på ett vagt och tveksamt formulerat problem är det näst intill omöjligt att bedöma förklaringsvärdet.

David Brax

Kan AI tänka? Om människor, djur och robotar

Peter Gärdenfors. Fri Tanke 2024, 283 s. ISBN 978-91-8973-249-0

Det här är en bok som är välskriven och full av information om forskningsläget inom de områden som har att göra med tänkande och andra mentala fenomen. Författaren beskriver inledningsvis mycket utförligt, och med hänvisning till en lång rad experiment, nutida teorier om människors och andra djurs tänkande.

Bokens fråga – "Kan AI tänka?" – har enligt författaren inget bestämt svar. Det beror förstås på vad man menar med "tänka", och Gärdenfors vill också visa att det AI kan göra är något ganska annorlunda, något mycket mer avgränsat och specialiserat, än vad människor och andra djur kan göra när de tänker. Åtminstone än så länge.

Mycket har förstås att göra med hur hjärnan fungerar. Men allt tänkande sitter inte i hjärnan. Gärdenfors anser att hjärnan inte är chefen, utan en betjänt som tjänar kroppens intressen (s. 77). Det är tänkvärt.

Det han skriver om sinnesintryck, känslor, minne, beslutfattande, inlevelseförmåga etcetera är mycket informativt, men mer filosofiskt intressant blir det framåt mitten av boken, när han kommer in på *medvetandet*. Han konstaterar att här finns det stora frågor: "Hur uppstår medvetandet? Vad består det av? Vad är poängen med att ha ett medvetande?" (s. 133).

Vad är medvetande för något? Gärdenfors tycks anse att det är en kombination, en "enhet", av sinnesintryck och känslokomponenter (s. 135), men framför allt betonar han en kombination av "kärnmedve-

tande” och ”självmédvetande”. Han säger att kärnmedvetandet ger oss ”en känsla av att veta – vi får upplevelser av världen och av vad som händer med oss själva” och när detta blandas med minnen ”så får vi en känsla av ett sammanhängande jag med en serie av förflutna upplevelser och med förväntningar på framtiden” (s. 134f.).

Men här låter det då (enkelt uttryckt) som att medvetandet *ger oss* dessa upplevelser, samtidigt som det *är* dessa upplevelser. Det går inte riktigt ihop!

Kanske ska man fatta det så att medvetandet är dessa upplevelser. Men upplevelserna varierar ju från tid till annan; ska man då anse att vi har olika medvetanden – och kanske också olika ”jag” – vid olika tidpunkter? Eller ska vi ta fasta på tanken att jaget är en illusion (s. 66)? Kan kanske även medvetandet vara en ”illusion”, även om upplevelserna är verkliga?

Det filosofiskt mest intressanta är kanske inte medvetandet, utan just upplevelserna. Varför har vi (medvetna) upplevelser och hur uppstår de?

Och vad är det vi upplever? Gärdenfors säger att det är ”viktigt att inse att det vi upplever är inte världen i sig utan världen för oss”, det vill säga ”vår hjärnas konstruktion av världen” (s. 136).

Är det verkligen viktigt att inse detta? I mina öron låter det alltför skumt – även om idén kommer från en mycket berömd tysk filosof. Såvitt jag förstår är det en dålig idé.

Att den värld vi upplever är en konstruktion håller jag med om, men varför ska vi tro att det dessutom finns en ”värld i sig”, vad det nu är för något? Den värld vi lever i och som vetenskapen försöker beskriva är ju, enligt vad vi tror, den värld vi upplever. Har vi fel? Varför ska vi tro – eller ”inse” – att det också finns en annan värld, som vi inte ska blanda ihop med den vi upplever?

Att hjärnan konstruerar världen är inte så konstigt. Gärdenfors har redan tidigare påpekat att även enkla sinnesintryck kan innebära tolkningar. Hans referenser här till fantomsmärtor etcetera verkar helt onödiga. Elektrokemiska impulser i nervbanor ska ju på något sätt ge upphov till upplevelser. Det måste rimligen involvera någon sorts tolkning eller konstruktion.

Men att världen är konstruerad betyder ju inte att den inte finns eller att den är något annat än den värld vi lever i. All evidens vi har om världen är våra upplevelser. Vi anser att upplevelserna – åtminstone vissa upplevelser – kan ge en korrekt bild av det som upplevs. Vetenskapen bygger på observationer.

Hur uppstår då upplevelserna? Eller medvetandet? Det vet vi ännu inte. Kanske aldrig. Gärdenfors diskuterar fem teorier om detta.

För det första ”beräkningsteorin”, som är vanlig bland forskare som ägnar sig åt AI. Enligt den är medvetandet, eller upplevelser, en viss typ av beräkningar (s. 138). Det låter ju helt vrickat, men det bygger på att beräkningar i en dator bevisligen kan åstadkomma typer av beteende – samtal, schackspel etcetera – som även människor uppvisar. Men beteende är ju inte detsamma som upplevelser.

De övriga fyra teorierna behöver knappast nämnas här. De verkar alla helt otillräckliga, och Gärdenfors konstaterar själv att vi inte har någon aning om hur medvetandet uppstår och att vi därför inte heller vet hur vi ska avgöra om AI har ett medvetande (s. 142).

Men *varför* har vi upplevelser? Det kräver, som Gärdenfors påpekar, en evolutionär förklaring. Hur kan upplevelser öka våra chanser att få livskraftiga barn? En idé, som Gärdenfors tycks godta, är att vissa beteenden hos däggdjur och fåglar tyder på att de eftersträvar lustfyllda upplevelser. Men det förklarar knappast hur detta är evolutionärt fördelaktigt.

Man kan tänka sig, som Gärdenfors tycks göra (s. 146), att det att vi har upplevelser, och upplevelser av dessa upplevelser, är en förutsättning för att vi ska kunna uppleva att andra har upplevelser – och även vad de upplever. Detta kan i sin tur vara nödvändigt för socialt samspel. Vilket är evolutionärt fördelaktigt (s. 151).

Kanske är det så. Men är verkligen upplevelser nödvändiga för socialt samspel? Bin och myror tycks ju klara sig utan upplevelser (såvitt vi vet).

Hur medvetandet kan uppstå ur fysisk materia är en svår fråga. Men Gärdenfors anser att en ännu svårare fråga är hur medvetandet kan påverka materien (s. 149). Men hur vet han att medvetandet alls *kan* påverka materien? Han påpekar själv att många forskare tror att det inte kan det.

Gärdenfors noterar och förklarar att intelligensmätning är något mycket tvivelaktigt och han vill därför undvika begreppet intelligens. I stället vill han ”fråga hur man kan jämföra människors och artificiella systems förnuft – om det över huvud taget är möjligt” (s. 160).

Han bemöter också de idéer om ”superintelligens” som på svenska framförts av Nick Bostrom, Max Tegmark och Olle Häggström. Han menar att ingen inom området har några klara idéer om hur ett superintelligent AI-program skulle kunna konstrueras eller identifieras (s. 164).

Han tror inte heller så mycket på människors förnuft. ”Med det begränsade förnuft vi har går världen inte att förstå, och inte människan

heller. Du kan inte ens förstå dig själv” (s. 169). Det låter dramatiskt, och lite väl överdrivet. Jag tycker inte att han har visat det på något övertygande sätt. Men delvis har han kanske rätt.

Han påpekar vidare, helt riktigt såvitt jag förstår, att till skillnad från människor saknar AI-program och robotar den sorts förnuft eller intelligens som visar sig i hantverksskicklighet och omdömesförmåga (s. 179f.). Tanken att hjärnan arbetar som en dator är också helt föråldrad, ”eftersom hjärnan består av så mycket mer än bara neuroner. Signalsubstanser som dopamin, adrenalin och oxytocin spelar en stor roll för hjärnans processer och [...] antagandet om att hjärnans aktiviteter kan fångas i ett system av ettor och nollor håller inte” (s. 191).

Vi vet inte hur hjärnan fungerar. Gärdenfors lyfter fram ett lustigt citat: ”Om hjärnan vore så enkel att vi kunde förstå den, skulle vi vara så enkla att vi inte kunde” (s. 194).

Senare i boken skriver han intressant och lärorikt om neurala nätverk, stora språkmodeller och robotar – och hur deras kognitiva prestationer skiljer sig från vad människor och djur kan göra. Hans överblick över dessa områden är imponerande. Han ser inga avgörande argument för att AI skulle kunna uppnå människors generella kognitiva kompetens – men inte heller något avgörande argument för motsatsen (s. 148f.).

Vissa tänkare har som bekant varnat för att AI kan komma att uppnå någon sorts ”superintelligens” och därefter hota mänskligheten och ta över världen. Det tror inte Gärdenfors på. ”AI och robotar kommer inte att ta över därför att de blir för intelligenta. Möjligen kommer de att ta över därför att människor är för dumma” (s. 265). Vi låter ju oss mer och mer påverkas av artificiella system. ”Vi bör inte vara rädda för AI, men vi bör vara rädda för hur AI används” (s. 267).

Lars Bergström

Skepticism

Henrik Lagerlund

Fri Tanke 2022. 419 s. ISBN 978-91-89139-70-1

Henrik Lagerlund har skrivit en filosofihistorisk exposé över skepticis-
men. Något förenklat är skepticis-
men uppfattningen att vi sällan eller
aldrig har någon kunskap. Boken är den första av sitt slag och fyller
därmed en lucka i utgivningen av filosofihistoria på svenska. Den in-
gående diskussionen av skepticis-
men under medeltiden ska dessutom