

Kunstig intelligens og etik

Af Lennart B. Jessen

Efter oplysningstidens og mekaniseringens indtog i Europa kom der et nyt syn på mennesket. Vi finder beskrivelser af mennesket, der mest af alt ligner dem, vi i dag kunne nedfælde for robotters vedkommende (formelle mekaniske systemer). *Descartes* blev kendt for at forstå dyr som "automater", og *La Mettrie* blev kendt for sin bog "*Menneske-maskinen*". Som en naturlig fortsættelse på den naturvidenskabelige forskning, der har været toneangivende efter Descartes, har vi set en tendens til at betragte mennesket som en maskine. Hvis man kunne afkode eller gennemskue funktionerne af hvert enkelt lem eller legemsdel af mennesket, så ville man til sidst sidde med den totale arbejdstegning for, hvorledes et menneske er konstrueret. (Det er det samme syn, der i dag driver projekter som *Human Genome Project*, hvor man samler oplysninger om alle menneskelige gener, for til sidst at have kortlagt alle menneskets arveanlæg). Hjernen var og er ingen undtagelse her. Da computeren var opfundet, mente man, at den var et egnet redskab til simulation af hjernen. Hjernen kunne huske, computeren har en hukommelse. Hjernen kan regne, det kan en computer også. Hjernen kan tage imod kommandoer og udføre et arbejde, det kan computeren også. Man mente altså, at stå med paradigmet på menneskets hjerne. Computeren har alle de faktorer, som hjernen også har, blot skulle det rette program skrives. Man skulle "bare" finde hjernens algoritme, så ville man kunne skabe kunstig intelligens (KI). En idé var, at finde ud af, hvordan viden ligger gemt i vor hjerne. Nu kunne man ikke åbne en hjerne og så finde viden. Neurologer har kendt hjernens materiale i mange år, og stadig ikke fundet ud af, hvorledes vor viden er repræsenteret i den. Man har kikket over skuldrene af lægevidenska-

ben for at se, om der ikke skulle være interessante opdagelser mht. vidensrepræsentation. Den har man endnu kun haft teorier om. Dem har der så til gengæld været en sand flod af. Det mest generelle har været hypoteserne omkring symboler. Hjernen skulle således gemme sine data i symbolsk form. På baggrund af denne teori, blev hele industrien omkring 5. generations programmeringssprog (især LISP og PROLOG) og ekspertsystemer opbygget. I 80'erne var optimismen på sit højeste, og der blev ofret milliarder på verdensplan for at finde den rette måde at simulere en hjerne på. Militæret har været en afgørende økonomisk faktor som støtte til projekterne. Efterhånden opstod der mange problemer med projekterne. Problemerne affødte ikke de ventede resultater, og den økonomiske støtte er blevet minimal. Der var i flere tilfælde problemer med at få afsluttet projekterne, fordi man ikke havde regnet med, at tiden ville stige eksponentielt i forhold til den grad af detaljering, som den viden skulle have, man ville lægge ind i et system. Projekterne blev uoverskuelige og for dyre.

Den svage KI-forskning

En del forskere indså efterhånden det udsigtsløse i den såkaldt stærke KI, dvs. troen på, at den menneskelige intelligens fuldt ud ville kunne genskabes i en computer. Da det endelige resultat ikke så ud til at være inden for rækkevidde i de første årtier, så blev desillusioneringen større, og begrebet KI blev belastet. Der lagde sig en stemning omkring KI i begyndelsen af 90'erne, der minder meget om den, man husker fra middelalderens alkymister, der ville fremstille guld. På den baggrund har nogle KI-forskere afvist den meget ambitiøse (stærke) KI. De nøjes med at antage, at visse dele af den menneskelige intelligens og bevidsthed vil kunne genskabes i en computer (evt. en robot), og at studiet af disse dele af vore mentale funktioner kan give os nyt-

tig viden om os selv. Man ville nu ikke gerne kaldes forsker i KI, men hellere kognitionsforsker. Denne forskning fik navnet *den svage KI-forskning*. Man kan med dette udgangspunkt forsøge at opstille logiske modeller for udvalgte dele af den menneskelige ræsonneren. Man kan også forsøge sig med en simulation af den menneskelige hjernes celler, for derved at opnå en model af hjernen, der af naturlige årsager selv ville fremvise hjernens egenskaber. Man benytter igen computeren som redskab, men denne gang ikke til en overordnet algoritme, men til simulationen af væsentlige træk af hjernens neurale netværk.

Den stærke KI-forskning har lidt under en stærk kritik fra filosoffer som bl.a. Hubert Dreyfus og John Searle², der har opstillet filosofiske tanker, som har været betragtet som kategoriske afvisninger og gendrivelselser af mulighed for KI. Jeg mener dog ikke, de er særlig overbevisende mht. kritik af KI. Deres kritik går meget på, at KI er baseret på den opfattelse, at verden kan beskrives formelt i en nutidig computer, hvilket de sikkert med god grund holder for umuligt. Der er dog et vigtigt modargument til stede, som holder forskningen i gang. Hvis intelligens og alt, hvad dertil hører, findes i vore hjerner, der er en materiel størrelse, og som man derfor kan underkaste naturvidenskabelige undersøgelser, hvorfor skulle man så ikke kunne finde ud af dens funktioner. Hvis man ville hævde, at der er mere i vore hjerner end det, der kan undersøges, så er vi jo tilbage i en middelalderlig situation, hvor man kan hævde alt (bl.a. sjælens eksistens, som et immaterielt objekt). Det kan selvfølgelig være den endelige sandhed, men det er ikke særligt videnskabeligt. Dermed sætter man sig uden for videnskabens rammer og må opgive at finde / søge efter mere faktuel viden om vore mentale evner. Den videnskab, der drives med de faktuelle data, kan stadig søge videre og gøre sig forhåbninger om at finde ny viden om os selv, hvem vi er, hvorfor vi gør, som vi gør, og i sidste ende, hvordan vi bør være. Vi kan så småt ane en fremtidig sammenkædning af forskellige teknologier, som ikke synes indeholdt i KI-kritikernes argumentation. Et eksempel kunne være, når bioteknologien og biokemien begynder at sammenføje sin viden / kunnen med den, vi kender fra fysikken og specielt elektronikken.

Man kan meget vel forestille sig fremtidige "computere", der delvist består af traditionel elektronisk teknologi og ny bioteknologi. Et trin i denne retning har været kendt i nogle år allerede. Selv om Joseph Weizenbaum³ mener, at den slags projekter inden for KI er umoraliske, så foretages de alligevel. Eksemplet her er en mand, der var blind fra fødslen, og har fået indlagt et elektronisk implantat i hjernen, så han via kamerasignaler fra sine briller er kommet til at "se". Man har også hørt om lignende beskrivelser om døve. Teknologien er her en mikrofon i øregangen, der sender signaler via ledninger til en chip⁴, der sidder under huden bag øret. Denne computer sender de bearbejdede signaler videre til hjernebarken nær ved hørecenteret ved hjælp af guldtråde, der tilfældigt er lagt ind over hjernen. Efter en periode med indøvning af lydmonstre, som blev forklaret visuelt for manden, begyndte hans hjerne selv at lave interface⁵ til signaltrådene fra computeren. Manden kom til at høre. Vi kan altså ikke afvise, at en kombination af naturens egne råstoffer og vor teknologi vil bringe KI nærmere virkeligheden. Mange af kritikerne vil da heller ikke kategorisk afvise dens fremkomst. Dem, der tør omtale KI som et absolut umuligt projekt, synes at tilhøre en polemisk fløj, som enten ikke selv fik indfriet de ønsker, de selv havde til KI, da de selv i 80'erne var engageret i KI, eller også gør de det pga. deres menneskesyn, der netop tilhører den kategori, som ikke tør eller vil stille spørgsmålstejn ved vor placering i Guds store skaberværk.

Der har været lavet mange systemer, som antageligt skulle have haft et eller andet niveau af intelligens. Det afgørende spørgsmål er, hvornår er det systemets eller den kunstige hjernes egen intelligens, der skinner igennem, og hvornår er det en afspejling / simulation af konstruktørens indlagte interpretation af virkeligheden? Dette bringer os tilbage til definitionen af intelligens. På dette punkt er forskerne ikke enige. Hvorledes kan denne intelligens tolkes eller defineres? Er det blot evnen til at lagre data i hukommelsen, er det evnen til at kombinere data, er det hjernens struktur med selvbevidstheden, eller er det hjernen i interaktivitet i en interaktiv krop med en interaktiv omverden, der udgør intelligensen? Man kan tilsyneladende ikke komme den generelle definition af intel-

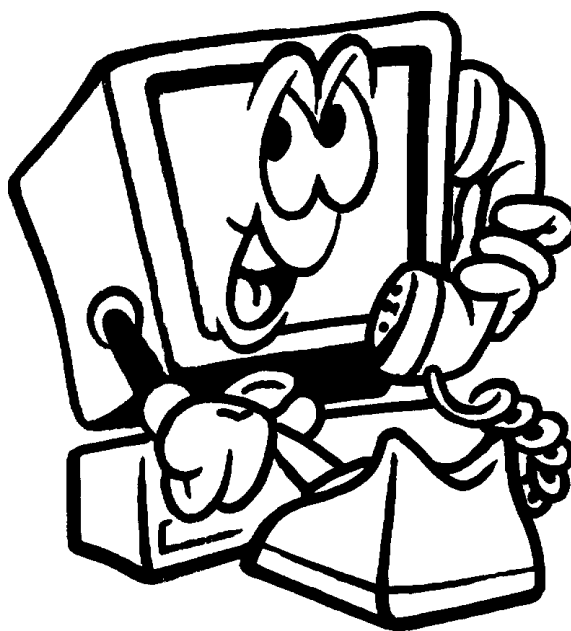
ligens nærmere ad denne vej. Forskerne har derfor koncentreret sig om delaspekter ved den menneskelige hjerneaktivitet. Ved at forske i menneskets kognitive egenskaber prøver man via den indvundne viden og de nye teorier at opstille modeller, der kan bruges i forsøget på at bygge "intelligente systemer".

Computeren som redskab til udvidelse af menneskets potentiale

Mennesket har altid haft et værktøj i hånden. Til tider kan man også se aber bruge genstande som redskaber, men disse er dog forhåndenværende genstande, som så vidt vides ikke bliver forarbejdet og gemt til specifikke formål. Det er også på den baggrund, at nogle har kaldt mennesket for *homo faber*. Synet på værktøj har ikke ændret sig meget igennem tiderne. En tømrer, der tager en hammer i hånden, får udvidet sine naturgivne evner til også at kunne slå et hårdt søm i en hård træbjælke. Værktøj eller redskab er således egenskabsudvidende for det enkelte menneske. Vi kan altså ikke her skille redskabets formål fra den, der har redskabet i hånden. Derfor har der heller ikke været særlige problemer omkring ansvarsplacering og tildeling af ros og ris for det udførte arbejde. Manden bag værktøjer var personen med ansvaret.

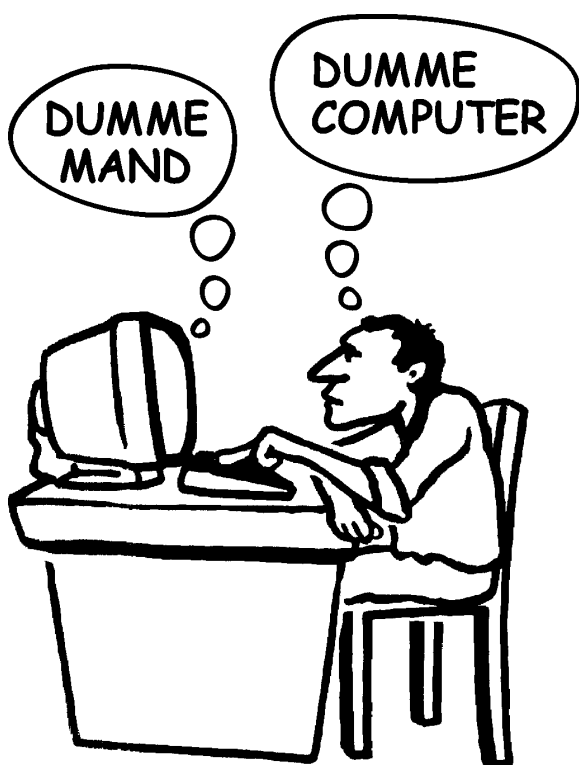
Hvis vi ser tilbage til en tid, hvor slavehandel stadig var lovlig, så finder vi situationer, hvor mennesker blev behandlet som redskaber. I dag vil det vække afsky, hvis vi herhjemme i nutidens Danmark ville behandle andre som redskaber. De flestes etik har heldigvis ændret sig til det bedre på dette område. Den kristne etik af i dag har op igennem historien taget teologiske og filosofiske drejninger, som har ledt os til det, vi mener i dag. En synsvinkel på etikken, som satte sig spor, var bl.a. den, Immanuel Kant beskrev med *det kategoriske imperativ*⁶, som skal indeholde respekt for et andet menneske. Man skal behandle mennesker som formål i sig selv og ikke blot som middel. Menneskerettigheder og andre generelle lovkomplekser giver de samme retningslinier. Hvis man behandler mennesker som redskaber eller midler til eget formål, så kommer disse i konflikt med den nutidige etik.

Nu ville man så sige, at det er et overstået



Figur 1. Kunne man forestille sig, at der kunne fremstilles en samtalerobot, som vi kunne gennemføre samtaler med af samme kvalitet, som vi kender fra vore samtaler med vore medmennesker? I givet fald ville samtalerobotten have bestået Turings test. Ville vi i givet fald måtte tillægge robotten etisk status? Ville vi have pligt til at behandle en sådan robot på en bestemt måde?

kapitel, og især her i Danmark. Vi har ikke problemer med at definere, hvem der er mennesker og kan tilkomme en behandling derefter. Vi er ikke et øjeblik i tvivl om, at vore værktøjer i hobbyrummet, bilen, cyklen, kaffemaskinen osv., at de er redskaber/værktøjer/hjælpe midler i vor hverdag. Selv computeren er vi ikke i tvivl om. Vi hører gentagne gange dem, der siger: "Computeren er og bliver ikke andet end et stykke isenkram". Alligevel kender vi også den situation, at der er ting, vi har personificeret i hverdagen. Der er mange, der f.eks. giver deres biler navne. Ja – man kan endda finde på at tale til den eller blive vred på den, hvis den ikke gør, som man forventer. Selv om bilens mekaniske indre ikke kan gøre for dens fejl og mangler, så går manges aggression ud over bilen, og ikke over dens konstruktør eller mekaniker. Der bliver sparket til den, smækket med dørene og banket på rattet. Vi er alle klar over, at bilen ikke forstår det, og vi ved alle, at den ikke er



Figur 2. Hvem er vi som mennesker i forhold til computeren? Har stærk kunstig intelligens i virkeligheden ret i, at et menneskes bevidsthed i princippet ikke er andet end et avanceret computerprogram?

levende osv. Men alligevel vælger vi i visse situationer at opføre os, som om den var. Ud fra dette kan vi tænke, hvor let det må være at personificere en computer, der både kan give talt sprog tilbage og opføre sig som levende. Netop dette observerede Weizenbaum, da han lod analyserne prøve Eliza, den computer han havde programmeret til at simulere psykoterapeut.⁷ Weizenbaum fortæller selv i en af sine bøger, at hans sekretær blev vred på ham, da han ville spørge computeren om de samtaler, den havde haft med sekretæren. Hun reagerede if. Weizenbaum, fordi hun havde haft fortrolige samtaler med computeren og var bange for, at den skulle afsløre hende. Systemer, der kan simulere den slags samtaler, findes i dag flere steder på Internettet, hvor man kan prøve at afsløre deres begrænsninger. Man finder disse kunstige samtalepartnere ved at søge efter kunstig intelligens eller Artificial Intelligens på mange af de kendte søgemaskiner (Alta Vista, Jahoo, osv.).

Selv om vi her stadig har en klar fornemmelse af, hvad der er redskab/computer, og hvad der er menneske, og det i kraft af, at vi har været det foregående stof igennem og evt. også selv sidder inde med faglig viden om computere, så kan vi iagttage et stigende problem for de mennesker, som ikke har en faglig baggrund for at forstå, hvor grænsen mellem såkaldte "dumme" værktøjer og intelligens ligger. Om end grænsen ikke bliver usynlig, så bliver den i hvert tilfælde meget utydelig, når man ser medierne bruge udtryk som intelligens på en meget lemfældig måde. Hvis man tager begrebet intelligens med i PR-materialer, så almindelige forbrugere får det indtryk, at computere bliver intelligente ved køb af specielle programpakker, så er det yderst betænkeligt. Det er måske direkte vildledende. Ved at vende sig til brugen af begrebet intelligens i denne forbindelse, så afsvækkes begrebets betydningsindhold, som det har haft i relation til mennesker. Når man taler om intelligens, bliver det med tiden en meget relativ størrelse, og vi får derfor et problem med den almindelige brug af det.

Med dette som baggrund mener jeg, at frygten for en igangværende lavineeffekt er begrundet. Frygten for at mennesket skaber sin egen totale afhængighed af teknologi. Er vi på vej til at forslave os selv? Tanken er ikke ny. Heidegger har filosoferet grundigt over dette emne. Efter hans mening har vi allerede forladt den totale naturlighed og er årsag til, at verden med alt, hvad der er i den, er indfældet i teknik. Selv den mest naturlige plet i Amazones jungle eller i Sibiriens ødemarker er indfældet og omgivet af vor verdensomspændende styring, kategorisering og værdisætning. Ifølge Heideggers egne ord finder vi den yderste fare netop i teknikken⁸. Selvom Heidegger ikke ser teknik som andet end det, mennesker frembringer som værktøj for menneskets viljeshandling, og derfor ikke kan appliceres over på en evt. kunstig intelligens, der også må påregnes at have vilje, så ser han dog den enorme fare, der ligger i dens denaturering af verden. Løsningen ses ikke i en kamp imod teknikken, men derimod i en tolkning af den (her igennem kunst).

Hvis værktøjet ikke mere kun er noget der sætter mennesker i stand til at gøre noget, - hvis

værktøjet bliver selvstændigt, gør sig fri af menneskets styring, - hvis vi skal til at dele det øverste trin i skabelshierarkiet med noget andet, - hvis mennesket må vige førerpladsen på podiet til fordel for værktøjet, der bliver selvstændigt, - hvad så? Selv om vi ikke har den kunstige hjerne endnu, så er der en hel del i vort samfund og i videnskabelige kredse, der opfører sig, som om det ligger lige om hjørnet. I denne sammenhæng er det faktisk ikke afgørende, om det nogensinde bliver virkeliggjort. Det er min opfattelse, at alene det faktum, at en del styrer imod dette mål, idet de antager, at det i princippet ville kunne nås (og at stærk KI dermed har ret), sætter en dagsorden for en nutidig og dermed også fremtidig etik.

Livskvalitet, kultur og teknologi

I verdenshistorien har vi set kulturer komme og gå. Vi kan bare tænke på Inka- og Mayakulturerne, den Ægyptiske og Babylonske kultur, og mange andre. Af hvilke grunde disse kulturer end er gået til grunde, blevet opslugt af andre eller blot blevet forladt, er ikke interessant her. I alle disse kulturer kan vi se nogle generelle fælles værdier, som til alle tider har været interessant for mennesker. En eller anden form for prioritering har altid spillet ind. Hvis vi i første omgang ser bort fra mere specielle religiøse situationer, så står føde og sikkerhed naturligt nok øverst på prioriteringslisten. Når de basale behov er tilgodeset, så begynder sikringen af fremtiden at spille en rolle. Hvis den også synes at være sikker, så står mennesket i en situation, hvor det behagelige bliver værdsat. Uden nogen særlig form for etik går det over i udnyttelse af omgivelserne. Magt bliver udøvet, og luksus bliver måske anskaffet på bekostning af andres velfærd. Vi oplever i dag en levende debat om værdier i det danske samfund, og man hører om velfærd og livskvalitet. Bortset fra en ustyrlig nysgerrighed hos mennesket, så kunne man frygte, at det, der driver forskningen ind i robotverdenen og den kunstige intelligens, var ønsket om et væsen, der kan sikre mennesket og dets behagelige eksistens ind i fremtiden. Nogle ville måske foretrække en god gammel dags slave, men det er jo ikke lovligt mere! Vil en menneskeskabt "androide" (dvs. en robot i mere eller mindre human form) da være løsningen? Hvis vi forestiller os, at vi kan skabe et

væsen som vor egen kopi, som vi vil have magt over, så afføder det omgående to alvorlige spørgsmål:

a) Hvis vi forestiller os, at vi faktisk laver en androide, som en kopi af os selv, har vi så ikke brugt de forhåndenværende materialer fra naturen. En computer består også af det, vi finder i naturen, selv om vi kalder det teknik. Når vi benytter os af gener i grise og avler nye typer grise, så kan vi stadig hævde, at vi blot bruger de midler, Gud giver os dertil. Gør vi ikke det samme, når vi laver en computer? Silicium-chippen, hvorpå computeren laves, er også et stykke af det univers, Gud har skabt. Har Gud i givet fald brugt os som fødselshjælper til en ny art? Hvis der faktisk er tale om en androide, som i enhver henseende er en kopi af os selv, har vi så ikke de samme etiske forpligtelser over for denne androide, som over for os selv? Bliver den etik, vi har hos os selv, ikke den samme over for androiden? Og hvad bliver vore etiske forpligtelser, hvis der er tale om en androide, som i nogle henseender er en kopi af mennesket? Hvor meget skal en robot ligne os for med rimelighed at kunne gøre krav på etik i forhold til mennesket.

b) Et andet spørgsmål kunne være: Vil vi med teknologiens hjælp sætte os på Guds trone? Vil vi skabe intelligente væsner for at blive ligeværdige med Gud? Vil en kopiering af vor egen hjerne være den ultimative proteste for at være helt uafhængig? Vil vi unddrage os selve døden og overleve i en virtuel verden, for til tider at holde ferie i den rigtige verden i en lejet højteknologisk krop? Hvor mange dele må vi skifte ud i vor egen krop, før vi ikke er mennesker mere? Er dette ikke den højeste form for Gudsforagt og Gudsbe-spottelse, der styres imod?

Enhver, der vil hævde, at mennesket i princippet ikke er andet end en maskine, stilles over for disse vanskelige spørgsmål. Men det skal bemærkes, at forudsætningen for spørgsmålene er, at stærk KI har ret i sin grundlæggende antagelse om, at den menneskelige intelligens, bevidsthed og funktionalitet i princippet fuldt ud ville kunne genskabes i en computer (robot) evt. af et helt andet design end det, vi kender i

dag. Hvis den antagelse ikke holder, falder spørgsmålene til jorden. Men så er udfordringen jo at vise, hvorfor stærk KI tager fejl, og hvordan det menneskelige principielt adskiller sig fra det, som i princippet ville kunne produceres med en eller anden form for informationsteknologi.

Noter

¹ *Maskinen, människan och doktor la Mettrie*, Doxa, Lund, 1984.

² Peter Kemp, *Det uerstattelige*, Spektrums Uglebøger, 1994, side 252.

³ Tom Forester og Perry Morison refererer til Joseph Weizenbaum (professor i computervidenskab, MIT Massachusetts Institute of Tech-

nology) i *Computer Ethics*, Oxford 1990, side 129.

⁴ En "chip", er normalt et stykke (ca. 4 mm²) silicium, hvorpå der er konstrueret et elektronisk kredsløb, eller en mikrocomputer.

⁵ Interface betyder i dette tilfælde noget mere end en blot og bar tilslutning. Hjernen udvikler de neuroner og synapser, som føler impulserne fra mikrofonen, og udvikler et system, der i sidste ende producerer et forståeligt lyd billede for personen.

⁶ Immanuel Kant, *Metaphysik der Sitten*, 1797.

⁷ Peter Kemp, *Det uerstattelige*, side 263.

⁸ Søren Gosvig Olesen, *Vejledning til Heidegger*, Odense Universitetsforlag, 1990, p. 79.

Lennart B. Jessen er cand.theol. og sognepræst.

Bog anmeldelse

Forlaget Perspektiv har netop udgivet Erik Fanøes spændende kommentar til Bibelens første bog, Første Mosebog. Fanøes kommentar er et glimrende værktøj til enhver, der vil fordybe sig i Bibel-teksten. Kommentaren er velskrevet og engageret. Stilen er åben og ydmyg over for de vanskelige fortolknings spørgsmål, som findes i forbindelse med Første Mosebog. Erik Fanøe har skrevet kommentaren ud fra sin solide faglige og medmenneskelige ballast. Det er oplagt, at Fanøe har en stor viden at øse af - ikke bare på de teologiske og sproglige områder, men også i forhold til moderne naturvidenskab.

Bogen er opdelt i lektioner, og til hver lektion er der spørgsmål og opgaver. Bogen vil egne sig glimrende som grundbog for studiekredse mv.

Første Mosebog.
Kommenteret af Erik Fanøe.
Perspektiv Bibelstudier.
Forlaget Perspektiv 2001

Peter Øhrstrøm

