

■ Denne artikel er inspireret af kapitel 9, Et spørgsmål om information, i John C. Lennox' bog **Guds bødde** (Origo 2012). Og konklusionen er: Information i en eller anden forstand er nødvendig for at forklare tilblivelsen af DNA og proteiner.

Biologi og informationsteori

Af Andreas Vedel & K. Aa. Back

Det grundlæggende problem i kommunikation er i et punkt enten eksakt eller tilnærmelsesvis at gengive et budskab udvalgt fra et andet punkt.

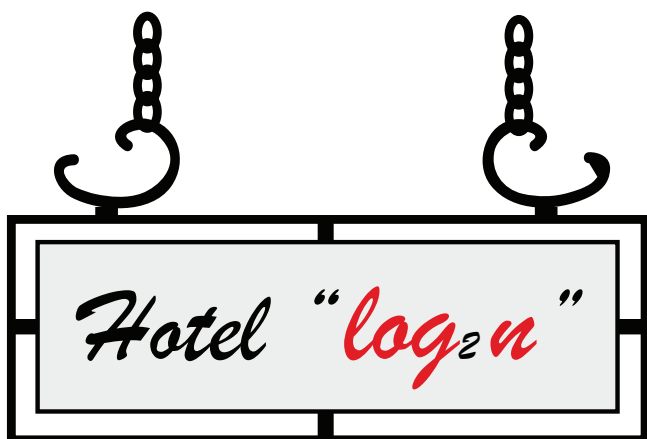
(Claude E. Shannon, 1948)

Dette citat er taget fra Claude E. Shannons klassiske afhandling *A Mathematical Theory of Communication* fra 1948. Denne skelsættende teori var for alvor startskuddet til etablering af den faglige disciplin vi kalder informationsteori, og som øjeblikkelig skabte verdensomspændende opmærksomhed. Shannon introducerede *en bit* som den mest grundlæggende informationsenhed, og han gjorde det muligt at sætte et eksakt mål på information.

Som titlen på afhandlingen antyder, er der tale om en matematisk formuleret teori. Det vil dog føre for vidt at redegøre for den her. Lad os i stedet kaste os over et par hverdagseksempler fra Lennox' bog der på bedste vis illustrerer værdien af Shannons teori.

På hotel for at måle information

Lennox tegner følgende billede af hvilke forestillinger vi har om begrebet information:



Omfanget eller mængden af information der er nødvendig for at angive et bestemt værelse på et hotel med n værelser kan beregnes vha denne matematiske formel, (III. Andreas Vedel).

»Vi ankommer fx til et lille hotel hvor vi har reserveret et værelse, og opdager at hotellet kun har otte værelser. Hvis vi nu antager at alle værelser er ens, og at vi ikke har bedt om et bestemt værelse, så er sandsynligheden for at få et bestemt værelse 1 til 8 ($1/8$). Denne sandsynlighedsberegning er et klart udtryk for vores usikkerhed. Modtager vi nu information

sagen kort



I almindelig sprogbrug anvender vi ordet "information" til at beskrive noget som vi véd nu, men som vi ikke vidste før. Vi siger at vi har modtaget information. – Den faglige disciplin *informationsteori* har noget væsentligt at sige for det vi omtaler som genetisk information. I denne artikel analyserer vi informationsbegrebet, der dels er et grundlæggende begreb og dels udfordrer den traditionelle neodarwinistiske opfattelse.

om at vi fx har fået tildelt værelse nr. 3, forsvinder denne usikkerhed.«

Men hvor meget information har vi egentlig modtaget? Det er her Shannon kommer os til hjælp.

»En af måderne hvorpå vi kan måle den informationsmængde vi har modtaget, er at lave en beregning af det mindste antal ja- eller nej-spørgsmål vi skal stille for at finde ud af hvilket værelse vi har fået tildelt. En lille eftertanke bør overbevise os om at dette tal er tre: Første spørgsmål bruger vi til at reducere antallet fra 8 til 4 alternativer ("ligger værelsesnummeret mellem 1 og 4?"), det andet spørgsmål fra 4 til 2 ("har værelset nr. 1 eller 2?"), og sidste spørgsmål afgør sagen ("er værelset nr. 3?"). Vi siger at vi har modtaget tre *bits* information, eller at vi har brug for tre bits information for at finde ud af hvilket værelse der er vores. Vi bemærker at 3 er den potens vi skal opløfte 2 i for at få 8 (dvs. $8 = 2^3$), eller, for nu at vende det om, 3 er logaritmen af 8 med grundtallet 2 (altså $3 = \log_2 8$). Ud fra det kan vi let udlede at hvis der er n værelser på hotellet, så er mængden af information der er nødvendig for at angive et bestemt værelse $\log_2 n$.«

Alfabet og information

Lennox bruger endnu et eksempel:

Vi bliver bedt om at tænke på en sms vi gerne vil sende. Husk at sms'er skrives i sætninger der består af ord og mellemrum. Vores "alfabet" består derfor af 29 bogstaver plus et mellemrum, altså 30 tegn. Den information der gives i en tekst på m tegn, bliver således m gange $\log_2 30$ ($m \cdot \log_2 30$).

Nu kan vores sms jo også indeholde tal, og så er vores "alfabet" udvidet til 40 tegn, altså en information for hvert tegn på $\log_2 40$.

Cifret "2" spiller åbenbart en særlig rolle her. Og det kan vi "sammenkøre" med det faktum at computerverdens symbolske "alfabet", det binære talsystem, består af de to tegn "0" og "1". Derfor kan man nøjes med blot 2 tegn hvis man vil indkode et hvilket som helst alfabet ved hjælp af det binære talsystem.

Det danske sprog bruger 29 bogstaver plus et mellemrum. Her vil, som *Guds bødde* anfører, »binære ord på bare 5 tegns længde (nemlig $2^5 = 32$ der jo er større end 30) være tilstrækkelige til at kode hele molevitten, oven i købet med et par pladser til overs: Vi kunne fx vælge at indkode symbolet for "mellemrum" som 00000, og sætte A = 00001, B = 00010, C = 00011, etc.«

Syntaktisk information

Lennox beder os antage at vi får følgende "besked" på vores mobiltelefon: ZXXTRQNJPWTRP. Men som han også konstaterer, ser vi ingen information "i den form for volapyk." Men vi skal lægge mærke til at "information" faktisk ikke har noget med en evt. meningsfuld betydning at gøre. Der kan være tale om *syntaktisk information*, og den er yderst brugbar for en kommunikationstekniker som alene er interesseret i kommunikationskanalens kapacitet, dens pålidelighed osv.

Vi tager endnu et eksempel fra Lennox for at sikre os at vi forstår hvad vi taler om:

På et bibliotek beder du bibliotekaren hjælpe dig med at finde en bog om *nefrologi*. Bibliotekaren får nu af dig de $9 \cdot \log_2 30$ informationsbits som dette ord indeholder (ifølge Shannon). Nu bliver ordet "nefrologi" behandlet på rent syntaktisk niveau, dvs. som en meningsløs række bogstaver. Men det er nok til at man kan finde frem til den bog om emnet biblioteket (måske) har stående. For en læge derimod, der véd at ordet "nefrologi" har noget med nyresygdomme at gøre, indeholder ordet ikke kun syntaktisk information, men også *semantisk*. Og hvad er så det?

Semantisk information

Lennox skriver: »Måling af semantisk information udgør et langt mere vanskeligt problem at få styr på rent matematisk, og det er endnu ikke lykkedes at finde en succesfuld måde at greje det på. At dette næppe kan komme som den store overraskelse, har at gøre med det velkendte faktum at betydningen af en tekst er meget afhængig af den pågældende teksts sammenhæng. Hvis du ser mig modtage et JA på min mobiltelefon, gætter du rimeligvis på at det er svaret på et spørgsmål jeg har stillet, men du har ikke en chance for at vide om dette spørgsmål f.eks. er: "Har du købt billet til fodboldkampen i aften?" Eller: "Vil du gifte dig med mig?" Meddelelsens betydning kan bare ikke bestemmes uden forudgående kendskab til den kontekst, den sammenhæng, den forekommer i. Med andre ord: Der skal en hel del mere information til før man kan tyde en given oplysning.«

DNA og information

Lennox overfører nu denne tankegang til molekylærbiologien. Han beder os tænke på den række "bogstaver" vi finder i DNA-molekylets kemiske alfabet (baserne A, C, G og T). For en molekylærbiolog kan strengen have en semantisk dimension ved at han/hun "forstår sproget", véd hvad den koder for osv.

For dem der ikke véd noget om gener og proteinsyntese, ser strengen ud som en lang liste af meningsløse symboler ACGGTCAGGTTCTA... Men det giver stadig god mening at tale om denne symbolstrengs informationsindhold i syntaktisk eller i Shannons forstand. Man behøver ikke forstå hvad

■ Darwin i vores videnshuller 0

Livets tilblivelse

Er skabelsestanken i bund og grund uvidenskabelig? Mange har i hvert fald den tro at Intelligent Design er en værre gang pseudovidenskab; at ID-forskerne bruger Gud-i-hullerne-argumenter: Vi kan ikke forklare et bestemt biologisk fænomen, fx øjets opståen i evolutionens løb, derfor "må Gud have skabt det".

Nu er det faktisk ikke sådan der argumenteres af ID-folk eller de forskere der tænker Skabelsen ind som en grundbetingelse (som *causa prima*) for deres videnskab. For modsætningen til skabelsestanken er ikke naturvidenskab, men en tro på at alting kan blive til af sig selv. Og der er ikke spor "pseudo" over ID-forskningen, blot en tør konstatering af fakta man ikke kan lade være med at snuble over med den nyeste forskning i bioinformatik. ID-forskerne stiller nemlig sig selv dette spørgsmål: »Hvor ligger den mest fornuftige forklaring? At livet er blevet til af sig selv i en "ursuppe" med tilfældige kemiske reaktioner? Eller at livets kode kræver en form for intelligent input, en form for information!«

Lennox siger i sin Guds bødde at den jødisk-kristne skabelsestanke finder mere og mere evidens i naturvidenskaben, lige fra universets fødsel til livets opståen; fra "der blive lys" til "i begyndelsen var Ordet/Logos". Altså at den bibelske beretning, ret forstået, faktisk i dag kommer med den bedste forklaring på altings tilblivelse. – Jo mere vi véd om naturens grundbetinger, desto klarere står skabelsestanken.

Darwin i vores videnshuller 1

For at komme uden om det faktum at de traditionelle forklaringer på livets opståen og udvikling mere og mere hænger i en tynd tråd, søger materialister tit & ofte at komme med forklaringer som bedst kan betegnes som "ikke-forklaringer", "sådan-er-det-bare-forklaringer" eller lige frem "Darwin-i-hullerne-forklaringer". Og det sidste kan defineres sådan: Selvom vi faktisk ikke har en forklaring på et bestemt biologisk fænomen, holder man sig til at Darwins evolution må have skabt det. Øjet fx. Nu er det selvfølgelig aldrig formuleret på denne (klare) måde; men derfor møder man alligevel masser af eksempler.

De afsløres tit ved hjælp af et simpelt spørgsmål: "Og hvordan er det så lige dét er foregået?"

Men pas på: Dette simple spørgsmål indhylles ofte i en uigennemsigtig forklaringsståge med et svar der bruger argumentet om at "det har været en fordel for organismen at ..." o.lign.

Bestil abonnement på Origo på abonnement@skabelse.dk

Eller giv et abbonemt som gave. Den perfekte gave til manden/kvinden der har alt.

strengen betyder. Man kan sagtens regne ud hvor meget syntaktisk information der er brug for for at kunne gengive strengen nøjagtigt: Det genetiske alfabet består af fire bogstaver, så hvert bogstav omfatter 2 informationsbits ($\log_2 4 = 2$). Så DNA'et i det menneskelige genom, der er omtrent 3,5 milliarder bogstaver langt, indeholder omkring 7 milliarder informationsbits.

Abekattestregere og forfattertekster

Der er en vigtig betragtning vedrørende begrebet kompleksitet vi lige skal forbi inden vi går videre:

Tag en (binær) streng på 6 mia. cifre, siger Lennox, og den skal være udformet sådan: 001001001001... osv. Er en sådan streng lige så kompleks som det menneskelige genom (som er "lidt længere", nemlig på førnævnte 7 mia. bits)? Overhovedet ikke! Vores 001-streng gentager blot et fastlagt mønster ("001"). Og selvom det er mange gange, er det alligevel forholdsvis simpelt. Et enkelt computerprogram kan skrive tallet ud, nemlig: "For $n = 1$ til $2 \cdot 10^9$ skriv 001. Stop."

Lennox beder os nu overveje følgende:

Hvis vi sætter en abe til at lege med et computer-tastatur, vil den måske få følgende streng ud af det: Mtl3(#8HJD[;ELSN29XITNSP]\@... Vi antager nu at denne streng også er 6 mia. bogstaver lang. Men bemærk i forhold til vores 001-streng: Den er algoritmisk ikke-komprimerbar, dvs. man kan ikke som før lave et simpelt program der skriver strengen ud nøjagtigt som aben (tilfældigvis) har gjort. Vi kan altså, siger Lennox, slå fast at denne streng er maksimalt kompleks ud fra vores definition på kompleksitet.

Nu tager vi så en streng nr. 3. Den er lavet af de første 6 mia. bogstaver som vi har kunnet finde i en hyldefuld biblioteks-bøger. Denne streng er lige så algoritmisk ikke-komprimerbar som streng nr. 2. Og i matematisk forstand er den også "tilfældig". Men der er en afgørende forskel: Strengen der er genereret af en abe, giver ingen mening. Vi forstår ikke hvad den har "skrevet".

Streng nr. 3 derimod indeholder semantisk information – vi forstår hvad ordene fra bøgerne betyder. Og dermed står vi med et eksempel på det der kaldes *specificeret kompleksitet*.

DNA – ikke en tilfældig rækkefølge af bogstaver

Lennox: »Alle DNA-molekylets A'er, C'er, G'er og T'er kan indtage en hvilken som helst position. De repræsenterer en "sprogbrug" der ud fra et matematisk synspunkt, er tilfældige (stokastiske). Men det må selvfølgelig ikke få os til at tro at DNA-sekvenser er helt vilkårlige. Langt fra! Det er faktisk sådan at kun en meget lille del af alle DNA-molekylets mulige sekvenser vil besidde den nødvendige kompleksitet der ligger i molekyler med en biologisk betydning. Det er i høj grad på samme måde som at det kun er en meget lille del af alle mulige bogstavsekvenser i et alfabet, eller endog ord i et hvilket som helst menneskeligt sprog, der vil besidde den specificerede kompleksitet der skal til for at give udsagn mening på det pågældende sprog.«

Lennox fortsætter med en beskrivelse af hvor ekstremt specifikke proteiner og DNA-koderne er.

De mindste proteiner vi kender til og som besidder en biologisk funktion, "bruger" ca. 100 aminosyrer. Så de DNA-molekyler der svarer til dem, har ca. 10^{130} mulige sekvenser, og kun en meget lille del af dem har biologisk betydning. Mængden af alle mulige sekvenser er derfor ufattelig stor. Da DNA'et ikke har en indbygget "rækkefølge" (lige som en bog ikke har en naturlov for hvordan ordene skal sættes sammen), er alle basesekvenser lige sandsynlige. Dvs. at sandsynligheden for at en bestemt sekvens skulle få biologisk betydning ved at den opstår tilfældigt, er så lille at den må afskrives.

Hertil kommer at proteiner er meget følsomme over for fejl i koden (mutationer). Så selv en udskiftning af en enkelt aminosyre i et brugbart protein kan få katastrofale følger. Se bare på hvilken ravage genetiske sygdomme kan lave. Med Lennox' ord: »Man kan derfor med god ret hævde at cellens molekylærbiologi (DNA, proteiner, enzymer etc.) er ekstremt finjusteret eller specificeret.«



I alle celler er der semantisk information skrevet vha. de 4 "bogstaver" i DNA'et: A, C, T og G. Cellers kromosomer kan opfattes som bøger med opskrifter eller koder der tilsammen indeholder information om den pågældende organismes opbygning og funktion. (Foto: Andreas Vedel, fra Udstillingen "Who am I?", Science Museum, London, maj 2012).

Og derfor har cellen (naturligvis, fristes man til at sige) den fantastiske korrekturfunktion som er beskrevet i andre sammenhænge.

Tilfældighed og ekstrem specificering

Lennox citerer Paul Davies for følgende: »Biologisk relevante makromolekyler som DNA og proteiner besidder altså på samme tid to afgørende egenskaber: tilfældighed og ekstrem specificering. En kaotisk proces kan muligvis nå frem til den førstnævnte egenskab, men vil have betydelige problemer med at nå frem til den sidstnævnte.«

Davies fortsætter: »Ved første øjekast synes dette at gøre genomet til et umuligt fænomen, uopnåeligt både ved hjælp af de kendte naturlove og de rene tilfældigheder.«

Lennox slutter med følgende citat af darwinisten(!) Davies: »Problemet er, for så vidt som det har med biogenese at gøre, at darwinismen kun kan fungere når livet (under en eller anden form) allerede er i funktion. *Evolutionsteorien kan ikke forklare hvordan livet er begyndt fra først af.*« (Lennox' fremhævelse.)

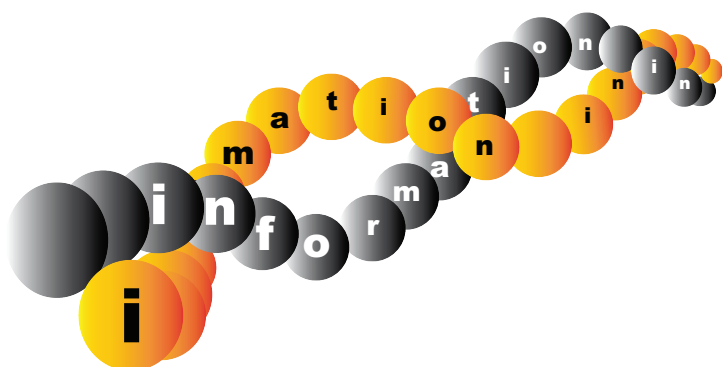
Altså kan vi foreløbig konkludere at intet tyder på at naturens processer har en tilbøjelighed til at skabe så komplekse og informationsbærende molekyler som DNA og proteiner. Hertil kommer at tilfældighed også er afskrevet som en gyldig forklaring.

biogenese = af *bio* (liv) + *genese* (skabelse) – altså livets tilblivelse

Tilførsel af information er eneste reelle alternativ

Det får så Lennox til at komme med følgende konstatering:

»Men hvilke andre muligheder findes der ud over tilfældighed og tilbøjelighed? [...] Hvis tilfældighed og tilbøjelighed, enten hver for sig eller tilsammen, ikke er i stand til at skabe livets kompleksitet, så må vi overveje muligheden for om der ikke har været en tredje faktor involveret. Og denne tredje mulighed er input eller tilførsel af information.«



Alle spiralsnoede DNA-molekyler indeholder store mængder information skrevet vha. livets alfabet på 4 bogstaver. Se også artiklerne om *cells proteinsyntese* på side 12-17. (III. Andreas Vedel).

Men som Lennox også gør opmærksom på, så bliver dette forslag »sædvanligvis mødt af et kor af protester om at det er i hvert fald et udtryk for anti-videnskabelig indstilling og intellektuel dovenskab at foreslå noget der reelt er en "Intelligens-i-hullerne-" eller en "Gud-i-hullerne-løsning". Nu skal denne anklage naturligvis tages alvorligt: Det er trods alt muligt at være intellektuel doven og så bare sige: "Jeg kan ikke forklare det, derfor må Gud have skabt det." – Men samtidigt er det vigtigt at understrege at dét med den intellektuelle dovenskab kan gælde begge veje: Det er også meget let at sige "evolution har skabt det", når man heller ikke har den fjerneste

■ Darwin i vores videnshuller 2

Altså hvordan blev livet til?

I Origo 122, side 5, citerer vi fra en dansk biologibog for folkeskolen. Den påstår at »også DNA som danner det "celleprogram" der styrer arbejdet i en celle, kunne dannes ud fra stofferne i "ursuppen".« – Vores hvordan-spørgsmål fra før afslører at dét har man faktisk ikke nogen forklaring på. "Men eftersom livet jo er opstået her på planeten, så må det jo være opstået af ursuppen!"

Vi konstaterer hermed at man her bruger en Darwin-i-hullerne-forklaring som ikke engang Darwin selv ville kendes ved. For Darwins evolutionsteori siger intet om hvordan livet er opstået. Den indeholder alene påstand om hvordan det har udviklet sig fra første begyndelse. Eller rettere *efter* første begyndelse.

Vi tager nogle flere Darwin-i-hullerne-forklaringer ved vingebeinet i dette blad. Find selv de små bokse rundt omkring i bladet.

idé om hvordan det er gået til. [...] Det er jo faktisk sådan [...] at en materialist *er nødt til* at sige at naturlige processer er eneansvarlige, eftersom der ifølge hans eller hendes overbevisning ikke findes noget acceptabelt alternativ. Som følge heraf ramler man lige så let ind i en "Evolution-i-hullerne-" som en "Gud-i-hullerne-forklaring".

Lennox bemærker så at det er langt farligere at bruge "evolution" som altingsforklaring frem for at bruge "Gud". For hvis man ikke bruger "Gud", vil der være langt færre protester, selv dér hvor argumentet har samme svage vægt.

Som det forhåbentligt fremgår af ovenstående argumenter, er information i en eller anden forstand nødvendig for at forklare tilblivelsen af DNA og proteiner. Hvis dette står til troende, så er det logisk at enhver forklaring der hævder at få information "gratis" vha. helt naturlige processer, på en eller anden måde får smuglet information ind udefra! Også selvom den påstår det modsatte. Og hvis logikken slår dette fast, har vi i det mindste et plausibelt argument for at et informationsinput er nødvendigt for livets oprindelse.

Den deraf følgende (ubehagelige) konsekvens

Hvis der er videnskabeligt belæg for at livet, som vi kender det, ikke i tilstrækkelig grad kan forklares uden inddragelse af et informationsinput, vil fokus uundgåeligt skifte over til at finde ud af hvad kilden til denne information så kunne være. Men det skal bemærkes at sidstnævnte er et helt andet emne – hvor svært det end kan være at holde de to ting ude fra hinanden i vores tænkning. Hvorvidt vi videnskabeligt kan fastslå informationskildens identitet, er logisk set irrelevant for spørgsmålet om hvorvidt et eksternt informationsinput er nødvendigt. ■