Normat **53**:2, 72–81 (2005)

Følger det fra Gödels ufullstendighetsteoremer at vi ikke er maskiner?

Ståle Gundersen

Filosofisk institutt
Universitetet i Bergen
Sydnesplassen 7
NO–5007 Bergen
stale.gundersen@fil.uib.no

Innledning

I denne artikkelen skal jeg gi en kortfattet presentasjon av Kurt Gödels to ufullstendighetsteoremer. Jeg skal ikke gi noe bevis for ufullstendighetsteoremene ettersom dette ville sprenge rammene for artikkelen.[1] Senere skal jeg vise hvordan disse to teoremene kan benyttes som premisser i et argument som konkluderer med at det ikke er mulig å programmere en robot eller computer slik at den kan oppvise de samme matematiske ferdigheter som mennesker. (Betegnelsene «maskin», «computer» og «datamaskin» vil i denne artikkelen bety det samme.) La oss kalle hypotesen som sier at det er mulig å programmere en robot til å oppvise de samme matematiske ferdigheter som mennesker for *computerhypotesen*. Computerhypotesen hevder at hjernen manipulerer symboler på grunnlag av visse regler; omtrent på samme måte som du flytter på («manipulerer») brikkene («symbolene») på sjakkbrettet. Ifølge computerhypotesen er våre matematiske ferdigheter et resultat av symbolmanipuleringer i hjernen i henhold til visse regler, og alt dette kan beskrives fullstendig ved hjelp av et dataprogram. Tilhengerne av computerhypotesen hevder vanligvis at en computer eller robot ikke bare kan oppvise de samme

[1] Originalartikkelen er Gödel (1931). Det finnes en enkel, men glimrende fremstilling av Gödels bevis i Nagel & Newman (1989). Smullyan (1992, kap. 1–3) er en god teknisk fremstilling.

matematiske ferdigheter som mennesker, men at den kan simulere (etterligne) all menneskelig adferd. En riktig programmert robot kan i prinsippet, for eksempel en gang i fremtiden, tenke, føle og oppføre seg som et menneske. Jeg skal i denne artikkelen begrense meg til å diskutere matematiske ferdigheter.

Oxford-filosofen John Randolph Lucas, og i senere tid også fysikeren og matematikeren Roger Penrose, har argumentert mot computerhypotesen.[2] De hevder at det følger fra Gödels ufullstendighetsteoremer at computerhypotesen må være feil. Lucas' og Penroses argument for denne påstanden er i kortformat at mennesker har visse matematiske ferdigheter som maskiner, i kraft av å være programmert på en bestemt måte, ikke kan ha. Svært få aksepterer dette argumentet til Lucas og Penrose, og jeg skal presentere noen innvendinger som har blitt rettet mot det. Diskusjonen omkring Lucas/Penrose-argumentet er til tider svært teknisk, men jeg skal kun beskrive de mest grunnleggende trekkene i dette argumentet.

Maskiner og aksiomsystemer

Vi kan aksiomatisere en matematisk disiplin (for eksempel aritmetikk) ved å konstruere en mengde med aksiomer og regler slik at vi kan utlede matematiske sannheter (teoremer) innenfor disiplinen ved å anvende reglene på aksiomene. (Euklids geometri er et godt eksempel på et aksiomsystem.) Et aksiomsystem er *fullstendig* hvis alle matematiske sannheter innenfor den matematiske disiplinen kan utledes innenfor aksiomsystemet. Moderne aksiomsystemer er rent *syntaktiske* eller *formelle* systemer. Dette innebærer at det ikke er nødvendig å tillegge symbolene noen bestemt mening for å kunne utlede teoremer. En kalkulator forstår ikke betydningen av symbolene på displayet, men den klarer likevel å gi oss svaret på ulike regnestykker i kraft av å følge visse regler. Her er et enkelt eksempel på et aksiomsystem.[3]

De eneste *symbolene* er «`p`», «`q`» og «`-`». Disse kan settes sammen til *setninger* eller *symbolstrenger*. Det er ett *aksiomskjema*: x`p-q`x`-`, hvor x er en sekvens av «`-`». (Dette kalles et aksiomskjema fordi det gir opphav til potensielt uendelig mange ulike symbolstrenger med status som aksiomer.) Den eneste *regelen* er: La x, y og z være sekvenser av «`-`». Hvis x`p`y`q`z er et teorem, så er xpy-qz- et teorem.

Her er et eksempel på en utledning innenfor aksiomsystemet.

`-p-q--`	(aksiom hvor vi har satt x lik «`-`»)
`-p--q---`	(fra 1 via regel)
`-p---q----`	(fra 2 via regel)

Linje 2 og 3 er *teoremer*, det vil si symbolstrenger som kan bevises eller utledes innenfor aksiomsystemet.

En computer kan lett foreta utledningene nevnt ovenfor. Slike utledninger krever kun at computeren kan gjenkjenne symbolstrengene ut fra deres fysiske form, og at vi har programmert computeren med aksiomskjemaet og regelen. Vi sier da at computeren eller maskinen *realiserer* dette aksiomsystemet. Når jeg i det følgende bruker betegnelsen «maskin» så vil jeg mene det samme som «maskin som realiserer et bestemt aksiomsystem». En slik maskin vil rent mekanisk fortsette å

[2]Se Lucas (1961) og Penrose (1990 & 1995).

[3]Basert på Hofstadter (1980, kap. 2).

utlede (bevise) teoremer. Dette innebærer at hvis en symbolstreng er et teorem, så vil maskinen finne et bevis for dette teoremet. Vi er imidlertid ikke garantert at maskinen vil finne et motbevis dersom symbolstrengen ikke er et teorem.

I et aksiomsystem er symbolene og symbolstrengene meningsløse i den forstand at de ikke representerer eller står for noe utenfor seg selv. Symbolstrengene tillegges ikke sannhetsverdier (sann eller usann) ettersom sannhet i denne sammenheng betyr det samme som korrespondanse mellom symbolstrengen og en virkelighet utenfor aksiomsystemet. Denne virkeligheten omtales gjerne som en *modell* eller *tolkning*. En symbolstreng er altså sann hvis og bare hvis den korresponderer (samsvarer eller er i overensstemmelse med) virkeligheten. Dette kan illustreres ved hjelp av et eksempel fra dagligspråket. Setningen «min bil er rød» er sann hvis og bare hvis bilen min faktisk er rød. Hvis bilen min er blå så vil ikke setningen korrespondere med virkeligheten, og den vil derfor være usann.

Symbolstrengene i et aksiomsystem beskriver i seg selv ingenting – de handler ikke *om* noe – og kan derfor ikke ha sannhetsverdier. Det er imidlertid mulig å gi symbolene en mening ved å *tolke* dem. Når vi tolker symbolene så gir vi symbolene mening ved å la dem stå for noe i virkeligheten utenfor aksiomsystemet. Symbolstrengene blir da setninger som kan være sanne eller usanne avhengig av om de korresponderer med tolkningen/modellen eller ikke. La oss med utgangspunkt i aksiomsystemet ovenfor foreta følgende tolkninger. («$\rightsquigarrow$» betyr «tolkes som».)

$$\mathtt{p} \rightsquigarrow + \quad \mathtt{q} \rightsquigarrow = \quad \mathtt{-} \rightsquigarrow 1 \quad \mathtt{--} \rightsquigarrow 2 \quad \mathtt{---} \rightsquigarrow 3 \quad \text{osv.}$$

Utledningen ovenfor blir ifølge denne tolkningen:

$1 + 1 = 2$

$1 + 2 = 3$

$1 + 3 = 4$

Målet eller hensikten med dette aksiomsystemet (som røpes først nå!) var å konstruere et aksiomsystem som skulle gjenspeile strukturen til addisjon av naturlige tall. De naturlige tallene med relasjonen addisjon blir derfor modellen til dette aksiomsystemet.[4] Tolkningen ovenfor var meningsfull, men vi kan også tolke symbolene på andre måter:

$$\mathtt{p} \rightsquigarrow \text{gener} \quad \mathtt{q} \rightsquigarrow \text{evolusjon}$$
$$- \rightsquigarrow \text{hest} \quad -- \rightsquigarrow \text{hest hest} \quad --- \rightsquigarrow \text{hest hest hest} \quad \text{osv.}$$

Utledningen ovenfor tolkes da på følgende måte:

Hest gener hest evolusjon hest hest

Hest gener hest hest evolusjon hest hest hest

Hest gener hest hest hest evolusjon hest hest hest hest

[4] Det finnes strengt tatt to måter å bedrive matematikk på: Enten så er man opptatt av å utlede logiske konsekvenser av en mengde aksiomer uten å gi disse aksiomene noen bestemt tolkning, eller så har man et håp om at en mengde aksiomer skal være sanne om en bestemt modell (for eksempel aritmetikken), og at vi på bakgrunn av dette skal kunne utlede alle aritmetiske sannheter fra aksiomene. Gödel beviste, som vi snart skal se, at dette ikke lar seg gjøre.

Dette kan betraktes som et særdeles mislykket forsøk på å aksiomatisere vår zoologiske kunnskap!

Alle utledninger innenfor et aksiomsystem er noe en computer kan utføre ettersom utledningene ikke er noe annet enn manipuleringer av symbolstrenger ved hjelp av visse fastlagte regler.[5] Det er stort sett dette datamaskiner er i stand til å gjøre.

Det er alltid mulig å omskrive et aksiomsystem til et dataprogram som deretter kan programmeres i en maskin. Et slikt dataprogram består av en mengde med regler som sier hvordan symboler og symbolstrenger skal manipuleres. I vårt eksempel består dataprogrammet av en mengde med aksiomer (bare ett aksiomskjema) og en mengde med regler (i vårt eksempel bare én regel) slik at maskinen rent mekanisk kan sette i gang med å bevise teoremer. Vi sier at en maskin som er programmert med dette dataprogrammet *realiserer aksiomsystemet*.

Det har vært et viktig mål å forsøke å konstruere aksiomsystemer som skal kunne gjenspeile vår matematiske kunnskap innenfor de ulike matematiske disiplinene. Dette innebærer at alle matematiske sannheter innenfor en bestemt matematisk disiplin (for eksempel aritmetikk) skal kunne utledes innenfor et aksiomsystem. Eller sagt på en mer presis måte: målet er at alle matematiske sannheter innenfor en matematisk disiplin skal være identisk med tolkninger av symbolstrenger som kan utledes innenfor et aksiomsystem. Bevisbare symbolstrenger kalles som sagt *teoremer*, og mange filosofer og matematikere var tidligere av den oppfatning at mengden av matematiske sannheter er identisk med mengden av teoremer, slik at matematisk sannhet kunne identifiseres med bevisbarhet. Gödel beviste i 1931 at mengden av matematiske sannheter er større enn mengden av teoremer. Det er derfor umulig å utlede alle matematiske sannheter innenfor en matematisk disiplin fra et aksiomsystem – uansett hvor mange aksiomer og slutningsregler vi legger til. Dette er essensen i Gödels første ufullstendighetsteorem, og dette er utvilsomt et av de viktigste og mest oppsiktsvekkende resultater i logikkens og matematikkens historie.

Gödels ufullstendighetsteoremer

Gödels bevis for de to ufullstendighetsteoremene er svært komplekse bevis. Bevisenes kompleksitet består i at Gödel opererer med tre ulike språklige nivåer, nemlig aksiomsystemet med dets «meningsløse» symboler og symbolstrenger, vanlig aritmetikk (de naturlige tallene, addisjon, multiplikasjon m.m.) og et metamatematisk nivå hvor vi snakker om aksiomsystemet «fra utsiden» – for eksempel om det er konsistent, fullstendig m.m. Utsagnet «dette aksiomsystemet inneholder fire aksiomer» er et eksempel på et metamatematisk utsagn fordi det sier noe om selve aksiomsystemet. I sine bevis benyttet Gödel en teknikk som i dag kalles for «Gödelnummerering». Gödelnummerering innebærer at vi kan uttrykke metamatematiske utsagn innenfor systemet selv, slik at systemet får en slags «selvbevissthet» – det vil si at det kan «snakke om seg selv». Symbolstrenger innenfor systemet kan med andre ord tolkes slik at de sier noe om systemet selv.

[5]Betegnelsene «system» og «aksiomsystem» vil i det følgende bety det samme.

Gödels *første ufullstendighetsteorem* sier at i ethvert konsistent (selvmotsigelsesfritt) aksiomsystem hvor vi kan utlede aritmetiske sannheter (det vil si at systemet inneholder blant annet operasjonene addisjon og multiplikasjon), så vil det finnes et aritmetisk utsagn G som ikke kan bevises i aksiomsystemet, men som vi kan bevise er sant. Dette utsagnet kalles aksiomsystemets *Gödelsetning*. Gödelsetningen G sin (litt merkelige) metamatematiske tolkning er at «G kan ikke bevises i dette aksiomsystemet».[6] (Jmf. poenget ovenfor om at symbolstrenger kan tolkes på svært ulike måter.) Det første ufullstendighetsteoremet innebærer altså at det finnes et utsagn G som er sant, men hvor G verken kan bevises eller motbevises i systemet. Gödel var i stand til å konstruere en slik Gödelsetning for et bestemt aksiomsystem for aritmetikk.

Ut fra G's metamatematiske tolkning er det lett å se at G må være sann. La oss først anta at G kan bevises. Dette medfører en selvmotsigelse ettersom G faktisk sier at G ikke kan bevises. (Vi forutsetter at systemet er konsistent og at vi bare kan utlede sannheter. Dette er et ufravikelig krav til ethvert aksiomsystem.) Følgelig kan ikke G bevises, men dermed må jo G være sann ettersom G jo sier at G ikke kan bevises! G er derfor en sannhet som ikke kan bevises innenfor systemet. Gödels argument har en viss likhet med det klassiske løgnerparadokset, ettersom det gjør bruk av et utsagn G som «snakker om» eller refererer til seg selv. Følgende setning er en versjon av løgnerparadokset: «Denne setningen er usann.» Dette er et paradoks, for hvis vi antar at setningen er sann så er den usann, og hvis vi antar at den er usann så er den sann. Gödelsetningen G er derimot ikke paradoksal i denne forstand, for vi kan lett se at den må være sann. Løgnerparadokset synes derimot ikke å ha noen sannhetsverdi. Vi kan ikke bevise Gödelsetningen G innenfor systemet, men vi kan ut fra argumentet ovenfor se at G må være sann. G spiller en viktig rolle i Lucas/Penrose-argumentet. Dette vil jeg komme tilbake til senere.

Gödels *andre ufullstendighetsteorem* sier at det ikke er mulig å bevise innenfor et aksiomsystem at det er konsistent (selvmotsigelsesfritt). Hvis aksiomsystemet er konsistent, så kan det med andre ord ikke bevise sin egen konsistens.[7] Gödel beviste det andre ufullstendighetsteoremet ved å konstruere et aritmetisk utsagn som kan tolkes (metamatematisk) som at «dette systemet er konsistent». Han beviste deretter at dette utsagnet ikke lar seg bevise innenfor systemet.

Gödels ufullstendighetsteoremer har utvilsomt hatt stor innvirkning på matematisk logikk og matematikkens filosofi. Gödel selv mente at det første ufullstendighetsteoremet kunne benyttes som et premiss i et argument for matematisk platonisme, det vil si det synspunkt at det finnes abstrakte objekter hinsides tid og rom, og at det er matematikerens oppgave å oppdage sannheter om disse objektene. Dette er et kontroversielt synspunkt.

Gödel synes imidlertid å ha bevist at bevisbarhet og sannhet ikke er det samme, og dette åpner for at sannhet i matematikk er «noe mer» enn bare å bevise noe ut

---

[6] Mer presist så er G et *aritmetisk utsagn* som også kan formuleres (eller har sitt motstykke) innenfor aksiomsystemet (det vil si at G kan uttrykkes ved hjelp av *aksiomsystemets symboler*), og utsagnet G kan dessuten gis en bestemt *metamatematisk tolkning*, nemlig at «G ikke kan bevises innenfor aksiomsystemet». Uttrykkene i kursiv blinker ut de tre ulike språknivåene som Gödel opererer med i sitt bevis.

[7] Fra dette følger at hvis vi kan bevise at et system er konsistent, så er det ikke konsistent. Dette tilsynelatende paradoksale resultatet følger fra det logiske prinsippet at fra en selvmotsigelse så kan vi utlede eller bevise hva som helst.

fra aksiomer. Også dette er imidlertid kontroversielt, og det faller utenfor rammene av denne artikkelen å drøfte disse problemene nærmere.

Lucas/Penrose-argumentet mot computerhypotesen

Gödel var den første som så at ufullstendighetsteoremene kunne benyttes som premisser i et argument mot computerhypotesen. Lucas og Penrose har imidlertid gjort dette argumentet mer eksplisitt og detaljert. Lucas og Penrose hevder at gitt en maskin M som realiserer et aksiomsystem S, så vil det alltid være mulig for et menneske å se at Gödelsetningen til M (eller S) er sann, men maskinen kan derimot ikke konkludere med at Gödelsetningen er sann fordi maskinen kan ikke bevise den. Vi kan da spørre om det ikke er mulig å konstruere en maskin som er så avansert at den kan gi som output *alle* de matematiske sannhetene som et menneske kan bevise. Argument A nedenfor er ment å vise at dette ikke er mulig, det vil si at computerhypotesen er feil. Argument A sier i kortformat at enhver maskin inneholder en Gödelsetning som ikke kan bevises av maskinen selv, men som vi mennesker kan se er sann. Vi har med andre ord en ferdighet som maskiner ikke kan ha. Ergo er computerhypotesen feil. Mine matematiske ferdigheter består med andre ord ikke i at jeg er en maskin M som realiserer et aksiomsystem, fordi jeg har en evne til å se G's sannhet som M ikke har. (G er Gödelsetningen til M.) La oss sette opp dette argumentet på en mer oversiktlig måte.

Argument A Hvis Kurt er en maskin M som realiserer et konsistent aksiomsystem S, så vil han ikke være i stand til å bevise at Gödelsetningen til M (eller S) er sann. (Dette følger fra det første ufullstendighetsteoremet som innebærer at M eller S ikke kan bevise sin egen Gödelsetning.) Kurt kan derimot bevise at Gödelsetningen til M (og S) er sann ved å anvende Gödels fremgangsmåte.[8] Ergo er Kurt ikke en maskin, og computerhypotesen er feil.

Det første premisset bygger på det faktum at en maskin bare kan holde det som kan *bevises* innenfor aksiomsystemet for å være sant. Det andre premisset fastslår at mennesker (for eksempel Kurt) kan bevise en sannhet, nemlig Gödelsetningen til en maskin, som maskinen selv ikke kan bevise. Dette viser at vi ikke er maskiner, det vil si at våre matematiske ferdigheter ikke er bestemt av at vi er maskiner som realiserer aksiomsystemer.

Man kan innvende mot dette argumentet at det burde være mulig å legge til et program P som kan konstruere Gödelsetningen G til M slik at denne nye maskinen M' (som er M programmert med P) kan oppvise de samme matematiske ferdighetene som Kurt. Lucas hevder imidlertid at M' vil ha en Gödelsetning G' som Kurt kan se er sann, men som M' ikke kan bevise. Vi kan da konstruere en ny maskin M'' ved å legge inn et program P' i M' slik at M'' kan bevise både G og G'. M'' vil derimot ha en Gödelsetning G'' som Kurt kan se er sann, men som M'' ikke kan bevise osv. For enhver maskin $M', M'', M''' \dots$ så vil altså Kurt «ligge et hode foran». (Det er for øvrig viktig å legge merke til at det ikke er mulig å konstruere et dataprogram som kan bevise G til enhver maskin.) Konklusjonen er følgelig at

[8] For eksempel den som Gödel selv benyttet i Gödel (1931).

vi har en ferdighet som maskiner ikke kan ha, og vi er derfor ikke maskiner. Dette innebærer at computerhypotesen er feil.

Argument B Et annet argument mot computerhypotesen (basert på Gödels andre ufullstendighetsteorem) sier at ingen maskin kan vite at den er konsistent, men vi mennesker kan derimot vite at vi er konsistente.[9] Følgelig har mennesker en egenskap som maskiner ikke har.
Ergo er mennesker ikke maskiner, og computerhypotesen er feil.

Jeg skal senere vise at begrepet konsistens utgjør selve akilleshælen til alle de nevnte argumentene mot computerhypotesen.

Argumentene A og B utgjør til sammen Lucas/Penrose-argumentet mot computerhypotesen.

Kritikk av Lucas/Penrose-argumentet

Det er vanskelig å skille kritikken av argumentene A og B fra hverandre, men jeg skal likevel forsøke å gjøre det for oversikten sin del. En kritikk som rammer argument A vil som regel også ramme argument B (og omvendt).

Kritikk av argument A

Det er viktig å legge merke til at ifølge Gödels første ufullstendighetsteorem så *kan vi bare bevise at en Gödelsetning er sann hvis vi kan bevise at systemet er konsistent*. Det andre premisset i argument A er sånn sett et tvilsomt premiss, og det er nettopp dette premisset jeg vil kritisere.

Anta at vi har konstruert et program (ut fra et bestemt aksiomsystem) som vi deretter programmerer en maskin med. Spørsmålet er nå om denne maskinen M vil være i stand til å simulere (etterligne) matematikernes ferdigheter. Nesten all matematikk kan utledes fra mengdelæren, og vi kan derfor late som om mengdelæren er det aksiomsystemet som best uttrykker matematikernes ferdigheter. Vi lar med andre ord M realisere mengdelæren. Det følger da fra Gödels andre ufullstendighetsteorem at M ikke kan bevise at mengdelæren er konsistent, og M kan følgelig ikke bevise at Gödelsetningen G til mengdelæren er sann. Men vi mennesker kan heller ikke bevise at mengdelæren er konsistent, og derfor kan vi (eller Kurt!) ikke bevise at den tilhørende Gödelsetningen G er sann. Det eneste vi og M kan bevise er at hvis systemet er konsistent, så er Gödelsetningen sann.

Det er mulig at mengdelæren faktisk er inkonsistent. (Den bygger blant annet på en del kontroversielle aksiomer.) Det er sånn sett mulig at vi befinner oss i samme situasjon som logikeren Gottlob Frege da han i 1902 mottok et brev fra filosofen Bertrand Russell som påviste en inkonsistens i hans mengdelære. Frege hadde arbeidet i flere tiår med sin teori, men hadde likevel oversett en inkonsistens. Verken vi eller en maskin kan med andre ord bevise eller «se» at mengdelæren eller matematikken som sådan er konsistent. Følgelig kan verken vi eller maskiner bevise at Gödelsetningene er sanne. Her stiller vi med andre ord likt med maskinene.

[9] Det faktum at en maskin i kraft av å realisere et aksiomsystem ikke kan bevise sin egen konsistens innebærer at den ikke kan bevise (eller vite) at den ikke en gang i fremtiden vil skrive en selvmotsigelse, slik som for eksempel «0 = 1».

Lucas hevdet i argumentet ovenfor at selv om vi konstruerer stadig mer avanserte maskiner (M', M'', M''' osv.), så kan vi stå utenfor en hvilken som helst av disse maskinene og konstruere en Gödelsetning som maskinen selv ikke kan bevise. Det er imidlertid tvilsomt om vi faktisk er i stand til å konstruere Gödelsetningen til enhver maskin. Denne vanskeligheten bunner blant annet i det faktum at det ikke lar seg gjøre å lage et dataprogram som vi kan bruke for å konstruere Gödelsetninger for vilkårlige maskiner. Det kan med andre ord bli svært vanskelig for oss å konstruere Gödelsetninger til komplekse maskiner. Som vi har sett kan vi heller ikke vite at Gödelsetningene som vi klarer å konstruere faktisk er sanne, ettersom dette forutsetter at vi må vite at maskinene er konsistente. For komplekse maskiner kan det imidlertid være svært vanskelig å avgjøre hvorvidt maskinene er konsistente eller ikke. Også her stiller vi derfor likt med maskinene.

La oss oppsummere og eksplisere argument A mot computerhypotesen. Anta at det finnes en maskin M som realiserer det samme aksiomsystemet som jeg, det vil si at M kan bevise alle de aritmetiske utsagnene som også jeg kan bevise. Ved hjelp av Gödels argument kan jeg bevise Gödelsetningen G til M, men denne kan imidlertid ikke M selv bevise. Ergo kan ikke mine matematiske ferdigheter bestå i at jeg er en maskin som realiserer et bestemt aksiomsystem, fordi jeg har en evne som M ikke har, nemlig en evne til å bevise at G er sann.

Jeg kritiserte dette argumentet fordi vi faktisk ikke kan bevise at Gödelsetningen G til M er sann. Vi kan bare bevise følgende kondisjonal:[10] *hvis M er konsistent, så er G sann.* Verken M eller jeg kan bevise G, men både M og jeg kan bevise kondisjonalen.[11] Vi stiller med andre ord helt likt med maskinene. Hvis M er en kompleks maskin, noe som i dette tilfellet er svært plausibelt, så vil det være svært vanskelig å bevise at M er konsistent, og dermed vil det være vanskelig for oss å bevise at G er sann. I motsetning til hva Lucas og Penrose hevder er det tvilsomt om vi har en evne til å «se» M's konsistens og G's sannhet direkte (hva nå det skulle bety). Det er vanskelig å spesifisere hva en slik evne skulle bestå i, særlig tatt i betraktning eksemplet ovenfor som viste at selv begavede matematikere har «viklet seg inn i» inkonsistenser. Heller ikke matematikere er ufeilbarlige!

Kritikk av argument B

I det andre ufullstendighetsteoremet beviser Gödel at aksiomsystemer eller maskiner ikke kan bevise sin egen konsistens. Men kan *vi* det? Lucas og Penrose svarer ja på dette spørsmålet, og fastslår dermed at vi har en egenskap som maskiner ikke har. Følgelig er computerhypotesen feil.[12]

Lucas presenterer et argument for at vi (i motsetning til M) har en evne til å vite at vi er konsistente, fordi et inkonsistent system vil generere «massevis med tull og tøys», noe vi vanligvis ikke gjør.[13] Dette følger fra det logiske prinsippet (som vi ikke skal bevise her) at fra en inkonsistens så kan vi utlede hva som helst. Men anta nå at mine oppfatninger er av en slik karakter at jeg kunne ha utledet

[10] En kondisjonal er kort fortalt en «hvis-så»-setning.

[11] Se Chihara (1972: 507–508).

[12] Lucas (1961) sier at vi må være «*basically consistent*», og Penrose argumenterer for det samme. Uttrykket «basically consistent» virker imidlertid litt merkelig ettersom konsistens er et enten-eller fenomen. Enten så er et system konsistent eller så er det inkonsistent. På samme måte er en kvinne enten gravid eller ikke-gravid. Konsistens og graviditet kommer ikke i grader.

[13] Se Lucas (1961).

både en setning P og dens negasjon (benektelse) ikke-P, men uten at jeg faktisk gjør det. Det er vanskelig å se at det fra dette skulle følge at jeg vil si omtrent hva som helst av tull og tøys. Det kan med andre ord finnes en inkonsistens gjemt et eller annet sted i oss eller i maskiner (en lokal eller isolert inkonsistens) uten at dette vil ha de dramatiske følgene som Lucas skisserer.[14]

Det er rimelig å anta at et normalt menneske ikke vil akseptere en inkonsistent mengde med oppfatninger som vedkommende har rettet oppmerksomheten sin mot. De fleste av våre oppfatninger har vi imidlertid *ikke* rettet oppmerksomheten vår mot. De fleste av oss tror at leger går med undertøy og at $11 + 11 = 22$, men det er svært sjelden vi retter oppmerksomheten mot disse oppfatningene. Det er derfor plausibelt at det kan gjemme seg minst én inkonsistens blant våre millioner av oppfatninger, men uten at dette vil ha slike radikale konsekvenser som Lucas skisserer. Det er faktisk stor grunn til å tro at det finnes slike inkonsistenser i vårt enorme nettverk av oppfatninger. Rasjonalitet er som kjent et ideal vi streber mot, men aldri når.

Anta at en maskin M simulerer en matematikers evner. Jeg har vist ovenfor at M ikke kan bevise at den er konsistent, men det kan heller ikke vi. Det er med andre ord like vanskelig for oss som for maskiner å påvise egen og andres konsistens. Mennesker synes med andre ord å ha de samme begrensninger som maskiner, og derfor *kan* vi være maskiner.[15]

Svakheten med argumentene A og B (Lucas/Penrose-argumentet) er ikke at disse argumentene fratar maskinene evner som de faktisk besitter, men at Lucas og Penrose gir mennesker en evne til å «se» matematiske sannheter og til å «se» at systemer er konsistente som vi faktisk ikke har. Feilen ligger ikke i en nedvurdering av maskinenes potensielle ferdigheter, men i et urealistisk syn på menneskets matematiske ferdigheter.

Konklusjon

Hensikten med denne kritikken av Lucas/Penrose-argumentet er ikke å argumentere for computerhypotesen. Det er en hypotese jeg ikke tilslutter meg. Det finnes mange gode argumenter mot denne hypotesen, men Lucas/Penrose-argumentet er ikke et av dem. Computerhypotesen er antakeligvis altfor enkel når det gjelder å forklare menneskers komplekse ferdigheter og adferd. Hvordan vi skal forklare våre ferdigheter, opplevelser og adferd er naturligvis omfattende psykologisk spørsmål, men det synes uansett å være urealistisk å tro at våre matematiske ferdigheter bare består i å utlede teoremer fra aksiomer.

[14] Man kan til og med argumentere for at inkonsistens i en viss forstand er ønskelig og rasjonelt. Det er for eksempel rasjonelt av deg å tro at du, som ikke er et allvitende vesen, har noen usanne oppfatninger. Men enhver person som har en oppfatning om at han eller hun har minst én usann oppfatning, vil ha en inkonsistent mengde med oppfatninger. En mengde med oppfatninger er konsistent hvis og bare hvis det er mulig for alle oppfatninger å være sanne, noe som i dette tilfellet er umulig. I denne forstand kan man si at det er rasjonelt å være inkonsistent. Man kan kanskje betegne dette som en «godartet inkonsistens». Det følger fra dette at mennesker som sier at de vet at de er konsistente ikke kan være rasjonelle!

[15] Se Putnam (1995).

Referanser

Chihara, C.S. (1972). On alleged refutations of mechanism using Gödel's incompleteness results. *Journal of Philosophy* **64**, 507–526.

Gödel, K. (1931). Über formal unentschiedbare Sätze per Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physic* **38**, 173–198.

Hofstadter, D.R. (1980). *Gödel, Escher, Bach: an Eternal Golden Braid.* London: Penguin Books.

Lucas, J.R. (1961). Minds, Machines and Gödel, *Philosophy* **36**, 112–127.

Nagel, E. & Newman, J.R. (1989). *Gödel's Proof.* London: Routledge.

Penrose, R. (1990). *The Emperor's New Mind.* London: Vintage.

Penrose, R. (1995). *Shadows of the Mind.* London: Vintage.

Putnam, H. (1995). (Anmeldelse av Penrose (1995).) *Bulletin of the American Mathematical Society* **32**, 370–373.

Smullyan, R. (1992). *Gödel's Incompleteness Theorem.* Oxford: Oxford University Press.