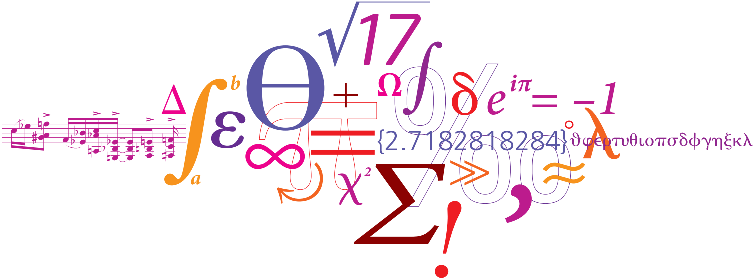
Gödels ufuldstændighedssætninger

Thomas Bolander, DTU Compute

UNF, DTU, 23. oktober 2019



DTU Compute
Institut for Matematik og Computer Science

Lidt om mig selv

Thomas Bolander

- Professor i logik og kunstig intelligens på **DTU Compute**.
- **Aktuel forskning**: Sociale aspekter af kunstig intelligens.
- Medlem af kommissioner og tænketanke indenfor kunstig intelligens og dets samfundsmæssige aspekter, herunder *SIRI-kommissionen* og *TechDK-kommissionen*.
- H.C. Ørsted sølvmedalje for fremragende videnskabsformidling, 2019.
- Med-arrangør og videnskabelig rådgiver for *Science & Cocktails*.

Pinocchios paradoks

Hvad kan Pinocchio sige som vil skabe en selvmodsigelse (et paradoks)?

Gödels Sætning

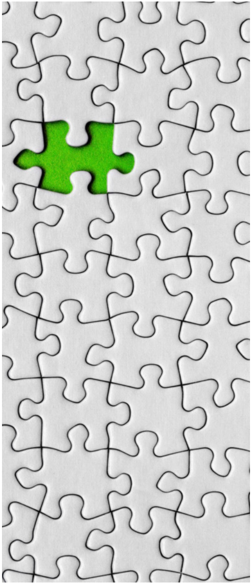
- Det fulde navn er "Gödels første ufuldstændighedssætning".
- Publiceret i 1931 af Kurt Gödel (som 25-årig!).
- En af det 20. århundredes mest berømte matematiske sætninger.

Populær formulering:

> *Der findes sande matematiske sætninger som ikke kan bevises.*

Truer **matematikkens grundlag** som netop er sikkerhed gennem bevis.

Revolutionerende: Bruger matematiske metoder til at vise begrænsningen i matematiske metoder.



Den oprindelige formulering

Den populære formulering (*sande sætninger som ikke kan bevises*) er naturligvis for upræcis til at kunne bevises. Hvorfor?

Oprindelig formulering af Gödels sætning:

> *Zu jeder ω-widerspruchsfreien rekursiven Klasse κ von Formeln gibt es rekursive Klassenzeichen r, so daß weder v Gen r noch Neg(v Gen r) zu Flg(κ) gehört (wobei v die freie Variable aus r ist).*

Denne formulering giver ikke mening for særligt mange mennesker, så lad os prøve at simplificere den lidt...

Simplificeret formulering

Simplificeret formulering af Gödels sætning:

> *I ethvert* **konsistent formelt system** *som indeholder talteori findes udsagn som hverken kan bevises eller modbevises.*

Intuitivt: Systemet kan udtrykke udsagn, som det ikke selv kan afgøre om er sande eller ej.

Eksempel. Godbach-formodningen (1742) siger at ethvert lige tal er summen af to primtal. Vi ved endnu ikke om formodningen kan bevises eller modbevises. Gödel siger: Der findes formodninger som **hverken** kan bevises eller modbevises, altså som vi aldrig kan få svar på om er rigtige eller forkerte (indenfor ethvert givet system).

Men formuleringen er stadig kryptisk. Hvad er et formelt system? Og hvad er talteori? Og hvad er konsistens?

Inden vi besvarer disse spørgsmål, så lad os se på nogen af de fortolkninger som resultatet har givet anledning til...

Formelle systemer

- Gödels sætning udtaler sig i første omgang kun om **formelle systemer**.
- Formelle systemer er formelle, matematiske strukturer som **modellerer** en del af verden.

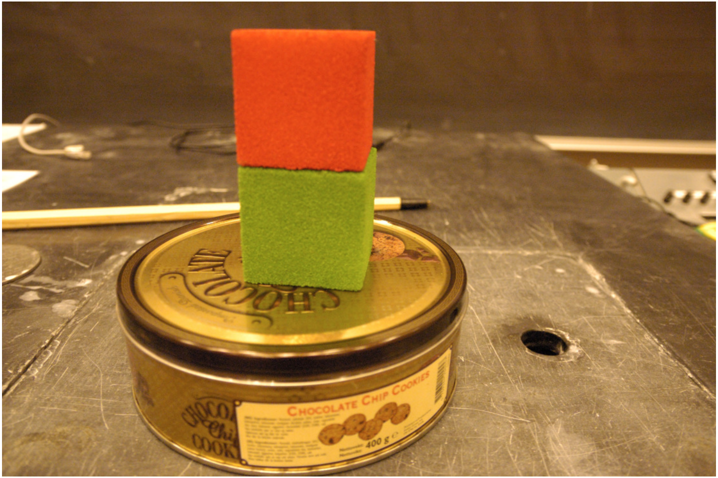
model

$\longrightarrow$

modelleret virkelighed

Modeller: Billeder

En model kan være et billede:



Modeller: Sproglige beskrivelser

En model kan være en sproglig beskrivelse:

1. Småkagedåsen ligger på bordet
2. Den grønne terning ligger på småkagedåsen
3. Den røde terning ligger på den grønne terning
4. Kagedåsen ligger over bordet
5. Den grønne terning ligger over kagedåsen
6. Den grønne terning ligger over bordet
7. Den røde terning ligger over den grønne terning
8. Den røde terning ligger over kagedåsen
9. Den røde terning ligger over bordet
10. Bordet ligger ikke på småkagedåsen
11. Småkagedåsen ligger ikke på den grønne terning

⋮

Modeller: Formler

En model kan være en **formaliseret** sproglig beskrivelse, altså en mængde af **formler**:

1. $\mathit{på}(\mathit{kage}, \mathit{bord})$
2. $\mathit{på}(\mathit{grøn}, \mathit{kage})$
3. $\mathit{på}(\mathit{rød}, \mathit{grøn})$
4. $\mathit{over}(\mathit{kage}, \mathit{bord})$
5. $\mathit{over}(\mathit{grøn}, \mathit{kage})$
6. $\mathit{over}(\mathit{grøn}, \mathit{bord})$
7. $\mathit{over}(\mathit{rød}, \mathit{grøn})$
8. $\mathit{over}(\mathit{rød}, \mathit{kage})$
9. $\mathit{over}(\mathit{rød}, \mathit{bord})$
10. $\neg\, \mathit{på}(\mathit{bord}, \mathit{kage})$ "$\neg$" læses "ikke" (negation)
11. $\neg\, \mathit{på}(\mathit{kage}, \mathit{grøn})$

$\vdots$

Modeller: Formelle systemer

En model kan være et **formelt system**: en mængde af **aksiomer** (formler) og en række **slutningsregler**:

Aksiomer:

(A1) $på(kage, bord)$

(A2) $på(grøn, kage)$

(A3) $på(rød, grøn)$

(A4) $på(x, y) \Rightarrow over(x, y)$ "$\Rightarrow$" læses "medfører" (hvis ... så ...)

(A5) $over(x, y)\ \&\ over(y, z) \Rightarrow over(x, z)$

(A6) $på(x, y) \Rightarrow \neg\, på(y, x)$

Slutningsregler:

(S1) Udfra $\phi \Rightarrow \psi$ og ϕ kan sluttes ψ.

(S2) Udfra ϕ og ψ kan sluttes $\phi\ \&\ \psi$.

(S3) Variable (f.eks. x) kan erstattes med konstanter (f.eks. $rød$).

Eksempel på bevis i det formelle system

Vi beviser at den røde terning er over kagedåsen:

1.	$på(grøn, kage)$	aksiom (A1)
2.	$på(rød, grøn)$	aksiom (A3)
3.	$på(x, y) \Rightarrow over(x, y)$	aksiom (A4)
4.	$på(grøn, kage) \Rightarrow over(grøn, kage)$	regel (S3) på linje 3
5.	$på(rød, kage) \Rightarrow over(rød, kage)$	regel (S3) på linje 3
6.	$over(grøn, kage)$	regel (S1) på linje 4, 1
7.	$over(rød, grøn)$	regel (S1) på linje 5, 2
8.	$over(rød, grøn)\ \&\ over(grøn, kage)$	regel (S2) på linje 6
9.	$over(x, y)\ \&\ over(y, z) \Rightarrow over(x, z)$	aksiom (A5)
10.	$over(rød, grøn)\ \&\ over(grøn, kage) \Rightarrow over(rød, kage)$	regel (S3) på linje 9
11.	$over(rød, kage)$	regel (S1) på linje 10,8

Formelle systemer

Formelle systemer består af en mængde **aksiomer** og en mængde **slutningsregler**.

Et **bevis** i et formelt system er en endelig sekvens af formler hvor der for hver formel gælder en af følgende:

1. Formlen er et **aksiom**.
2. Formlen følger af de foregående formler ved brug af en af **slutningsreglerne**.

En formel ϕ kan **bevises** (er **bevisbar**) hvis der findes et bevis for den. Den kan **modbevises** (er **modbevisbar**) hvis negationen $\neg\phi$ (det modsatte udsagn) kan bevises.

Eksempler. Formlen $\mathit{over}(\mathit{rød}, \mathit{kage})$ er bevisbar (se foregående slide). Formlen $\mathit{på}(\mathit{bord}, \mathit{kage})$ er modbevisbar (pga aksiom (A6)).

Eksempel: MUI-systemet (Douglas Hofstadter)

Alle formler bygges op af bogstaverne **M**, **I** og **U**. Eksempelvis: **MU**, **UIM**, **MUUMUU**.

Aksiomer. Kun ét aksiom, formlen **MI**.

Slutningsregler:

Regel 1. Hvis en formel slutter på **I**, kan man tilføje et **U** til enden.

Regel 2. Udfra **M**x, hvor x er en sekvens af bogstaver, kan man danne **M**xx.

Regel 3. Udfra x**III**y kan man danne x**U**y (man kan altså erstatte delsekvensen **III** med **U**).

Regel 4. Udfra x**UU**y kan man danne xy (man kan altså slette delsekvensen **UU**).

Eksempel. **MI** $\xrightarrow{\text{regel 1}}$ **MIU** $\xrightarrow{\text{regel 2}}$ **MIUIU**.

I par:

1. Udled **MUIIU**.
2. Prøve at udlede **MU**. Kan det lade sig gøre eller ej?

Løsninger

Husk slutningsreglerne:

Regel 1. Hvis en formel slutter på **I**, kan man tilføje et **U** til enden.
Regel 2. Udfra **M**x, dan **M**xx.
Regel 3. Udfra x**III**y, dan x**U**y.
Regel 4. Udfra x**UU**y, dan xy.

Løsninger:

1. Udledning af **MUIIU**: **MI** $\xrightarrow{\text{regel 2}}$ **MII** $\xrightarrow{\text{regel 2}}$ **MIIII** $\xrightarrow{\text{regel 1}}$ **MIIIIU** $\xrightarrow{\text{regel 3}}$ **MUIU** $\xrightarrow{\text{regel 2}}$ **MUIUUIU** $\xrightarrow{\text{regel 4}}$ **MUIIU**.
2. **MU** kan ikke udledes: Antag **MU** kunne udledes. Så kunne vi altså udlede en formel uden **I**er. Antallet af **I**er ville således være et multiplum af 3. Men så må antallet af **I**er heletiden have været et multiplum af 3, for regel 1 og 2 kan slet ikke ændre på antallet af **I**er, og reglerne 2 og 3 kan ikke lave en formel hvor antallet af **I**er **ikke** er et multiplum af 3 om til en formel hvor det er. Dermed må alle formler i beviset have et antal **I**er som er et multiplum af 3. Men det er en modstrid, for vores eneste aksiom **MI** har kun ét **I**!

Formelle systemer for talteori

Gödel ser på formelle systemer for **talteori** (teorien om de naturlige tal og deres egenskaber).

Formelt system for talteori:

- **Aksiomerne** inkluderer bl.a.:
  - $x + 0 = x$
  - $x + 1 \neq 0$
  - $x \cdot (y + 1) = x \cdot y + x$
- **Slutningsreglerne** inkluderer bl.a.:
  - Udfra $\phi \Rightarrow \psi$ og ϕ kan sluttes ψ.

Konsistens: Et formelt system kaldes **konsistent** hvis der ikke findes formler som både kan bevises og modbevises.

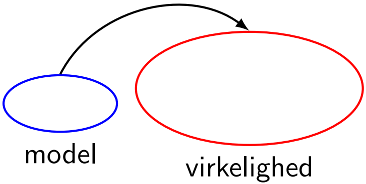
Nu kan vi bedre forstå den simplificerede udgave af Gödels sætning:

> *I ethvert* **konsistent formelt system** *som indeholder talteori findes formler som hverken kan bevises eller modbevises.*

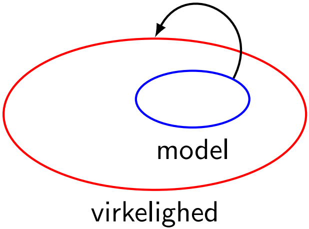
Modeller der modellerer sig selv

Formelle systemer er som sagt en slags **modeller**. To typer af modeller:

1. **Modeller som er komplet adskilt fra den virkelighed de modellerer**. F.eks. Legolands model af Big Ben eller det formelle system for kagedåserne.



2. **Modeller som selv er en del af den virkelighed de modellerer**. F.eks. en model af Legoland inde i Legoland. Eller et menneske der forsøger at skabe en model af sig selv.



Hvilken type model er mon et formelt system for talteori?

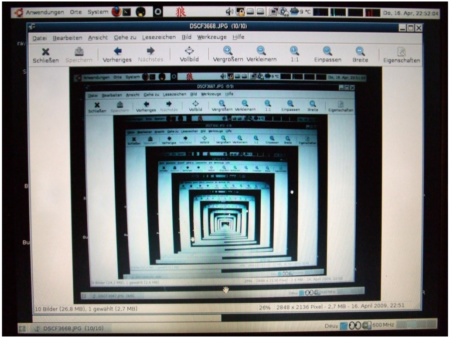
Modeller som modellerer sig selv

Modeller af den anden type giver ofte anledning til problemer i form af **selvreference**.

Selvreference bruges om objekter som **refererer til sig selv**. Det sker ved at en **del** af objektet **refererer til** objektet som helhed.

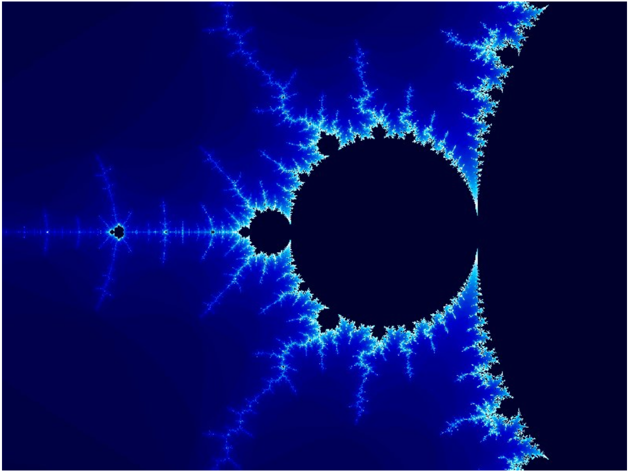
Eksempler på selvreference

- **Sætninger**: “Denne sætning indeholder 5 ord.”
- **Billeder**:



Flere eksempler på selvreference

- **Regler**: "Alle regler har en undtagelse".
- **Fraktaler**:



Problemet med selvreference (og modeller som modellerer sig selv)

Problemet med selvreference er at det kan lede til **paradokser**.

Paradoks: Et tilsyneladende korrekt ræssonement der, baseret på tilsyneladende korrekte antagelser, leder til en modstrid.

Et af de mest berømte paradokser er **løgnerparadokset**...

Løgnerparadokset bygger på **løgnersætningen** L:

L: Denne sætning er falsk.

Spørgsmålet er nu: Er løgnersætningen L sand eller falsk?

Egenskaber: Selvrefererende + paradoksal.

Trump på Twitter

Fra løgnersætning til Gödelformel

Gödels bevis bygger på **Gödelformlen** G som udtrykker:

G: Denne formel kan modbevises.

(Bemærk selvreference & relation til løgnersætning).

Spørgsmålet er nu: Kan formlen G bevises eller ej?

1. G **kan** bevises $\Rightarrow$ G er sand $\Rightarrow$ G kan modbevises.
2. G **kan ikke** bevises $\Rightarrow$ vi kan ikke bevise at G kan modbevises $\Rightarrow$ vi kan ikke modbevise G.

Konklusion. Der må gælde en af følgende:

1. G kan **både** bevises og modbevises.
2. G kan **hverken** bevises eller modbevises.

Altså: Hvis systemet er konsistent (ikke 1), så er det findes der formler som hverken kan bevises eller modbevises (2). Dette er netop formuleringen i **Gödels sætning**!

Fra Gödelformel til Gödels sætning

Hvad vi har (fra foregående slide):

> *I ethvert konsistent system som kan udtrykke Gödelformlen findes der formler som hverken kan bevises eller modbevises.*

Hvad vi ville vise (jvf. starten af forelæsning):

> *I ethvert konsistent formelt system som indeholder talteori findes der formler som hverken kan bevises eller modbevises.*

Hvad vi mangler at vise: Formelt system indeholder talteori ⇒ kan udtrykke Gödelformel. (Altså at talteori er tilstrækkeligt til at systemet kan modellerer sig selv og dermed referere til sig selv).

Gödelformel ⇒ ufuldstændighed	+	talteori ⇒ Gödelformel	=	talteori ⇒ ufuldstændighed

Talteori

Formelt system indeholder talteori hvis:

- Kan udtrykke egenskaber ved de naturlige tal $0, 1, 2, \ldots$.
- Har passende **aksiomer** for de naturlige tal, f.eks. $x + 1 \neq 0$.

Mål: Indenfor sådanne systemer at udtrykke Gödelformlen:

Denne formel kan modbevises.

Udfordring:

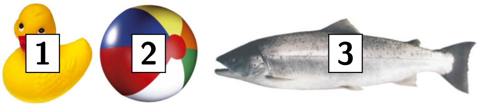
- Formler i talteori "taler om" tal og egenskaber ved tal.
- Vi skal have dem til i stedet at "tale om" formler og egenskaber ved formler (så vi kan danne Gödelformlen).

Antag de eneste objekter du må tale om er tal. Hvordan vil du da tale om objekterne i dette rum?

Gödel-nummerering

Gödel-nummerering: Hver formel φ i det formelle system får et entydig nummer (et "CPR-nummer"). Dette nummer kaldes **Gödelnummeret** af formlen, og betegnes $\langle\varphi\rangle$.

Pointen er: Formelle systemer indeholdende talteori kan da udtrykke egenskaber ved formler via deres Gödelnumre.



$$\mathbf{3} > \mathbf{2} \wedge \mathbf{2} > \mathbf{1}$$

Selvreference i formelle systemer

Vigtigste egenskab ved formler: Er de bevisbare eller ej?

Denne egenskab indfanger Gödel i en formel $\mathit{bevisbar}(x)$.

Der gælder:

$$\text{formlen } \varphi \text{ kan bevises} \quad \Leftrightarrow \quad \text{formlen } \mathit{bevisbar}(\langle\varphi\rangle) \text{ kan bevises.}$$

Intuitivt: Formlen $\mathit{bevisbar}(x)$ tillader systemet at "tale om" hvilke formler der er bevisbare **i systemet selv** (via deres Gödelnumre)!

Formlen $\mathit{bevisbar}(x)$ er et "alvidende orakel" for systemet.

Hvad er relationen til selvreference?

Gödelformlen

Gödel konstruerer en formel G så følgende er bevisbart:

$$G \Leftrightarrow bevisbar(\langle \neg G \rangle).$$

Intuitivt: G udtrykker at dens egen negation er bevisbar. Altså: "Denne formel kan modbevises." Dvs. Gödelformlen.

Konklusion: Gödelformlen kan udtrykkes i ethvert system som indeholder talteori.

Opsummering af Gödels sætning

Gödels sætning. *I ethvert konsistent formelt system som udvider talteori findes der formler som hverken kan bevises eller modbevises.*

Bevisskitse:

- Introducér Gödelnummerering i systemet (via talteori).
- Konstruér formlen *bevisbar*(x) som udtrykker bevisbarhed indenfor systemet selv (via Gödelnumre).
- Konstruér formlen G som udtrykker sin egen modbevisbarhed.
- Der må nu gælde en af følgende:
  1. G kan **både** bevises og modbevises.
  2. G kan **hverken** bevises eller modbevises.
- Da vi har antaget konsistens, må 2 gælde. Dette er den ønskede konklusion, og beviset er hermed fuldført.

Konsekvenser af Gödels sætning

Umiddelbart siger Gödels sætning kun noget om **formelle systemer**.

- Svarer sædvanlig matematisk praksis til et formelt system? I så fald siger Gödels sætning også noget om begrænsninger i den matematiske metode.
- Er mennesker underlagt de samme begrænsninger i ræssonering som formelle systemer? I så fald siger Gödels sætning potentielt noget om grænser for vores erkendelse. Hvis ikke, siger den noget menneskers overlegenhed i forhold til maskiner.

Gödels sætning er med sikkerhed sand: et matematisk resultat omkring matematiske strukturer (formelle systemer). Men matematisk praksis og menneskelig tænkning er ikke matematiske strukturer, så dem kan vi ikke sige noget om med samme matematiske sikkerhed.

Albert Einstein: *"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."*

Fortolkninger af Gödels sætning

Gödels sætning er blevet givet meget **vidtrækkende fortolkninger**, f.eks.:

- *Der findes sande matematiske sætninger som ikke kan bevises.*
- *Der er grænser for vores mulighed for erkendelse af sandhed gennem logiske argumenter.*
- *Der er grænser for den rationelle tankegangs rækkevidde.*
- *Mennesker er klogere end maskiner (John R. Lucas, Roger Penrose).*
- *Modsat computere har mennesker en fri vilje (John R. Lucas).*
- *"Gödel beviste at mennesker ved mere end de kan vide, hvorfra de ved" (Tor Nørretranders i Mærk Verden).*

Åbent problem: Er alle paradokser uden selvreference ens?