BEREGNBAR METODOLOGI

Et studie
i
Gärdenfors'
metodologi

Klaus Frovin Jørgensen, Modul 2, Forår 1999.
Matematik, Roskilde Universitetscenter.

Abstract

Gärdenfors-metodologien undersøges i denne tekst med hensyn til beregnbarhed og pålidelighed. Der angives rekursivt en beregnbar Gärdenfors-opdatering, og det vises, at den netop kan identificere åbne mængder i Borel-hierarkiet. Dernæst vises det, at metodologien ikke er en fuldstændig arkitektur, da kravet om data retention til metoden giver anledning til restriktivitet.

En stor tak til mine eksterne vejledere
Vincent R. Hendricks og Stig Andur Pedersen.

Indholdsfortegnelse

Indledning

Den matematiske metode bærer præcise resultater med sig, hvilket man med stor succes gør brug af i videnskaberne—især de natur- og samfundsvidenskabelige. Disse resultater opnået ved hjælp af matematiske modeller skal selvfølgelig altid ses i lyset af de præmisser, modellerne bygger på. Men accepteres præmisserne, følger resultaterne med sikkerhed—givet at kun gyldige slutningsregler er blevet anvendt. Inden for erkendelsesteorien og videnskabsfilosofien derimod er det ikke de definitive og præcise resultater, der hersker. Det kan måske heller ikke være radikalt anderledes, da disse videnskaber nok i deres natur er mindre absolutte end eksempelvis fysikken. Det er dog min hypotese, at der inden for erkendelsesteorien og videnskabsfilosofien kan opnås mange nye og interessante resultater gennem en kritisk brug af matematik. Denne tekst, som giver en matematisk analyse af et sæt metodologiske forskrifter, skulle gerne ses som et eksempel på dette.[1] En af pointerne er, at det er muligt, at valget af metode er afhængigt af en individuel balance mellem eksempelvis pålideligheds- kohærens- og konservativitets-overvejelser, og at valget derfor til en vis grad er subjektivt. Derimod er strukturelle uoverensstemmelser mellem intuitivt rationelle principper og overvejelser med hensyn til logisk pålidelighed genstand for en objektiv erkendelsesteoretisk analyse. Ikke at pålidelighed altid skal have det sidste ord, men der er strukturelle egenskaber, som kendetegner den enkelte metode, der ikke er subjektive.

De metodologiske forskrifter, jeg tager udgangspunkt i, er konkret fremsat af Peter Gärdenfors i [**Gae88**]. Her bygger Gärdenfors videre dels på eget arbejde, dels på arbejder af Carlos Alchourrón og David Makinson. Forskrifterne, som Gärdenfors når frem til, kaldes i litteraturen ofte for AGM-aksiomerne, med reference til de tre videnskabsfilosoffer. Disse aksiomer kan betragtes som rationelle induktive principper for videnstilegnelse og er primært motiveret af konservativitets-overvejelser. Når en konservativ metodologi skal revidere en teori i lyset af ny data, søger den at bibeholde så meget af den gamle teori som overhovedet muligt. Om aksiomerne er pålidelige, det vil sige, om de med logisk garanti før eller siden vil lede til sandheden, er et mere sekundært emne for Gärdenfors. Pålidelighed er til gengæld helt centralt inden for *formal learning theory* (FLT), der primært kan ses som studiet af induktive metoders evne til at fremsætte sande teorier givet en baggrundsviden. Jeg vil i denne tekst vurdere AGM-aksiomernes pålidelighed ved hjælp af begreber og (meta)metoder fra FLT.

1. Formal learning theory

E. Mark Gold og Hilary Putnam udviklede i 1960erne uafhængigt af hinanden grundstenene til FLT [**Go67**], [**Pu63**]. Gold brugte FLT til at studere børns sprogindlæring, hvorimod Putnam benyttede FLT til at kritisere Carnaps bekræftelsesteori (*confirmation theory*). Siden har man generaliseret begreberne og metoderne, og FLT er i dag generelt studier af induktive problemers løsbarhed.

Ved et induktivt problem forstås der følgende. En metode studeres, og der angives en mængde af mulige verdner, over hvilke metoden kræves at fremsætte korrekte outputs. Hvad der er et korrekt output for metoden, skal også specificeres. Er man en realistisk orienteret videnskabsteoretiker, vil dette typisk være sandhed, hvorimod en antirealistisk orienteret

[1] Som et andet eksempel, se Hendricks' erkendelsesteoretiske undersøgelse af viden som konvergens. [**He99**].

nok vil holde sig til empirisk adækvathed. Pointen med FLT er, at der kan undersøges for alle former for korrekthed. Enhver af de mulige verdner producerer nu en datastrøm, som metoden læser, alt imens den giver et output, der kan variere, mens der læses. Ydermere skal der gives et konvergenskriterie for metoden. Dermed angives det, hvad det vil sige, at metoden stabiliserer til et svar. Dette kan være inden for et endeligt stykke tid, men kan også være i grænsen. En metode løser nu det induktive problem, man har specificeret, netop når den stabiliserer til et korrekt svar over alle de mulige verdner, man lagde ud med at kræve succes over. På denne måde har man givet en præcis definition af begrebet pålidelighed: Metoden er pålidelig, hvis og kun hvis den løser det specificerede induktive problem. En metode kan nu på to måder være 'mere' pålidelig end en anden: Enten løser den flere induktive problemer, eller også løser den det samme problem blot over flere mulige verdner.[2]

Definér nu mængden af opdagelsesmetoder til at være den mindste mængde indeholdende de metoder, der på grundlag af et endeligt initialsegment af en datastrøm (det vil sige, de har set verden fra et tidspunkt til et andet) fremsætter teorier om verden. Metoden (kan) antages at have en baggrundsviden, hvilket svarer til ovennævnte afgrænsning af mulige verdner, der kræves succes i. På denne led er det nu muligt at undersøge 'opdagelsernes logik' (*logic of discovery*), nemlig hvorvidt en opdagelsesmetode – eller et sæt af principper for videnstilegnelse – altid vil fremsætte korrekte teorier direkte fra data og baggrundsviden. Forskellige metoder kan undersøges for pålidelighed og sættes overfor hinanden.[3]

Det interessante er nu en strukturel lighed mellem rekursionsteori og FLT, da man inden for FLT anskuer løsningen af et induktivt problem på nøjagtig samme måde, som man inden for datalogien anskuer den situation, at et computerprogram møder sine specifikationer. Givet et input x fra en given mængde af tilladte inputs producerer programmet altid et korrekt output y. Ligheden kan anskueliggøres på en lidt anden måde: Et problem er rekursivt løsbart, hvis en Turingmaskine for alle input terminerer med et output. Sikkerheden, for at en løsning til et beregnbarhedsproblem nu også er en løsning, er givet i problemets indre matematiske struktur, og at kræve pålidelighed af en induktiv metode er ikke andet end, hvad der inden for rekursionsteorien kræves af en algoritme.

2. Belief revision

Det vigtige spørgsmål inden for belief revision handler om, hvorledes man i lyset af ny viden A opdaterer hele sin mængde af epistemiske overbevisninger (*beliefs)* K. Det centrale begreb her er ikke pålidelighed – som inden for FLT – men derimod minimal forandring og konsistens. Som Gärdenfors skriver det på side 16 i [**Gae88**]: "The central rationality criterion on revisions is that the revision of K should be the minimal change of K that is consistent and includes A", og på side 67 "when changing beliefs in response to new

[2] Dette er en stringent definition af pålidelighed. En anden forståelse af begrebet går på, at en metode er mere pålidelig end en anden, hvis den har succes for flere hypoteser i den *aktuelle* verden. Men her har man ikke taget højde for hypotesernes *kompleksitet*; jævnfør et universelt udsagn overfor 117 instantieringer af dette. Et andet problem, der ofte optræder sammen med kompleksitetsproblemet, er at forveksle pålidelighed med evidens herfor. Men inden for FLT handler det om at definere, hvad pålidelighed er, ikke hvordan vi i den aktuelle verden får indikationer på, at en metode kan være pålidelig. Se også [**Ke96**]:44.

[3] At der findes en 'opdagelsernes logik' er på ingen måde noget, man entydigt har tilsluttet sig op igennem dette århundrede. Reichenbach afviste, at det gav mening at tale om noget sådant. Selvfølgelig er det på ingen måde trivielt at gøre store opdagelser. Men en af pointerne med FLT er at få afklaret og nuanceret spørgsmålet således, at vi ikke overlades til opfattelsen, at alt der kan siges om den sag er, at det er overladt den store kreative intuition at gøre opdagelserne.

evidence, you should continue to belive as many of the old beliefs as possible". Som sådan kan Gärdenfors' metodologi forstås som principper for en rationel induktiv videnstilegnelse.

> As I see it, *inductive reasoning* belongs to the process of adjusting a state of belief to an equilibrium. The active "force" here is the desire to have a representation of the current information that is as parsimonious as possible, and for this purpose inductive generalizations may play an important role. However, the theme of induction is not pursued here. [**Gae88**]:10.

Induktiv resoneren er derimod hjertet af de indlæringsteoretiske undersøgelser. Og belief revision er en induktiv metode, der på baggrund af endelig evidens fremsætter og reviderer teorier om verden. Det er derfor oplagt at gøre det, som Gärdernfors ikke gør: Undersøge belief revision som en induktiv metode med hensyn til pålidelighed.

3. Ideale og beregnbare metoder

Indtil nu har jeg ikke skelnet mellem ideale og beregnbare metoder.[4] En ideal metode er en metode, der kan afgøre en hvilken som helst matematisk relation (og derved beregne alle funktioner). Dette gælder dermed også for alle ikke-konstruktive principper, eksempelvis Skolemfunktioner introduceret ved brug af udvalgsaksiomet. En ideal metode bliver præsenteret for verden på samme måde som alle andre metoder men er ikke kognitivt hæmmet. Det er en beregnbar metode til gengæld. Af matematiske relationer kan en (Turing-)beregnbar metode ikke afgøre mere end de rekursive. Hvad menneskers kognitive evner kan forstås som (eksempelvis ideale overfor rekursive), er en diskussion, der primært hører hjemme i den kognitive psykologi. Men en diskussion af relevansen af metoders beregnbarhed er til gengæld relevant her. En metodes brugbarhed afhænger i høj grad af, hvorvidt den giver anledning til en algoritme, der kan følges skridt for skridt for alle tilladte input. For en metode, der går i stå og kræver guddommelig indsigt, er ikke meget værd.

En af de helt centrale pointer i Kellys bog [**Ke96**] og det vigtigste resultat i Schultes Masters' Thesis [**Sc95**] er, at principper, som er fornuftige set fra et idealt synspunkt, kan være katastrofale for beregnbare metoder. Jeg giver her to eksempler, der begge handler om bekræftelsesmetoder (*assessment methods*). En bekræftelsesmetode er en metode, der – givet en hypotese og et endeligt initialsegment af en datastrøm – vurderer korrektheden af hypotesen. Sig, at en metode er restriktiv, hvis den ikke kan løse et induktivt problem, mens en anden kan.

Lad en 'poppersk' bekræftelsesmetode fungere på den måde, at den givet en hypotese beregner hypotesens forudsigelser og tester dem imod dataen. Hvis der ikke er overensstemmelse forkastes hypotesen, ellers bliver man ved med at samle data ind. Dette svarer godt til, hvad Popper så som *den* videnskabelige metode, hvorfor han krævede, at de fremsatte hypoteser var refutérbare [**Po63**]. Som Kelly og Schulte har vist, er denne metode restriktiv for beregnbare metoder, da der findes en hypotese, hvis forudsigelser ikke er aritmetisk definérbare. Derfor kan den beregnbare popper-metode ikke beregne hypotesens forudsigelser og går i stå i sin vurdering. Ikke desto mindre findes der en beregnbar metode, der ikke opfylder Poppers metodologi, som kan refutere hypotesen inden for en endelig tid. [**Ke96**]:181, [**Sc95**]:23.[5]

[4] Jeg skelner generelt ikke mellem en metode og en videnskabsmand, der benytter metoden.

[5] I notationen, som senere vil blive introduceret, svarer dette til, at der eksisterer en hypotese $h = \{e\}$, for hvilken det gælder $\forall n(e \notin \Sigma_n^A)$, men $\{e\} \in \Pi_1^A$.

Det andet eksempel viser, at konsistens er et metodologisk princip, der også giver anledning til restriktivitet af beregnbare metoder. Sig, at en bekræftelsesmetode er konsistent med data, hvis og kun hvis den forkaster en hypotese, så snart hypotesen er inkonsistent med data. Denne relation mellem hypotese og data er udelukkende formal og kan derfor ikke være restriktiv for ideale metoder. Men den kan være ikke-rekursiv og derfor sætte en beregnbar metode i stå. Kelly viser, at noget stærkere gør sig gældende, nemlig at der findes en hypotese, som kan refuteres beregnbart inden for en endelig tid, men *ingen* aritmetisk definérbar konsistent metode kan refutere eller verificere hypotesen gradvist, hvilket er et meget svagere stabiliseringskriterie. [**Ke96**]:185.

Moralen af disse to eksempler er klar. Drejer det sig om beregnbare metoder, skal man være varsom med sine metodologiske anbefalinger. Popper-metodologien, som megen videnskabsfilosofi har arbejdet med, viser sig at have nogle uheldige konsekvenser, man ikke umiddelbart var opmærksom på. Og et (rationelt) princip så intuitivt indlysende som konsistens med data har konsekvenser, der også kan forhindre metoden i at nå sandheden. I hvert fald når det drejer sig om beregnbare metoder.

Nu er pointen ikke, at man *kun* skal undersøge beregnbare metoder, for "Ideal methodology is not bad in itself. What a demigod can't do, a computer can't do" [**Ke96**]:160. Men i lyset af de to eksempler giver det mening, når Kelly skriver:

> Ideal results are fine so long as they are contrasted with and qualified by their computational counterparts. I depart company, however, when ideal considerations are portrayed as what is essential to methodology while computational considerations are brushed aside as bothersome details of application. [**Ke96**]:160.

4. Problemformulering

Gärdenfors er også interesseret i beregnbare aspekter vedrørende de metodologiske principper. Ikke så meget på grund af overvejelserne, som de er fremført i sidste afsnit, men mere på grund af en mulig computer-implementering i forbindelse med kunstig intelligens (AI).

> The epistemic states are modeled in a logically idealized way, which entails that they are amenable to computer implementations. [**Gae88**]:2-3.

Det er klart, at Gärdenfors mener, at metodologien muliggør en beregnbar implementering, men det undersøges ikke nærmere i bogen. Den første af denne teksts opgaver er således, at finde ud af, hvorvidt dette er muligt. Den anden opgave går ud på at give en pålidelighedsvurdering af Gärdenfors-metodologien. Min *problemformulering* er derfor:

- **(a):** Giver Gärdenforsmetodologien anledning til en beregnbar metode?
- **(b):** Givet metoden er beregnbar er Gärdenforsmetodologien da en restriktiv arkitektur?

5. Litteratur

Tidligere er der skrevet om dette emne, påliceligheden af Gärdenforsmetodologien, i [**He96**] og [**KSH96**]. Jeg vil undervejs referere til disse, når det er relevant.

Jeg har forsøgt at skrive teksten, så kan læses uden de store forudsætninger. Men da det drejer som om en matematisk modellering af et videnskabsfilosofisk problem, er det nok nødvendigt, at læseren med hensyn til filosofien har en vis viden inden for erkendelsesteori og videnskabsfilosofi. På den matematiske side kræves først og fremmest lidt elementær logik og topologi. Men mest af alt forudsætter jeg et kendskab til rekursionsteori op til og med det aritmetiske hierarki. Dette er indtroduceret af Schoenfield på de første 60 sider i [**Sc93**].

Karakterisering af beregnbare metoder

Logikken og matematikken kan være med til at skærpe og præcisere de begreber, der arbejdes med. Brugt inden for videnskabsfilosofien kan der skabes nye resultater. De to eksempler fra indledningen viser dette; principper, man i lang tid har betragtet som fornuftige (i ordets kantianske forstand), har nogle overraskende og nok kontraintuitive konsekvenser, for hvorfor skulle man ikke huske den data, man har set?

Selvfølgelig kan modellen, som resultaterne er skabt inden for, blive for simpel. Jeg vil nu præsentere og diskutere præmisserne for den model, jeg benytter.

1. Generelle antagelser

Nærværende tekst handler om videnkabelige undersøgelser, hvorfor der må gøres nogle elementære antagelser omkring de komponenter, der ingår deri. Videnskabelige undersøgelser handler frem for alt om verden, metoder, videnskabsmænd og -kvinder, korrekthed og om, hvorledes disse ingredienser interagerer med hinanden.

1.1. Superveniens. Det antages, at datastrømmen, som metoden modtager supervenerer asymmetrisk på verden. Det vil sige, at to mulige verdner ikke kan producere to forskellige datastrømme, med mindre de to mulige verdner er forskellige. Pointen med denne antagelser er kort og godt, at gøre observationer meningsfulde. Men med asymmetrisk superveniens udelukkes det ikke, at to forskellige verdner kan producere samme datastrøm. For at udelukke dette, har vi brug for antagelsen om, at der ikke findes global underbestemthed.

1.2. Global underbestemthed. Man har siden antikken været opmærksom på, at hvis sandhed ikke alene afhænger af verdnen, som vi ser den, men mere af en underliggende 'realitet' er der mulighed for en form for underbestemthed af vore hypoteser om verden, som er undergravende for erkendelsen. I dette tilfælde kan to forskellige mulige verdner tildele "falsk" henholdsvis "sand" til en given hypotese til trods for, at de to mulige verdner producerer nøjagtig den samme data til al evig tid. Når dette er tilfældet, siger vi, at hypotesen er *globalt underbestemt.* Når der argumenteres for global underbestemthed, er det som regel i forbindelse med en diskussion af måden, hvorpå vi perciperer verden—for hvad relationen mellem vore sanseorganer og verden *an sich* er, er svært at svare på. Ideen, om at muligheden for global underbestemthed er eksisterende, har ført vidt omkring: Vi er blevet præsenteret for alle mulige scenarier, lige fra Platons og Descartes' illusions- og drømmeargumenter til science fiction setups i stil med hjerner i syrebade og film fra 1999 som *Matrix.* Der kan ikke argumenteres deduktivt for, at der ikke findes global underbestemthed. Men der er to gode måder at komme forbi det klassiske skeptiske problem.

For det første: Rent pragmatisk giver det mening at antage, at der ikke findes global underbestemthed, at erkendelse *er* mulig. For det opleves sådan, og der er evidens for denne oplevelse: Verden fungerer nu engang, blandt andet fordi vi antager, at erkendelse om dette og hint er mulig.

For det andet: En moderne måde at undgå problemet med global underbestemthed er ved simpelt hen at antage, at der ikke kræves succes over mængden af alle mulige mere eller mindre mærkelige verdner men kun over en passende delmængde. Eksempelvis kræver man kun pålidelighed i verdner, som er tættest på den aktuelle verden.[1] Under denne antagelse undersøger man, hvor langt metoden kan nå. Man kan antage baggrundsviden på andre måder, men essensen er, som Glymour skriver det på side 249 i [**Gl92**], at det er et af de mest alvorlige nutidige svar til sketicisten.

På baggrund af disse to argumenter antager jeg, at ingen af de hypoteser, der i det følgende vil blive undersøgt, er globalt underbestemte.

1.3. Verden givet diskret. Verden viser sig i kraft af en datastrøm. Denne strøm er en række observationer givet diskret. Dette svarer måske ikke helt til vores umiddelbare oplevelse af verden som givet ved et flydende kontinuum. Men for en metode er det en rimelig simplificering, da metoden altid vil tage forskellige observationer til forskellige tider i betragtning. Samtidig viser moderne psykologiske undersøgelser, at vi ikke kan andet end opfatte verden diskret. Når virkeligheden er givet som et kontinuum, og vi kun har det at henholde os til, kan vi ikke skelne enkelte hændelser fra hinanden. Samtidig ved vi, at beskrivelser af verden altid bliver diskrete: Til tiden t_0 blev x_0 observeret, til tiden t_1 blev x_1 observeret og så videre.

1.4. Passiv observans. Slutteligt antager jeg, at datastrømmen på ingen måde afhænger af metodens opførsel, med andre ord afspejles videnskabsmandens handlinger ikke i datastrømmen. Dette er selvfølgelig en kraftig antagelse, der ikke passer helt med eksempelvis opdagelser inden for kvantemekanikken. Her ved man, at hvorvidt et system grundlæggende har (eller ser ud for os til at have) en bølge-karakter eller en partikel-karakter afhænger af den måde, vi har valgt at måle systemet på. Kelly viser i [**Ke96**], at den indlæringsteoretiske model godt kan relativeres til at tage højde for eksempelvis teori-ladethed, men jeg har valgt at holde mig til det mere simple tilfælde her.

2. Setup

Definition 2.1. En datastrøm er en total funktion $e : \mathbb{N} \to \mathbb{N}$.
Domænet til en datastrøm (DS) er diskrete tidsenheder, og værdimængden er talkoder for observationerne til disse tider. En DS antages at være total, således at der til enhver tid er givet en observation. En DS kan ækvivalent defineres som en uendelig følge af kodede diskrete observationer.

Notation 2.2.

1. *e, d, skrives for en DS. Universet er givet ved*
$$U := \{e \mid e \text{ er en DS}\}.$$
2. *$e(n)$ er observationen til $t = n$ på e.*
3. *$e|n := (e(0), \dots, e(n-1))$ skrives for det endelige initialsegment af e. Skrives også med græske bogstaver ε, τ.*
$$U^* := \{\varepsilon \mid \varepsilon \text{ er et endeligt initial segment af en DS } e\}.$$
4. *$lh(e|n) = n$, skrives for længden af den endelige DS, $e|n$, der har længden n.*
5. *$n|e := (e(n), e(n+1), \dots)$, halen af e fra $t = n$.*

[1] Nozick benytter sig af denne strategi i [**No81**].

6. *Lad $K \subseteq U$, da kaldes K for* ***baggrundsviden*** *og*
$$K^* := \{\varepsilon \in U^* \mid \varepsilon \subset e \in K\}$$
er mængden af alle endelige initialsegmenter inden for bagrundsviden.
7. *$*$ skrives for en generel operator der kæder to udtryk sammen. Eksempelvis er $e = e|n * n|e$.*
8. *Lad U og en endelig DS ε være givet. Sæt da*
$$[\varepsilon] := \{d \in U \mid \varepsilon \subset d\}.$$
$[\varepsilon]$ er altså mængden af de DS, der er ekstensioner af ε. $[\varepsilon]$ kaldes en ***tulipan*** *og ε kaldes* ***stilken****.*

2.1. Hypoteser, empiriske spørgsmål og opdagelsesmetoder. Det antages indtil videre, at sandhed er en uspecificeret relation C mellem hypoteser og datastrømme og kaldes efterfølgende – sålænge den er uspecificeret – for korrekthed, da det er en vigtig pointe, at det ikke behøver at dreje sig om sandhed, men at det ligeså godt kan være empirisk adækvathed eller en anden form for korrekthed, der er tale om. C er i det endnu uspecificerede tilfælde givet ved: $(e, h) \in C$ hviss h er korrekt for den data, der udgøres af e. Herved kan hypoteser identificeres med den mængde af DS, som netop gør hypoteserne korrekte:

Definition 2.3. En hypotese h er bestemt ved følgende:
$$h := \{e \mid h \text{ er korrekt i } e\}.$$

Ind imellem kan det dog have et formål at skelne imellem en hypotese A formuleret eksempelvis som en lukket formel i et formelt sprog og mængden af de DS, $\langle A \rangle$, som gør A korrekt. Teorier udgøres af hypoteser. Derved kan teorier også identificeres med en mængde af DS. Som eksempel kan man tage teorien T, som hævder at enten er h_1 korrekt, eller også er h_2 og h_3 korrekte. Denne teori er identisk med $h_1 \cup (h_2 \cap h_3)$. Eller hvis hypoteserne A, B og C er udtrykt i et formelt sprog og $T = A \vee (B \wedge C)$, så er $\langle T \rangle = \langle A \rangle \cup (\langle B \rangle \cap \langle C \rangle)$. På denne måde anskueliggøres det, hvorledes de logiske konnektiver $\neg$, $\vee$, $\wedge$, $\rightarrow$ kommer til udtryk ved de mængdeteoretiske operationer kompliment, $\cup$, $\cap$, $\subseteq$, henholdsvis. Men medmindre der angives en eksplicit skelnen mellem en hypotese A og mængden af DS $\langle A \rangle$, for hvilke A er korrekt, skelner jeg i det følgende ikke mellem en hypotese og mængden af DS, der gør hypotesen korrekt. Derved bliver hypoteser og teorier delmængder af U.

Et eksempel på en hypotese er $h = \{e\}$. I dette tilfælde siges h at være *empirisk fuldstændig*, da der netop er én DS, som gør h korrekt. Et eksempel på en empirisk fuldstændig hypotese er, at vi kun vil observere 1-taller:
$$h = \{e \mid \forall n(e(n) = 1)\}.$$
Videnskabelige undersøgelser har to kognitive mål: At finde sandheden og undgå fejl. Lad Θ være et par af to elementer: Dels en mængde af hypoteser, dels en partition af disse hypoteser "op i enkelte korrekte celler". Hypotesemængden forstås som en mængde udsagn om verden, vi ønsker at undersøge og fremsætte korrekte teorier udfra. Θ er nu, hvad vi vil kalde for et *empirisk spørgsmål*, og de enkelte celler i partitionen er potentielle svar til spørgsmålet.

Definition 2.4. $\Theta = (H, C)$ er **empirisk spørgsmål**, hvis Θ består af:

1. En mængde $H = \{h \mid h \text{ er en hypotese}\}$.

2. En korrekthedsrelation C mellem H og U.

Hvor H er lukket under komplement.

Definition 2.5. $\Theta(e)$ er det korrekte svar til det empiriske spørgsmål i datastrømmen e.

Θ og $\Theta(e)$ skal specificeres nærmere før bestemte paradigmer træder frem. Som eksempel kan vi tage, hvad man kunne kalde for det sværeste empiriske spørgsmål, nemlig at skulle identificere korrekte fuldstændige hypoteser. Som beskrevet i forrige afsnit findes der til empiriske fuldstændige hypoteser en og kun en DS, som gør hypotesen korrekt, det vil sige $(e, h) \in C$ hviss $h = \{e\}$. Således er

$$\Theta_0(e) = \{\{e\}\}.$$

Som et andet eksempel er der, hvad man i [**KSH96**] kalder for *den fuldstændige H-teori til e*. Lad en mængde af hypoteser H være i givet. I første omgang kræves der ikke andet, end at H er tællelig og – som ovenfor – lukket under komplement. Tænk på H som en mængde af udsagn, vi ønsker at undersøge med hensyn til korrekthed. En H-teori er givet ved et snit mellem en række hypoteser fra H, det vil sige som konjuktionen af en række udsagn. Givet $e \in U$ er den fuldstændige H-teori til e givet ved:

$$\Theta_f(e) = \bigcap\{h \in H \mid e \in h\}.$$

Som nævnt i indledningen læser en opdagelsesmetode et endeligt stykke af en DS og fremsætter på grundlag heraf en hypotese (eller en teori) om verden. Som motivation for den kommende formelle definition defineres:

$$Eq(n, e(n)) = \{e' \mid e'(n) = e(n)\}.$$

Epistemologisk er $Eq(n, e(n))$ netop, hvad der til $t = n$ opdateres på. Altså det udsagn der siger, at til det enkelte tidspunkt n observerede man lige nøjagtig $e(n)$. Definér:

$$Up(e|n) = (Eq(0, e(0)), \ldots, Eq(n-1, e(n-1))).$$

$Up(e|n)$ er med andre ord den følge af opdateringer, man har gjort sig indtil $t = n$. Da $[e|n] = \cap Up(e|n)$ er det rimeligt at forlange, at en opdagelsesmetode efter at have læst det endelige initialsegment $e|n$ af e fremsætter sit gæt på grundlag af $[e|n]$. Men da tulipaner er entydigt bestemt af deres stilke, giver det mening med følgende definition, hvor $\mathcal{P}(U)$ er potensmængden af U:

Definition 2.6. En opdagelsesmetode er en funktion

$$f : U^* \to \mathcal{P}(U)$$

2.2. Identifikation af empiriske spørgsmål. Givet et empirisk spørgsmål Θ kunne man håbe, at den anvendte metoden på et tidspunkt med garanti standser i sine undersøgelser og giver et korrekt svar. Denne form for opdagelser er, hvad man kalder pålidelig opdagelse *i endelig tid*. Jeg vælger i denne tekst kun at arbejde med et svagere stabiliseringskriterie, nemlig hvad Gold kalder *identification in the limit*, det vil sige stabilisering i grænsen.[**Go67**] Motivationen for dette succes-kriterie er to-sidet. *For det første* er kriteriet understøttet af en peirceiansk opfattelse af videnskab og sandhed: Sandheden om verden er det, som videnskaberne i grænsen hævder om verden. Det kan dog være tilfældet, at videnskaberne har fundet sandheden uden at vise tegn herpå. *For det andet* viser det sig, at problemer, som hidtil har været anskuet som uløselige, bliver principielt løselige i grænsen.

Vi siger, at f stabiliserer i grænsen til et korrekt svar til Θ på e hviss f identificerer Θ på e hviss $\exists n \forall m \geq n : f(e|m) \subseteq \Theta(e)$. Uformelt kan man sige, at Θ kan identificeres, netop når det gælder, at vi for alle DS altid finder en korrekt hypotese til Θ. Det forudsættes ikke, at f efter at have nået sit modulus, det vil sige det n sådan at for alle $m \geq n : f(e|m) \subseteq \Theta(e)$, holder sig til en bestemt hypotese; med andre ord at $f(e|m) = f(e|m')$ for alle $m, m' \geq n$. Det er tilstrækkeligt, at f kun fremsætter korrekte hypoteser.

Antagelse af baggrundsviden kan, som tidligere nævnt, være en måde at løse problemer, som ellers ikke kunne løses. Baggrundsviden kan nu forstås, som den mængde af mulige verdner for hvilke metoden er garanteret succes, det vil sige pålidelig. Men man kan også tænke på baggrundsviden, som den epistemiske tilstand agenten (eller metoden) starter med.[2]

Definition 2.7. Identifikation under antagelse af baggrundsviden K:

$$f \text{ identificerer } \Theta \text{ givet } K \equiv \forall e \in K, \exists n \forall m \geq n : f(e|m) \subseteq \Theta(e).$$

3. Topologi og Baire space

Den opmærksomme matematiske læser vil måske have bemærket, at ovenstående model er bygget over det topologiske rum Baire space. Senere i denne tekst bliver forskellige induktive metoder karakteriseret i præcise topologiske termer, hvorfor disse termer nu introduceres.

Definition 3.1. Et topologisk rum er en mængde X af punkter sammen med en topologisk struktur, hvilken er bestemt af en klasse $\mathcal{T}$ af delmængder af X, kaldet for åbne mængder af X, som opfylder følgende:

1. $\emptyset, X \in \mathcal{T}$.
2. Hvis $A, B \in \mathcal{T}$ så $A \cap B \in \mathcal{T}$.
3. For alle $\mathcal{E} \subseteq \mathcal{T}$ gælder der $\cup\mathcal{E} \in \mathcal{T}$.

Lukkede mængder er komplementet af åbne mængder relativt til X, det vil sige $X \setminus G$, med G åben. Mængder, der både er åbne og lukkede, kaldes clopen.

3.1. Baire space. I notationen introduceret tidligere bliver Baire space til parret $(U, \mathcal{B})$, hvor de åbne mængder i $\mathcal{B}$ er defineret ved:

$$P \in \mathcal{B} \text{ hviss } \exists \mathcal{G} \subseteq U^* : P = \cup\{[\varepsilon] \mid \varepsilon \in \mathcal{G}\}.$$

At dette er et topologisk rum indses ved følgende. Lad o være den tomme følge. Da er $U = [o]$, og $\emptyset = \cup\emptyset$, så U og $\emptyset$ er elementer i $\mathcal{B}$, det vil sige åbne mængder. Det gælder for et hvilket som helst par af tulipaner $[\varepsilon]$, $[\varepsilon']$, at enten er den ene indeholdt i den anden eller også er de disjunkte, hvilket med andre ord vil sige $[\varepsilon] \subseteq [\varepsilon']$ hviss $\varepsilon' \subseteq \varepsilon$. Derfor gælder der, at $[\varepsilon] \cap [\varepsilon']$ enten er en tulipan eller den tomme mængde. Heraf følger der, at en endelig fællesmængde af to mængder, der hver især er en forening af tulipaner, enten er en foreningsmængde af tulipaner eller den tomme mængde. I begge tilfælde en åben mængde.

[2]Den sidste forståelse af baggrundsviden bliver naturlig, når jeg senere (side 17) formelt definerer Gärdenfors-metodologien.

Slutteligt er en vilkårlig foreningsmængde af foreningsmængder af tulipaner pr. definition en åben mængde.

U^* er en tællelig mængde, da den kan ordnes leksikografisk.[3] Da en tulipan $[\varepsilon]$ er bestemt alene ved ε følger der, at der kun findes et tælleligt antal tulipaner, hvorfor enhver åben mængde højst vil udgøres af et tælleligt antal tulipaner. Der gives nu en meget naturlig interpretation af en åben mængde: En åben mængde P svarer til en hypotese, der er verificérbar i endelig tid. En datastrøm observeres, og 'genkendes' en stilk ε til en tulipan fra P, ved man, at P er korrekt, da en hvilken som helst ekstension af ε er en datastrøm i P. De lukkede mængder svarer omvendt til hypoteser, der er refutérbare i endelig tid, og mængder som er clopen svarer til hypoteser, der er afgørbare i endelig tid. Det viser sig, at genereres der et hierarki ud fra tællelige foreningsmængder og komplementer af mængder, som er clopen, får man en naturlig katalogisering af forskellige induktive problemers kompleksitet. Denne katalogisering er tæt knyttet til de succes- og stabiliserings-kriterier, der anvendes, og man ser med hierarkiet, at et induktionsproblem ikke bare er et induktionsproblem, men at induktionsproblemer kan måles i grader af kompleksitet. Se eventuelt [**Ke96**]. Den matematiske struktur, som beskriver dette, er det endelige Borelhierarki.

Lad $\overline{P}$ være komplementet til P. Da er *Borelhierarkiet* induktivt defineret på følgende vis:

Definition 3.2.

$$\begin{array}{rcl} P \in \Sigma_0^B & \equiv & P \text{ er clopen.} \\ P \in \Sigma_{n+1}^B & \equiv & \text{der eksisterer tællelig } \Gamma \subseteq \Sigma_n^B : P = \cup\{\overline{R} \mid R \in \Gamma\}. \\ P \in \Pi_n^B & \equiv & \overline{P} \in \Sigma_n^B. \\ P \in \Delta_n^B & \equiv & P \in \Sigma_n^B \cap \Pi_n^B. \end{array}$$

Bemærk at $\Delta_0^B = \Delta_1^B = \Sigma_0^B = \Pi_0^B$, at de åbne mængder er Σ_1^B mængder, og de lukkede mængder er Π_1^B mængder.

3.2. Begrænset Baire space. At begrænse topologien svarer til at antage baggrundsviden. Lad $(U, \mathcal{B})$ være givet som i sidste afsnit. Restriktionen af $(U, \mathcal{B})$ til $K \subseteq U$ er da givet ved følgende: Erstat U med K og erstat $\mathcal{B}$ med mængden $\mathcal{B}|K$, der består af alle fællesmængder $A \cap K$ hvor $A \in \mathcal{B}$. Det begrænsede rum er nu $(K, \mathcal{B}|K)$. Det følger, af grundlæggende samme argument som ovenfor, at dette rum er topologisk. Elementer fra $\mathcal{B}|K$ kaldes K–åbne, mængder af formen $K \setminus G$, hvor G er K–åben, er K–lukkede, og mængder, der er begge dele, er K–clopen. Et endeligt Borelhierarki med en restriktion af U til K er defineret som ovenfor, hvor bunden er mængder, som er K–clopen. Hvis P er K–clopen, ligger P i $\Sigma[K]_0^B$, og notationen for resten af hierarkiet, herunder K–åbne og K–lukkede mængder, er tilsvarende.

3.3. Det effektive endelige Borelhierarki. Det endelige Borelhierarki beskriver almindeligvis induktive problemers kompleksitet i tilfældet, hvor metoden er ideal. I denne tekst er jeg dog mere interesseret i det tilfælde, hvor den induktive metode er beregnbar, og dertil benyttes, hvad man til tider kalder for det effektive endelige Borelhierarki. Men indledningsvist er der forhold omkring forståelsen af hypoteser, som skal specificeres.

[3]Metoden er følgende: Først numererer man en følge af længde 1, derefter endnu en følge af længde 1 og en af længde 2, derefter en af længde 1 en af længde 2 en af længde 3, og så fremdeles. Til sidst er man nået igennem alle endelige følger.

Korrekthedsrelationen C er en relation mellem en datastrøm e og en mængde af datastrømme h. I det generelle tilfælde forstås en hypotese h altid som en mængde af datastrømme. Men i denne model, givet at vi har formaliseret opdagelsesmetoder som funktioner, kan vi faktisk skabe en slags kodning af de hypoteser, vi vælger at undersøge, ind i $\mathbb{N}$. Som sagt identificerer vi normalt en hypotese med den mængde af datastrømme, som gør hypotesen sand. Dette muliggør ikke en kodning ind i $\mathbb{N}$, men kun at en hypotese kan kodes som en enkelt datastrøm, da en enkelt datastrøm koder en mængde af datastrømme (metode i stil med metoden for nummereringen af U^*). På denne måde kan man eksempelvis have en enkelt datastrøm som indeks for en borelmængde. I det generelle tilfælde har vi derfor, at en opdagelsesmetode er en funktion:

$$f : U^* \rightarrow U.$$

Men da opdagelsesmetoden skal opfylde funktionsegenskaberne, og da U^* er tællelig, må værdimængden til f også være tællelig. Derfor giver det i denne sammenhæng kun mening at arbejde med et tælleligt antal af hypoteser. Kald værdimængden for H. Da H er tællelig, kan vi (idealt set) nummerere den, hvorfor enhver hypotese i H indekseres. Derved bliver indekset til $h \in H$ en slags kode for h. Og vi har, at givet en indeksering af H kan vi arbejde med H, som om den var en mængde af indekser, og opdagelser kommer derved til at handle om identifikation af *indekser* for hypoteser. Hermed menes der ikke, at den enkelte hypotese er tællelig, for en hvilken som helst åben mængde i Baire space er overtællelig. Pointen er blot, at antallet af hypoteser, som vi vælger at undersøge, må være tælleligt.[4]

Et aritmetisk hierarki af mængder af relationer skal nu konstrueres på en sådan måde, at det dels er en naturlig generalisering af det aritmetiske hierarki, som man kender det fra rekursionsteorien, dels så det kan bruges til at karakterisere beregnbare induktive metoder. Ideen er, at vi skal have en Turingmaskine til at acceptere en uendelig datastrøm som input og ikke kun endelige mængder af naturlige tal. Dette kan lade sig gøre ved at give maskine et ekstra input-bånd, som den læser på, alt imens der gives outputs. At det er muligt skyldes, at der netop ikke kræves læsning af en hel datastrøm, førend der produceres hypoteser, men kun endelige initialsegmenter.

Korrekthedsrelation $C(e, h)$ er, jævnfør forrige afsnit, en relation mellem to typer objekter: kodetal for hypoteser (type 1) og datastrømme (type 2). Generelt set kan en relation R have et hvilket som helst antal argumenter af type 1 og type 2. Sig, at R er af type (k, m), netop når R har k argumenter af type 2 og m argumenter af type 1. Vi skal nu have defineret aritmetisk kompleksitet for en vilkårlig relation af type (k, m). Det viser sig, at det sædvanlige aritmetiske hierarki er et specialtilfælde af det generaliserede, nemlig hvor relationerne har type $(0, m)$.

En relation R af type (k, m) er rekursiv, hvis der findes en Turingmaskine, der givet de m endelige argumenter og et bånd for hvert af de k uendelige argumenter før eller senere stopper med 1, hvis R er sand og ellers stopper med 0. Lad $\overrightarrow{e}$ betegne en k-vektor af datastrømme og $\overrightarrow{x}$ en m-vektor af naturlige tal.

[4]Bemærk, at dette er en nødvendig *forudsætning* for at få modellen til at fungere. Men det er en rimelig antagelse, da jeg i denne tekst vælger at arbejde med beregnbare metoder, der ville få svært ved at 'håndtere' hver og en hypotese fra en mængde af hypoteser, som er overtællelig.

Definition 3.3. Det generaliserede aritmetiske hierarki:

$$\begin{aligned} R \in \Sigma_0^A &\equiv R \text{ er rekursiv.} \\ R \in \Sigma_{n+1}^A &\equiv \exists Q \in \Sigma_n^A \text{ af type } (k, m+1) \text{ sådan at} \\ & \quad \forall \overrightarrow{e} \in U^k, \overrightarrow{x} \in \mathbb{N}^m : R(\overrightarrow{e}, \overrightarrow{x}) \leftrightarrow \\ & \quad \exists x_{m+1} : \neg Q(\overrightarrow{e}, \overrightarrow{x}, x_{m+1}) \end{aligned}$$

Π_n^A og Δ_n^A er defineret, på sædvanlig vis, på basis af Σ_0^A.

En karakterisering af beregnbare metoder gives i termer af kompleksiteten af korrekthedsrelationen C. Men også kompleksiteten af metodens baggrundsviden K og hypotesemængden H er af betydning. Derfor må vi relativere aritmetisk kompleksitet til begge disse mængder:

Definition 3.4.

$$\begin{aligned} R \in \Sigma[K, H]_n^A &\equiv \exists Q \in \Sigma_n^A \text{ sådan at} \\ & \quad \forall \overrightarrow{e} \in K, \overrightarrow{x} \in H^m : R(\overrightarrow{e}, \overrightarrow{x}) \leftrightarrow Q(\overrightarrow{e}, \overrightarrow{x}). \end{aligned}$$

4. Karakterisering af beregnbare metoder

Der kan nu gives en karakterisering af de beregnbare opdagelsesmetoder, det vil sige, at der gives en bestemmelse af omfanget og kompleksiteten af de problemer som er *pålideligt* løselige. Kelly giver karakteriseringen for et identifikationskriterie, som er lidt stærkere end det, som jeg har valgt at benytte. Definér derfor identifikation$_K$ i for den kommende sætning således:

$$f \text{ identificerer}_K\ \Theta \text{ givet } K \equiv \forall e \in K, \exists n \forall m \geq n : f(e|m) = \Theta(e).$$

Sætning 4.1. *(Kelly). Θ er identificérbar$_K$ givet K ved en beregnbar opdagelsesmetode hviss $\forall e \in K(\Theta(e) \neq \emptyset)$, $H \in \Sigma_2^A$ og $C \in \Sigma[K, H]_2^A$.*

Bevis. Se [**Ke96**]:253. ⊣

Som nævnt tidligere er det et relativt strengt krav til metoden, at den skal identificere den fulde hypotese. Jeg vælger derfor i det følgende at arbejde med det svagere kriterie for identifikation, det vil sige med mængdeinklusionen, hvor der kun kræves fremsættelse af korrekte hypoteser.

En beregnbar Gärdenfors-metode

Gärdenfors er selv interesseret i en mulig computer-implementering af sin metodologi. Med andre ord, om den giver anledning til en beregnbar metode.

> I formulate rationality postulates for revisions and other types of epistemic changes. I also present some explicit constructions of ideally rational changes [...] Although I do not develop computer implementations of the constructions, I believe that my theory is of interest to AI (artificial intelligence) researchers. After all, the problem of finding an appropriate knowledge representation is a key problem for AI. But a solution to this problem is of little help unless one also understands how to update the epistemic states in the light of new information. [**Gae88**]:ix.

Gärdenfors formulerer sine principper for rationel videns-opdatering i et relativt simpelt udsagnslogisk setup, men er vi interesseret i det beregnbare aspekt ved opdateringen, er det naturligt at se på, hvilken metode principperne giver anledning til, når de formuleres rekursionsteoretisk. Et af mine arbejdsspørgsmål undervejs har derfor været: Hvordan ser en beregnbar Gärdenfors-opdatering ud?

Metodologien har i det ideale tilfælde været undersøgt inden for FLT. Kelly, Schulte og Hendricks viste i [**KSH96**], at givet 'den rigtige' baggrundsviden er metodologien fuldstændig, det vil sige, at ethvert empirisk spørgsmål, som kan identificeres idealt, kan identificeres (idealt) af en Gärdenfors-metode. Det er ikke nogen simpel sag at formulere Gärdenfors-metodologien indlæringsteoretisk; et problem, som jeg senere vil komme ind på. I første omgang går jeg ud fra den formulering Kelly, Schulte og Hendricks har givet.

1. Gärdenfors-metodologi repræsenteret ved AGM-aksiomer

Lad T være en teori og P en empirisk proposition, det vil sige, korrektheden af P afhænger kun af datastrømmen. En belief revision operator $+$ tager en teori T og en empirisk proposition P som input og har altid en teori T' som output. Sagt på en anden måde:

$$T + P = T'.$$

Som nævnt i indledningen bygger Gärdenfors blandt andet videre på arbejde af Makinson og Alchourrón, hvorfor de metodologisk principper kaldes for AGM-aksiomerne. I et indlæringsteoretisk setup er de i [**KSH96**] givet ved følgende fire askiomer:

(AGM 1) $T + P \subseteq P$.
(AGM 2) Hvis $P \neq \emptyset$ så $T + P \neq \emptyset$.
(AGM 3) Hvis $P \cap T \neq \emptyset$ så $T + P = P \cap T$.
(AGM 4) Hvis $P \cap (T + Q) \neq \emptyset$ så $P \cap (T + Q) = T + (P \cap Q)$.

Disse principper kan repræsenteres i et såkaldt Grove system [**Gr88**]. Et Grove system er en samling $\Gamma^Q \subseteq \mathcal{P}(U)$ centreret omkring en proposition Q, som opfylder:

1. $\subseteq$ definerer en fuldstændig ordning af Γ^Q.
2. For alle konsistente propositioner P eksisterer der et unikt $\subseteq$-minimalt element $\gamma_P \in \Gamma^Q$, sådan at $\gamma_P \cap P \neq \emptyset$.
3. $U \in \Gamma^Q$.

4. Q er det mindste element i Γ^Q.

Grove har vist, at for ethvert par bestående af en AGM-operator $+$ og en proposition K eksisterer der et Grove system Γ^K sådan at, for alle propositioner P gælder der $\gamma_P \cap P = K + P$. Omvendt gælder der: Lad Γ^K være et Grove system for en proposition K. Da eksisterer der en unik AGM-operator $+$ sådan at, for alle P gælder der $\gamma_P \cap P = K + P$.

En AGM-operator bestemmer nu entydigt en iterativ AGM-opdagelsesmetode f efter følgende induktive metode:

1. $f(\emptyset) = K$.
2. $f(\varepsilon * \tau) = f(\varepsilon) + [\varepsilon * \tau]$.

Det bliver nu vist i [**KSH96**], at for 'passende'[1] definitioner af *data retention, timidity* og *stubbornness* er de to udsagn "f er data rententive, stubborn og timid" og "f er en iterativ AGM-opdagelsesmetode" ækvivalente. Dette er af stor praktisk nytte, når AGM-aksiomerne skal undersøges.

De uformelle defintioner af de tre egenskaber er følgende: f er *data retentive,* hvis det altid er tilfældet, at de hypoteser, det vil sige, mængden af epistemiske overbevisninger (*belief set*) der fremsættes af f implicerer den data, f har set. f er *timid,* netop når det er tilfældet, at hvis dataen er konsistent med den tidligere hypotesefremsættelse, så implicerer den tidligere hypotese den nye hypotese. Slutteligt er f *stubborn*, når f's hypotesefremsættelse implicerer den forrige, med mindre denne er refuteret. De formelle definitioner, der er generaliserede med hensyn til de to sidste, ser således ud:

Definition 1.1. f er *data retentive* $\equiv$ for alle $e|n \in U^* : f(e|n) \subseteq [e|n]$.
Hermed implicerer data retention konsistens i al almindelighed; herunder lokal konsistens, hvilket vil sige, at fremsættelsen skal være konsistent med den sidste observation. I denne definition huskes al data simpelthen inden for bagrundsviden, hvad man andre steder kalder for *perfect memory* [**Ke98**].[2]

Definition 1.2. f er *timid* $\equiv$ for alle $e|n+m \in U^*$ sådan at $[e|n+m] \cap f(e|n) \neq \emptyset$ medfører $f(e|n) \cap [e|n+m] \subseteq f(e|n+m)$.

Definition 1.3. f er *stubborn* $\equiv$ for alle $e|n+m \in U^*$ sådan at $[e|n+m] \cap f(e|n) \neq \emptyset$ medfører $f(e|n) \cap [e|n+m] \supseteq f(e|n+m)$.
Tilføjer man som krav til f, at den ikke må give trivielle gæt, det vil sige, hvis ε ikke er den tomme følge er $f(\varepsilon) \neq \emptyset$, ser man overensstemmelsen med AGM-aksiomerne med det samme. Data retention er (AGM 1), og stubbornness og timidity tilsammen giver (AGM 3) og (AGM 4). Det sidste aksiom på grund af associativitet og kommutativitet af $\cap$, da $[e|n+m] = ([e|n+1] \cap \ldots \cap [e|n+m])$.[3] Forstår vi nu – i overensstemmelse den tidligere diskussionen af baggrundsviden (side 10) – baggrundsviden, som den mængde af epistemiske overbevisninger metoden begynder med, kan vi definere AGM-metoderne.

---

[1] Jeg skriver 'passende', da min definition af *data retention* inkluderer de to definitioner i [**KSH96**] af *data retention* og *consistency.*

[2] En anden definition kunne lyde: f er data retentive$_2$ inden for baggrundsviden $\equiv \forall n < m, \forall e \in K : f(e|n) \cap Eq(n-1, e(n-1)) = \emptyset \rightarrow f(e|m) \cap f(e|n) = \emptyset$. Data retention$_2$ implicerer her ikke lokal konsistens, der er defineret ved $\forall n \forall e \in K : f(e|n) \cap Eq(n-1, e(n-1)) \neq \emptyset$. Data retention$_2$ betyder, at hvis en hypotese en gang er blevet refuteret af data, så vil denne hypotese aldrig igen dukke op i senere hypotesefremsættelser. På denne led kan man nuancere beskrivelsen af en AGM-metode. Jeg har dog valgt at holde mig til definitionen af data retention, som den er givet i brødteksten.

[3] Et fuldstændigt argument går dog via Grove systemer, se appendiks til [**KSH96**].

Definition 1.4. En *AGM-metode* er bestemt som værende følgende induktive metode:

1. $f(\emptyset) = K$.
2. f opdaterer i overensstemmelse med *data rentention, timidity, stubbornness* og giver ikke-trivielle gæt.

I hvert fald ét forhold ved en AGM-metodologien er umiddelbart værd at bemærke, nemlig at den er meget konservativ, og at den derfor er følsom over for den initiale mængde af overbevisninger. Det er stubbornness og timidity, der giver anledning til konservativiteten. Følgende eksempel, som er hentet fra [**KSH96**], viser dette. Lad os betragte en videnskabsmand, som er stubborn, og som til at begynde med er overbevist om, at når man slår plat eller krone med en mønt, vil mønten med tiden mindst en gang vende kronen opad. Lad os yderlige antage, at dette aldrig sker. Hvis videnskabsmanden kun er stubborn, kan han med tiden snævre sin hypotese ind på en sådan måde, at den oprindelige hypotese stadigvæk er impliceret. Han kan for eksempel efter 53 kast tilføje hypotesen om, at inden de første hundrede kast har mønten vendt kronen op. Men er videnskabsmanden også timid, er dette ikke muligt. Han vil være fastlåst til evighed på en forkert hypotese. Men det er ikke noget problem at finde ud af sandheden omkring hypotesen, der kræves blot én revision. Man hævder hypotesens negation – at kronen aldrig kommer op – og *hvis* kronen kommer op, hævder man hypotesen i al evighed. Eksemplet viser, hvor følsom AGM-metoden – eller rettere en stubborn og timid videnskabsmand – er overfor den initiale mængde af overbevisninger. Men det illustrerer også, hvorledes timidity og stubbornness virkelig tvinger metoden til at gøre minimale forandringer—i Gärdenfors' ånd.

2. En beregnbar Gärdenfors

Jeg angiver nu, hvorledes en beregnbar AGM-opdatering foregår. Bemærk først, at givet data retention indskrænkes stubbornness og timidity inden for baggrundsviden til:

$$\forall e \in K, \forall n, m \in \mathbb{N} : [e|n+m] \cap f(e|n) \neq \emptyset \rightarrow f(e|n+m) = f(e|n) \cap [e|n+m].$$

2.1. AGM-rekursionen.

2.1. AGM-rekursionen. Førend det er muligt rekursivt at angive, hvorledes AGM-opdateringen foregår, er det nødvendigt at antage, dels at for den initiale mængde af overbevisninger F_0 er $e \in F_0$ rekursivt afgørbart (oplagt at tage en tulipan eller foreningsmængden af et endeligt antal tulipaner), dels at

$$\forall n \in \mathbb{N}, \forall e \in U : f(e|n) \cap [e|n+1] \neq \emptyset. \tag{1}$$

Da foregår en opdatering i overensstemmelse med Gärdenfors-metodologien således:

1. $f(\emptyset) = F_0$.
2. $f(e|n) = \underbrace{F_0 \cap [e|n]}_{(*)}$.

$(*)$ er i høj grad en beregnbar procedure, da det gælder for fællesmængden, at det er afgørbart om en datastrøm d er element i den. Hvis F_0 er en endelig mængde af tulipaner er proceduren sågar primitiv rekursiv. Sættes $F_0 = [e|0]$, får vi at $f(e|n) = [e|n]$. Så med tiden snævrer metoden længere og længere ind omkring hypotesen $\{e\}$, da $\cap[e|n] = \{e\}$.

2.2. Karakterisering af beregnbar AGM-metode. Det interessante er nu at se, hvad der bliver identificeret.

Sætning 2.1. *Antag (1) ovenfor, samt at initialmængden er således, at* $\forall e \in F_0, \exists n : [e|n] \subseteq F_0$*, samt at* F_0 *er rekursiv afgørbar. Da gælder der, at hvis beregnbar AGM-metode* f *identificerer* Θ *givet* F_0 *så er de enkelte celler til* Θ F_0*-åbne i Borelhierarkiet.*
Bevis. Genkald definitionen af identifikation:

$$f \text{ identificerer } \Theta \text{ givet } F_0 \equiv \forall e \in F_0, \exists n \forall m \geq n : f(e|m) \subseteq \Theta(e)$$

Heraf følger der for beregnbar AGM-metode:

$$\forall e \in F_0, \exists n \forall m \geq n : [e|m] \cap F_0 \subseteq \Theta(e)$$

hvilket på grund af antagelsen $\forall e \in F_0, \exists n : [e|n] \subseteq F_0$, samt at F_0 er rekursiv afgørbar medfører

$$\forall e \in F_0, \exists n' \forall m \geq n' : [e|m] \subseteq \Theta(e) \text{ hviss}$$

$$\forall e \in F_0, \exists n' : [e|n'] \subseteq \Theta(e)$$

som medfører

$$\forall e \in (F_0 \cap \Theta(e)), \exists n' : [e|n'] \subseteq \Theta(e)$$

som netop er, at $\Theta(e)$ er F_0-åben, det vil sige $\Theta(e) \in \Sigma_1^B$ i Borelhierarkiet over F_0. ⊣

Omvendt kan det nu vises, at hvis F_0 antages at være hele universet, det vil sige, at f starter ud med at vide ingenting, så kan AGM-metoden identificere alle åbne mængder.

Sætning 2.2. *Antag* $F_0 = U$*. Så kan en beregnbar AGM-metode identificere åbne mængder.*

Bevis. Lad $\Theta(e)$ være åben, samt d element heri. Heraf følger, at $\exists k \in \mathbb{N} : [d|k] \subseteq \Theta(e)$. Samtidig gælder der, da $F_0 = U$ at $\forall d' \in U : f(d'|n) = [d'|n]$, hvorfor vi har

$$\forall m \geq k : f(d|m) = [d|m] \subseteq [d|k] \subseteq \Theta(e).$$

Hvilket var, hvad der skulle vises. ⊣

Hermed er der givet en karakterisering af en beregnbar AGM-metode. Givet en 'velvalgt' initialmængde, F_0, samt antagelsen (1) ovenfor kan en AGM-metode indentificere et empirisk spørgsmål, hvis og kun hvis de enkelte celler i partitionen er F_0-åbne.

3. Diskussion af karakterisering og model

Ovenstående karakterisering er overraskende i lyset af [**Ke96**]. For det første karakteriseres den beregnbare AGM-metode i termer af det (fulde) endelig Borelhierarki og ikke i termer af det aritmetiske. For det andet er det overraskende, at de identificérbare mængder er Σ_1^B-mængder. Disse mængder karakteriserer ideale bekræftelsesmetoder, der over en datastrøm e verificerer h i endelig tid, hviss h er korrekt over e. Det overraskende element er, at der findes mange hypoteser, som er idealt verificérbare i endelig tid, men som en beregnbar metode går i stå på. Lad X være en mængde af tilstrækkelig aritmetisk kompleksitet, eksempelvis $X \in \Pi_1^A \setminus \Delta_1^A$. Lad hypotesen h_1 være hypotesen, der er korrekt over e hviss $e(0) \in X$. Denne kan enhver ideal metode afgøre efter at have læst den første observation i e, men hvis $e(0)$ ligger i X går en beregnbar metode måske ind i et uendeligt loop. Dette synes at være i modstrid med ovenstående karakterisering, for man kunne sætte $\Theta = (H, C)$, med $H = \{h_1, \overline{h_1}\}$ og C som angivet. En forklaring på denne umiddelbare uoverenstemmelse er nok karakteriseringernes følsomhed overfor de præcise definitioner. Gøres definitionen af identifikation stærkere, så der kræves $f(e|m) = \Theta(e)$ – ligesom det tidligere blev krævet i definitionen før karakteriseringsbeviset for den beregnbare opdagelsesmetode uden metodologiske krav – gælder sætning 2.7 ikke mere. Så er det kun tulipaner, der kan identificeres. Denne følsomhed over for definitionerne er i sig selv en interessant problemstilling, som jeg desværre ikke har haft tid til at forfølge.

I [**KSH96**] blev det vist at den ideale AGM-metodologi er fuldstændig. Det vil sige, at AGM-metoden kan identificere Borelhierarkiet til og med Σ_2^B. I denne sammenhæng er det interessant, at kravet om beregnbarhed gør, at de identificérbare mængder går fra at være Σ_2^B til at være Σ_1^B. Det vigtige i sammenligningen af disse to resultater er, at de begge bygger på det samme sæt AGM-aksiomer og den samme definition af identifikation.

Et andet aspekt, jeg har valgt at se bort fra, er selve AGM-aksiomernes udformning. Disse er formuleret således, at metodens reaktioner på dataen kun angives i det tilfælde, hvor dataen er konsistent med den mængde af overbevisninger, som skal opdateres. Givet dette er tilfældet angives en Gärdenfors-metode. Det er imidlertid en vigtig pointe for Gärdenfors at angive, hvorledes der opdateres, når mængden af epistemiske overbevisninger og dataen er inkonsistente. For Gärdenfors er der tre typer af opdatering: udvidelse, revision og kontraktion. For disse tre typer lykkes det for Gärdenfors kun at få entydigt bestemt en funktion for udvidelse.[4] For de to sidste typer er postulaterne, som Gärdenfors fremsætter ikke tilstrækkelige til at bestemme unikke funktioner. Problemet med de to sidste typer af opdateringer er, at der mangler en præcis logisk struktur på mængderne af epistemiske overbevisninger, og Gärdenfors må indføre mere eller mindre uspecificerede begreber som eksempelvis *entrenchment* for at give beskrivelser af mængderne. I denne sammehæng viser Gärdenfors, at *hvis* der eksisterer en kontraktionsfunktion, så eksisterer der en fuldstændig ordning med hensyn til entrenchment [**Gae88**]:81. At en sådan ordning til enhver tid og til alle mængder eksisterer, tvivler jeg på kan beskrives matematisk logisk, og endnu mindre at denne ordning er beregnbar. Da dette er et matematisk projekt, har jeg valgt ikke at gå dybere ind i dette spørgsmål, som jeg betragter som værende først og fremmest filosofisk i dets natur.

[4]Denne er at lægge dataen til ens mængde af overbevisninger og lukke under deduktion.

Restriktivitet og Gärdernfors-metodologi

Det er nu blevet vist, at Gärdenfors-metodologien *kan* implementeres på en computer, og at den netop kan identificere åbne mængder. Dette resultat hviler på den konkrete formulering af metodologien, som den kommer til udtryk i AGM-aksiomerne. Det gør den næste undersøgelse ikke, hvor det undersøges, om arkitekturen er restriktiv for beregnbare opdagelsesmetoder. Gärdenfors-metodologien *er*, som det nu vises, restriktiv, og det er den i kraft af kravet om, at opdateringeren af mængden af epistemiske overbevisninger skal ske således, at den opdaterede mængde implicerer den data, metoden har set. Restriktiviteten påviser et problem ved Gärdenfors-metodologien, da et krav til et rationelt induktivt princip må være, at det bidrager til opdagelsen af korrekte teorier, eller som et minimum ikke stiller sig i vejen for disse opdagelser. At Gärdenfors-metodologien er restriktiv er ikke noget nyt resultat. Hendricks har i sin Ph.D.-afhandling [**He96**] vist dette, hvor han benytter et resultat fra [**OSW86**].

Ideen er at specificere korrekthedsrelationen og give metoden en passende baggrundsviden. Metoden skal over e identificere den r.e. mængde, der udgøres af værdimængden til e, hvor en mængde X er r.e. hviss $X \in \Sigma_1^A$. Metoden antager derfor som baggrundsviden, at den ikke vil blive præsenteret for andet en r.e. mængder. Motivationen for at specificere det generelle setup i overensstemmelse med dette er følgende.

Man kan se en opdagelsesmetode som en metode til identifikation af naturlige sprog. Metoden må derfor stabilisere til en korrekt grammatik for det sprog, som metoden bliver præsenteret for. En grammatik giver en positiv test for et sprog. Dette vil sige, at hvis en sprogbruger har forstået grammatikken og bliver præsenteret for en sætning, siger han at sætningen tilhører sproget, hvis og kun hvis sætningen rent faktisk tilhører sproget. Derfor modeleres sprog som r.e. mængder, og det kræves af metoden, at den før eller senere stabiliserer til et korrekt indeks for den r.e. mængde, som metoden bliver vist.

Aksiomatisérbare teorier er også en del af motivationen. Lad Y være en mængde af lukkede formler og lad R_Y være den egenskab, at $R_Y(x)$ gælder hviss $x \in Y$. Inden for rekursionsteorien definerer man nu en teori T som aksiomatisérbar, hvis der findes en række postulater P for T således at R_P er r.e., hvilket—som i ovenstående afsnit—vil sige, at man har en positiv test. Dette betyder samtidigt, at givet et sæt postulater P for en aksiomatisérbar teori T kan vi effektivt finde et r.e. indeks for R_P. Man kan nu rekursionsteoretisk vise, at en teori T er aksiomatisérbar, hviss R_T er r.e. Og det følger, at en teori er aksiomatisérbar, netop når postulaterne for teorien effektivt kan genereres.[1]

1. Identifikation af r.e. mængder

Lad W_i være domænet til den rekursive funktion med indeks i, det vil sige φ_i. Da en mængde X er r.e., hviss den er domænet til en rekursiv funktion, kan indlæringsparadigmet for *identifikation af r.e. indekser* angives.

Definition 1.1.

1. Korrekthed: $C_{r.e.}(e, i) \equiv range(e) = W_i$.
2. Baggrundsviden: $K := \{e \mid range(e) \text{ er r.e.}\}$.

[1] Se eventuelt [**BM77**] side 332 for en rekursionsteoretisk diskussion af aksiomatisérbare teorier.

I dette setup bliver data retention til:

Definition 1.2. Data retention$_{\text{r.e.}}$:

$$f \text{ er data retentive}_{\text{r.e.}} \text{ inden for } K \equiv \forall \varepsilon \in K^* : range(\varepsilon) \subseteq W_{f(\varepsilon)}.$$

Strategien er nu at kræve, at f er beregnbar og data retentive, og herefter undersøge hvor meget f med disse to krav kan identificere. Dette svarer til det andet af mine arbejdsspørgsmål: Givet at metoden *er* beregnbar og opfylder den del af Gärdenfors-metodologien, som kræver data retention, er metodologien så fuldstændig?

Da korrekthedsrelationen er specificeret, ændres definitionen af identifikation, som den blev givet på side 11, således at f nu identificerer mængder og ikke længere empiriske spørgsmål Θ. Den ændrede definition er:

$$\begin{aligned} f \text{ identificerer } \Sigma_1^A \text{ givet } K \quad &\equiv \\ \forall T \quad &\in \quad \Sigma_1^A, \forall e \in K, \exists n \forall m \geq n : W_{f(e|m)} = T \leftrightarrow range(e) = T. \end{aligned}$$

Der arbejdes også med, at f identificerer enkelte r.e. mængder, hvilket defineres som ovenstående, hvor, Σ_1^A skiftes ud med $T \in \Sigma_1^A$ før ækvivalenstegnet og tage $T \in \Sigma_1^A$ væk efter ækvivalensen.

Lemma 1.3. *(Blum og Blum). Antag, at f identificerer $T \in \Sigma_1^A$. Da eksisterer der $\varepsilon \in U^*$ sådan at:*

1. $range(\varepsilon) \subseteq T$.
2. $W_{f(\varepsilon)} = T$.
3. $\forall \tau \in U^*$ *gælder der, at* $range(\tau) \subseteq T$ *medfører* $f(\varepsilon * \tau) = f(\varepsilon)$.

Bevis. (Som i [**OSW86**]). Antag lemmaet er falsk, hvilket giver

(*): For alle $\sigma \in U^*$, sådan at $range(\sigma) \subseteq T$ og $W_{f(\sigma)} = T$, eksisterer der $\tau \in U^*$ hvor om det gælder at $range(\tau) \subseteq T$ og $f(\sigma * \tau) \neq f(\sigma)$.

Der gælder nu at **(*):** implicerer eksistensen af en datastrøm e sådan, at $range(e) = T$ men f identificerer ikke T, hvilket er i modstrid med antagelsen om, at f indentificerer T. Lad $d \in U$ sådan at $range(d) = T$. e konstrueres nu induktivt.

Trin 0. Lad σ_0 være sådan at $range(\sigma_0) \subseteq T$ og $W_{f(\sigma_0)} = T$. σ_0 eksisterer, da f identificerer T.
Trin $n+1$. Givet σ_n er der to tilfælde. Hvis $W_{f(\sigma_n)} \neq T$ så $\sigma_{n+1} = \sigma_n * (d(n))$. Ellers følger der af **(*):** eksistensen af $\tau \in U^*$ sådan at $range(\tau) \subseteq T$ og $f(\sigma_n * \tau) \neq f(\sigma_n)$. Sæt $\sigma_{n+1} = \sigma_n * \tau * (d(n))$.

Lad $e = \bigcup_n \sigma_n$. Da gælder der, at $range(e) = T$ og f stabiliserer aldrig til noget indeks for T, da det for alle n gælder, at $W_{f(\sigma_n)} \neq T$ eller $f(\sigma_n * \tau) \neq f(\sigma_n)$. ⊣

Lad $T \in \Sigma_1^A$, f opdagelsesmetode og $\varepsilon \in U^*$ være givet. Hvis ε opfylder punkterne 1 til 3 i lemma 1.3 kaldes ε for en *låsende følge for T og f*.

Sætning 1.4. *(Osherson, Stob og Weinstein). Lad f identificere $\Lambda \subseteq \Sigma_1^A$ givet $K := \{e \mid range(e)$ er r.e.$\}$. Hvis f er data retentiv og rekursiv så $\Lambda \subseteq \Delta_1^A$.*
Bevis. Lad $T \in \Lambda$. Ifølge ovenstående lemma eksisterer der en låsende følge ε for T og f. Hermed kan vi skabe en effektiv afgørbar test for elementer i T, hvorfor $T \in \Delta_1^A$. Lad $x \in T$. Da gælder der $f(\varepsilon * (x)) = f(\varepsilon)$, da ε er en låsende følge. På den anden side gælder der, at

hvis $x \notin T$, så er $f(\varepsilon * (x))$ ikke et indeks for T, da f er data retentive. Altså $x \in T$ hviss $f(\varepsilon * (x)) = f(\varepsilon)$. Da f er rekursiv og defineret for alle naturlige tal n, er dette en afgørbar test. $\dashv$

Der findes mængder, der kan identificeres af beregnbare opdagelsesmetoder inden for paradigmet for identifikation af r.e. indekser, hvor mængderne indeholder ikke-rekursive elementer.[2] Heraf følger det, at Gärdenfors-metodologien er en restriktiv arkitektur og derfor ikke fuldstændig.

2. Metodologiske krav og restriktivitet

Ovenstående resultat er som nævnt ikke originalt. Men i en diskussion af Gärdenfors' metodologi er det relevant at angive, da det viser, at et 'rationelt' princip for videnstilegnelse som data retention kan forhindre en metode i at nå sandheden. Det virker overraskende, at et simpelt princip som at huske den data, man har set, og fremsætte teorier der gør denne data sand, er et restriktivt krav. Pointen er ikke desto mindre, at denne procedure kan være ikke-beregnbar. For ideale metoder derimod er data retention et fornuftigt krav, som det eksempelvis er bevist i [**KSH96**], hvor fuldstændigheden af AGM-aksiomerne for ideale metoder er vist. I visse tilfælde kan lignende principper som konsistens endda være et nødvendigt krav for at optimere eksempelvis en ideal bekræftelsesmetodes udnyttelse af daten, som Kelly har vist det i [**Ke96**].

[2] Definér $K = \{i \mid \varphi_i(i) \downarrow\}$. K er en 'ægte' r.e. mængde. $T = \{K \cup \{x\} \mid x \in \overline{K}\}$ kan identificeres af en beregnbar opdagelsesmetode, se [**OSW86**]:49.

Videre arbejde

De formelle undersøgelser af de beregnbare aspekter ved Gärdenfors' metodologi er bestemt ikke afsluttet med mine undersøgelser. Mange vigtige spørgsmål trænger sig på.

I første omgang må det dreje sig om at få afdækket de præcise sammenhænge mellem den beregnbare AGM-metode og beregnbare metoder uden metodologiske krav i almindelighed. De to sætninger, der giver karakteristiker af de to metoder, er i første omgang ikke helt sammenlignelige, da kriterierne for identifikation er lidt forskellige. Yderligere gives karakteristikken i det ene tilfælde i Borel-termer, mens den i det andet gives i aritmetiske. I denne sammenhæng kunne det være interessant at få en præcis analyse af restriktiviteten af AGM-aksiomerne, som ikke går via paradigmet for identifikation af r.e. indekser. Ovenstående resultat vedrørende restriktivitet af AGM-aksiomerne ville være mere værdifuldt i den generelle model, hvor det ikke kun er r.e. indekser, der identificeres og r.e. mængder, som antages som baggrundsviden.

I anden omgang må det handle om at formulere nogle nye aksiomer for Gärdenfors' metodologi. Aksiomer, som ikke kun beskriver AGM-metodens opførsel i det tilfælde, hvor dataen så at sige kommer konsistent ind. Jævnfør min diskussion af AGM-modellen ser jeg dette som en vanskelig opgave. Men i arbejdet med udarbejdelsen af nye formuleringer, kunne kravene til opdateringen formuleres således, at disse i en efterfølgende analyse af metoden kunne svækkes. Derved kunne man klarlægge hvilke krav, der var årsag til eksempelvis restriktivitet, og hvad der i givet fald skulle til af svækkelser, for at få en revideret AGM-metode til at være pålidelig. I den nuværende model anser jeg det for sandsynligt, at en fuldstændig metode kan opnås, ved at svække kravet om data retention til et krav om lokal konsistens, hvilket er et noget svagere krav. I et ambitiøst og matematisk tungt projekt ville nye aksiomer blive formuleret inden for den rekursive deskriptive mængdelære.

Slutteligt ville det være interessant at få undersøgt, hvorledes AGM-metoden relaterer sig til andre belief revision metoder med hensyn til pålidelighed. Kelly sammenligner i artiklen [**Ke98**] en række forskellige metoder fremsat af blandt andre Spohn og Pearl og får en række generelle strukturligheder og -forskelligheder frem. En undersøgelse i den stil med hensyn til AGM-aksiomerne ville have vigtige metodologiske implikationer.

REFERENCER

[Gae88] Gärdenfors, P. (1988). *Knowledge in Flux*, Cambridge: MIT Press.

[BM77] Bell, J.L og Machover, M. (1977). *A Course in Mathematical Logic*, Amsterdam: North Holland.

[Gl92] Glymour, C. (1992). *Thinking Things Through*, Cambridge: MIT Press.

[Go67] Gold, E.M. (1967). "Language Identification in the Limit", *Information and Control*, **10**: 447-474.

[Gr88] Grove, A. (1988). "Two Modellings For Theory Change", *Journal of Philosophical Logic*, **17**: 157-170.

[He96] Hendricks, V.F. (1996). "Epistemology, Methodology and Reliability", Ph.D.-Dissertation, Dept. of Philosophy, University of Copenhagen.

[He99] Hendricks, V.F. (1999). *The Convergence of Knowledge*, Draft.

[Ke96] Kelly, K. (1996). *The Logic of Reliable Inquiry,* Oxford: Oxford University Press.

[Ke98] Kelly, K. (1998). "Iterated Belief Revision, Reliability, and Inductive Amnesia", Carnegie Mellon Philosophy Department Tech Report CMU Phil-88.

[KSH96] Kelly, K., Schulte, O. and Hendricks, V.F. (1996). "Reliable Belief Revision", *Proceedings of the 10th International congress of Logic, Methodology and Philosophy of Science,* ed. Ciara, K.L.D.

[No81] Nozick, R. (1981). *Philosophical Explanations*, Cambridge: Harvard University Press.

[OSW86] Osherson, D., Stob, M. and Weinstein, S. (1986). *Systems That Learn*, Cambridge: MIT Press.

[Po63] Popper, K. (1963). *Conjectures and Refutations*, London: Routledge & Kegan Paul.

[Pu63] Putnam, H. (1963). "Degree of Confirmation and Inductive Logic", *The Philosophy of Rudolf Carnap*, ed. Schlipp, La Salle. Open Court.

[Sc93] Schoenfield, J.R. (1993). *Recursion Theory*, Berlin: Springer.

[Sc95] Schulte, O. (1995). "The Computable Testability of Uncomputable Theories", Master's Thesis, Dept. of Philosophy, Carnegie Mellon University.