Eric Johannesson

Goodmans induktionsproblem

1. Det gamla och det nya induktionsproblemet

För Hume var induktionsproblemet ungefär följande: hur kommer det sig att vi utifrån ett begränsat antal observationer kan dra slutsatser beträffande det som ännu inte observerats? Hur vet vi exempelvis att solen går upp imorgon? Mer allmänt: hur vet vi att naturen är regelbunden, dvs. att framtiden liknar det förflutna? Vi verkar nämligen inte kunna rättfärdiga sådana slutsatser på rent logiska/deduktiva grunder. Ty betrakta härledningsregeln

$$\frac{\forall x(\text{Observerad}(x) \wedge F(x) \rightarrow G(x))}{\forall x(F(x) \rightarrow G(x))} \tag{1}$$

Även då antalet observationer är stort, är det lätt att inse att de flesta instanser av denna regel kommer vara deduktivt ogiltiga: det kommer finnas möjliga världar där premissen är sann men slutsatsen falsk. Om exempelvis $F(x)$ betyder att x är en smaragd, och $G(x)$ betyder att x är grön, så finns det en möjlig värld där alla observerade smaragder är gröna, men vissa icke-observerade smaragder är blå. Denna värld utgör då ett motexempel.

Vad Goodman (1955) visade med ”the new riddle of induction” var att den ovan föreslagna härledningsregeln inte bara är *deduktivt* ogiltig, men också *induktivt* ogiltig. För även om vi begränsar oss till världar som *är* regelbundna i Humes mening (världar där solen alltid går upp, alla smaragder är gröna, allt bröd är närande, etc.), så är den föreslagna härledningsregeln *ändå* ogiltig. Givet en viss tolkning av F och G, och även om vi begränsar oss till världar där både premissen och slutsatsen i så fall är sann, går det nämligen att konstruera en *annan* instans av härledningsregeln enligt vilken en av dessa världar bildar ett motexempel. Det räcker att definiera ett predikat G' så att

$$G'(x) \Leftrightarrow (\text{Observerad}(x) \wedge G(x)) \vee (\neg\text{Observerad}\ (x) \wedge \neg \mathrm{G}(x)) \tag{2}$$

Filosofisk tidskrift 2014 nr 4, 18–22

För i en värld där alla F är G, men vissa F inte är observerade, så kommer

$$\forall x(\text{Observerad}(x) \wedge F(x) \rightarrow G'(x)) \tag{3}$$

vara sann, men

$$\forall x(F(x) \rightarrow G'(x)) \tag{4}$$

vara falsk.

2. DEFINITIONEN AV ”GRUE”

Enligt Goodman (1955) är ett objekt *grue* om och endast om det har observerats före en viss tidpunkt t och är grönt, eller inte har observerats före denna tidpunkt och är blått. Som ni ser skiljer sig detta något från hur G' definieras i (2) ovan. Skillnaden är emellertid obetydlig. Argumentet fungerar nämligen även om vi med Observerad(x) menar att x har observerats *vid minst en tidpunkt* (*vilken som helst*). Tidpunkten t är alltså en överflödig ingrediens i Goodmans definition. Förekomsten av den verkar dessutom ha givit upphov till en del onödiga missförstånd.

Om man läser Goodman slarvigt är det nämligen lätt att få intrycket att ”grue” betyder ”grön före t, och blå därefter”. Som kandidatstudent gjorde jag själv detta misstag (vilket noterades av min dåvarande handledare). Då ansåg jag misstaget vara harmlöst. Det anser jag fortfarande. Därvidlag befinner jag mig i gott sällskap. Gärdenfors (1990) använder sig nämligen av precis denna definition. Han påpekar visserligen att den inte är exakt som Goodmans, men att det inte har någon betydelse för argumentet som följer. Dock stöter man ibland på uppfattningen att Gärdenfors definition gör att det nya induktionsproblemet reduceras till det gamla. Det är kanske lätt att förstå varför. Om den alternativa hypotesen går ut på att smaragder efter en viss tidpunkt ändrar färg, låter ju detta onekligen mer som Humes problem: hur vet vi att smaragder kommer vara gröna även i morgon? Skillnaden mellan det nya och det gamla induktionsproblemet kan emellertid vara ganska subtil. Om det gamla problemet är hur vi kan veta att naturen är regelbunden, så är det nya problemet hur vi kan veta att naturen är regelbunden på vissa sätt snarare än andra.

För att förvissa oss om lämpligheten hos Gärdenfors definition, låt $G(x,u)$, $B(x,u)$ och $G'(x,u)$ betyda att x är grön vid tidpunkten u, blå vid tidpunkten u, respektive *grue* vid tidpunkten u, där G' definieras enligt

$$G'(x,u) \Leftrightarrow (u < t \rightarrow G(x,u)) \wedge (t \leq u \rightarrow B(x,u)) \tag{5}$$

Notera att om $u < t$, så gäller att $G(x,u)$ om och endast om $G'(x,u)$. Allt som har observerats som grönt före t, kommer alltså även ha observerats som grue, och vice versa.

Det finns nu två hypoteser:

1. Alla smaragder är gröna vid varje tidpunkt.
2. Alla smaragder är grue vid varje tidpunkt.

I den mån det finns några smaragder över huvud taget efter tidpunkten t, så är dessa hypoteser naturligtvis ömsesidigt uteslutande. Låt $F(x,u)$ betyda att x är en smaragd vid tidpunkten u, och betrakta följande härledningsregel:

$$\frac{\forall x \forall u(\text{Observerad}(x,u) \wedge F(x,u) \rightarrow G(x,u))}{\forall x \forall u(F(x,u) \rightarrow G(x,u))} \tag{6}$$

Om vi väljer att kvantifiera över mängden ordnade par av objekt och tidpunkter (mängden objekt-tidsdelar), snarare än över objekten och tidpunkterna var för sig (vilket är logiskt ekvivalent), så kan vi betrakta (6) som en instans av (1). Om vi vill kan vi sålunda även omformulera våra hypoteser:

1′. Alla smaragd-tidsdelar är gröna.
2′. Alla smaragd-tidsdelar är grue.

Förutsatt att alla observationer har ägt rum före t, och att alla smaragd-tidsdelar faktiskt är gröna, så kommer

$$\forall x \forall u(\text{Observerad}(x,u) \wedge F(x,u) \rightarrow G'(x,u)) \tag{7}$$

vara sann, men

$$\forall x \forall u(F(x,u) \rightarrow G'(x,u)) \tag{8}$$

vara falsk. Vi har därmed visat att Gärdenfors definition ger upphov till ett argument med exakt samma logiska form som Goodmans. Det misstag jag begick som kandidatstudent var således inte allvarligare än om jag hade råkat byta ut smaragder mot exempelvis gräs.

Gärdenfors definition har dessutom fördelen att vara immun mot Frank Jacksons lösning (Jackson, 1975), som går ut på att på att begränsa induktionsschemat till egenskaper F och G som uppfyller det kontrafaktiska villkoret att objektet hade haft egenskaperna ifråga även om det inte hade observerats. En observerad grön smaragd hade inte varit grue enligt Goodmans definition om den inte hade observerats, den hade bara varit grön. Goodmans ”grue”, till skillnad från Gärdenfors, uppfyller alltså inte detta villkor.

3. UNDERBESTÄMNINGSPROBLEMET

Egentligen är Goodmans induktionsproblem bara en instans av det mer generella underbestämningsproblemet som uppstår i förhållandet mellan observation och hypotes. Insikten att varje mängd observationer kan förklaras (förutsägas) av ett oändligt antal ömsesidigt uteslutande hypoteser brukar ofta förknippas med Quine och Duhem. Men Leibniz gjorde väsentligen samma observation redan 1703. I ett brev till Jakob Bernoulli (Brev XIII, 3 december 1703) nämner han att han kan visa att det genom varje uppsättning punkter löper ett oändligt antal "regelbundna" (i modern terminologi: beräkningsbara) kurvor.

Likaledes kan observationen av gröna smaragder förklaras både av hypotesen att alla smaragder är gröna, såväl som att alla smaragder är grue. Eller, för att tala med Descartes, att vi är lurade av en ond demon. Så även om Goodmans induktionsproblem är nytt i förhållande till Hume, är det knappast nytt i förhållande till Quine, Duhem eller Leibniz. Frågan är om det är nytt ens i förhållande till Descartes.

Ur ett Bayesianskt perspektiv handlar underbestämningsproblemet om att tillskriva de olika hypoteserna rimliga initiala sannolikheter (på engelska: *prior probabilities*). Enligt Bayes teorem är nämligen sannolikheten för en viss hypotes givet en viss observation proportionerlig mot produkten av två faktorer: dels observationens sannolikhet enligt hypotesen, dels hypotesens initiala sannolikhet. Vi vill på förhand kunna utesluta orimliga hypoteser (såsom att alla smaragder är grue), dvs. tillskriva dem låga initiala sannolikheter, vi vet bara inte hur vi ska rättfärdiga det.

För att se hur detta fungerar i praktiken, definiera

$$\begin{aligned} H_1 &= \text{Alla smaragder är gröna} \\ H_2 &= \text{Alla smaragder är blå} \\ H_3 &= \text{Alla smaragder är grue} \\ E &= \text{Alla observerade smaragder är gröna} \end{aligned} \tag{9}$$

I den mån det finns smaragder som ännu inte observerats, är hypoteserna H_1, H_2, och H_3 ömsesidigt uteslutande. Låt oss för enkelhetens skull också anta att det inte finns några andra hypoteser. För de initiala sannolikheterna gäller då att $P(H_1) + P(H_2) + P(H_3) = 1$. Enligt Bayes teorem får vi följande:

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E|H_1)P(H_1)+P(E|H_2)P(H_2)+P(E|H_3)P(H_3)} = \frac{P(H_1)}{P(H_1)+P(H_3)} \tag{10}$$

$$P(H_2|E) = \frac{P(E|H_2)P(H_2)}{P(E|H_1)P(H_1)+P(E|H_2)P(H_2)+P(E|H_3)P(H_3)} = 0 \quad (11)$$

$$P(H_3|E) = \frac{P(E|H_3)P(H_3)}{P(E|H_1)P(H_1)+P(E|H_2)P(H_2)+P(E|H_3)P(H_3)} = \frac{P(H_3)}{P(H_1)+P(H_3)} \quad (12)$$

Som ni ser är de betingade sannolikheterna för H_1 och H_3 helt proportionerliga mot sina respektive initiala sannolikheter (den betingade sannolikheten för H_2 är däremot 0 alldeles oavsett).

Bayesianer kritiseras ibland för att de inte säger något om hur man erhåller rimliga initiala sannolikheter. Det är egentligen en ganska märklig kritik. För, som vi har sett, om man kunde lösa *det* problemet, så skulle man också kunna lösa induktionsproblemet (det nya såväl som det gamla). Och det är det, mig veterligen, ännu ingen som har gjort.

REFERENSER

Goodman, N. 1955. *Fact, Fiction and Forecast*. Cambridge, Mass.: Harvard University Press.

Gärdenfors, P. 1990. Induction, Conceptual Spaces and AI. *Philosophy of Science* 57, nr 1, s. 78–95.

Jackson, F. 1975. Grue. *Journal of Philosophy* 72, nr 5, s. 113–31.