BERTIL ROLF

Om trovärdighet

1. Typer av trovärdighetsproblem

Det finns olika typer av problem som förknippas med trovärdighet. Kunskapsproblem handlar om skäl för att lita på information vi får genom expertis och vittnen. Hur kan man veta så mycket om världen, trots att man själv sett en så begränsad del av den?

En annan typ av problem har behandlats av spelteoretiker och ekonomer. Problemen gäller trovärdighet hos löften eller hot. Är exempelvis en ekonomisk politik mer trovärdig med en oberoende riksbank?

En tredje typ av problem avser social och politisk ordning. I "Making Democracy Work" (1993) studerar Robert Putnam två sociala ordningar i Italien i ett försök att knyta samman tillit ("trust"), mellanmänskliga normer och ekonomisk-politiska institutioner. En svensk studie över problematiken är Per Molanders vackra bok "Akvedukten vid Zaghouran" (1994). Till denna grupp hör också studier över den minskande tillit till folkvalda politiker som visar sig i flertalet västliga länder.

En fjärde typ av problem handlar om moralisk auktoritet och trovärdighet. Ibland förutsätts att talesmän för en moralisk ståndpunkt har en "trovärdighet" som bl a innebär att de efterlever normer som de pläderar.

I det följande skall jag diskutera de två första typerna av trovärdighetsproblem. Jag skall visa att trovärdighet av löften eller hot till stora delar kan behandlas med samma slags begrepp som trovärdighet av information.

2. Det s k "genetiska felslutet" och bevisvärdemodellen

Begreppet "trovärdighet" kom jag i kontakt med för många år sedan då jag höll kurser i argumentation, källkritik och vittnespsykologi.

Filosofisk tidskrift 1995 nr 4, 151-159

Läroböckerna i argumentationsanalys dömde ut argument ad hominem. Vittnespsykologi och källkritik tycktes systematiskt tillämpa något som liknar argument ad hominem.

I stället för termen "ad hominem" talar argumentationsteoretiker mer precist om "det genetiska felslutet". Det förmodade felslutet består i att "man angriper personen *istället* för hans påståenden". Men vittnespsykologi och källkritik bygger på att det inte finns någon motsättning mellan granskning av person och granskning av sak. Båda har utvecklat systematiska rutiner för att studera hur en utsaga uppkommit. Utsagans uppkomst anses väsentlig för den tillit vi bör fästa till den. Det tycks ligga något rimligt i att en person som har starka intressen att få oss att tro att p kanske inte är den mest tillförlitliga källan rörande p.

Det s k "genetiska felslutet" kan analyseras med metoder som hämtas från bevisvärdering. I början av 1970-talet utvecklades en teori om juridisk bevisvärdering vars grundläggande tankar återfinns hos Per-Olof Ekelöf (1964), Sören Halldén (1973) och Martin Edman (1973).

Ett centralt problem avser förhållandet mellan frekvens och evidens. Låt säga att en anlagd brand bryter ut i det lilla samhället. På orten bor en person som tidigare dömts för mordbrand och haft tillfälle och kunskaper att utföra dådet. Antag att återfallsfrekvensen är 90% och att det är ytterst osannolikt att någon annan haft nog lokalkännedom för att anlägga branden. Varför inte gripa och döma honom för mordbranden? Det rimliga svaret är att personen inte "kan bindas till" brottet, trots att vi av allmänna samband kan sluta oss till en hög sannolikhet för att han begått dådet.

Vad betyder då denna fras "kan bindas till"? Det som binder en person till brottet är evidens i form av fingeravtryck på en bortkastad fotogendunk eller vittnesuppgifter om möten med den tilltalade i närheten av brandplatsen, utsändande en stark röklukt.

Bevisvärdeteorin går ut på att domstolens bedömning vilar på evidens för att något slags "indiciemekanism" eller "informationskedja" (termerna är Halldéns), eventuellt i form av en kausalkedja, finns mellan den anlagda branden och de senare vittnesmål eller den tekniska bevisning som förevisas domstolen. Domstolen överväger inte (eller bör inte överväga) den direkta sannolikheten för att X begick H. Domstolen överväger istället sannolikheten för att vittnesmål och teknisk bevisning beror av att X begick H. Evidensen för "Per tände

på Påls lada" stärks via evidensen för "Pers antändning av Påls lada är kausalt länkat till förekomsten av Pers fingeravtryck på den fotogendunk som undersökts av rättsteknikerna".

Bevisvärdeteorin löser problemen med det s k "genetiska felslutet". Bakom en utredning som uttalar sig om tobaksrökningens oskadlighet finns urval, metoder och procedurer. Om utredningen sponsrats av en cigarettfabrikant kan man inte utesluta att rapportens utsaga om oskadligheten har annat upphov än den påstår. Uppkomsten är relevant för vår tilltro. Argumentationsteoretiker som dömt ut det s k "genetiska felslutet" tycks ha trott att det genetiska betraktelsesättet försökt argumentera från "undersökningen har sponsrats av Philip Morris" till "rökning är skadligt" vilket givetvis vore ett felslut.

3. Generalisering av teorin till trovärdighet av information

Under 1980-talets början kom Nils-Eric Sahlin (1986) och jag själv (Rolf 1983) att, oberoende av varandras insatser, generalisera bevisvärdeteorin till en generell modell för kunskap eller kunskapsöverföring. Med viss förenkling kan modellen uttryckas:

> (K) X vet att p om och endast om (1) X tror att p; (2) Det finns en fungerande informationskedja mellan p och X:s tro att p och (3) X har evidens för (2).

Medan reliabilitetsteorier för kunskap nöjer sig med villkor (1) och (2) ställer (K) också krav på evidens. Motivet för krav (3) framgår genom fabeln om pojken och vargen. För att lura byborna ropade vallpojken "Vargen kommer" nio gånger utan att vargen kom. Den tionde gången kom vargen men byborna hade slutat lita på pojkens rop. Byborna har evidens mot (2) och pojken kan inte längre överföra sin kunskap till dem, trots att informationskedjan faktiskt fungerar i det tionde fallet.

Genom kombinationen av evidens och reliabilitet på ett sätt som icke är ad hoc, kringgår (K) Russell-Gettier problematiken. Om det finns falska mellanled mellan personens övertygelse och sakförhållandet p, så är villkor (2) inte uppfyllt.

Idag känner jag tveksamhet inför (K) som ett bidrag till samtida kunskapsteori. Dels förstår jag inte hur genren relaterar sina teorier till preteoretiska kunskapsproblem och kunskapsbegrepp eller hur man använder data, argument och metoder. Dels vill jag fokusera de procedurer och institutioner som levererar information snarare än det slut-

resultat "kunskap" som eventuellt uppkommer i den som mottar informationen och då till följd av ett flertal andra faktorer, exempelvis bakgrundsinformation. Jag föreslår i stället en modell som fångas i följande definition:

> DEF: Låt A vara en person eller institution som i situationen s utfärdar påståendet p som mottas av B. A:s utsägande av p i s skall sägas vara "trovärdigt" för B om och endast om B har evidens för att det finns en tillförlitligt fungerande informationskedja mellan p och A:s utfärdande av p i s

I förenklande syfte bortser modellen från gradering av trovärdighet.

4. Tillförlitliga informationskedjor

Ett problem gäller vad som skall menas med att en informationskedja eller indiciemekanism fungerat "tillförlitligt" och hur man etablerar evidens för det. En vanlig tanke är att bestämma tillförlitlighet genom en kontrafaktisk villkorssats: Vittnesmålet om p skulle inte förelegat såvida inte p varit fallet.

Denna lösning kan medföra problem. Säg att en notorisk skattefuskare till taxeringsmyndigheterna varje år deklarerar uppgift D(p) om nollinkomst. Att nollinkomst förelegat kallar vi "p". Men 1994 gjorde han stora spekulationsförluster och hade faktiskt nollinkomst. Den kontrafaktiska teorin tycks då medföra att informationskedjan fungerat endast om:

> (S) Skattefuskaren skulle inte ha deklarerat nollinkomst såvida han inte haft nollinkomst.

Men (S) är falsk, trots att informationskedjan fungerat. Det kan rent av vara så, att taxeringsmyndigheten har evidens för att informationskedjan fungerat vid 1994 års taxering, trots att skattefuskaren alltid tidigare bluffat. Han kan vid denna taxering förete korrekt bestyrka yrkanden på avdrag.

Problemet med den kontrafaktiska kopplingen är att utan p kan något annat komma att ge upphov till D(p). Den lösning som förefaller mig mest rimlig är att betrakta uppgiften D(p) som en rigid designator: just *denna* deklarationsuppgift jämte underlag, D(p), som uppger nollinkomst, skulle inte ha förelegat såvida inte nollinkomst förelegat. Vi kan då precisera:

DEF: Säg att det föreligger information om p, I(p). En informationskedja bakom I(p) sägs ha fungerat "tillförlitligt" om och endast om i varje tillstånd där *denna* informationskedja föreligger, föreligger också p.

Säg att bevismaterial för p i form av en blodig handske företes tillsammans med undersökningsprotokoll av handsken jämte protokoll om att handsken hittats i mordoffrets bil. För domstolen är detta yttersta ändpunkten av en informationskedja vars närmare egenskaper man inte känner. Man överväger då om *denna* informationskedja som levererat handsken skulle kunna finnas i en värld där p är falsk. Frågan man bör ställa sig är *inte*: "Skulle detektiven ha planterat bevismaterial om han inte funnit något". Utan den för bevisningen relevanta frågan är: "Planterade han *detta* bevismaterial *den här* gången?" Svaret på frågan beror av de möjligheter som är förenliga med *denna* indiciemekanism.

5. Trovärdighet av utfästelser

Det finns trovärdighetsforskning som sysslar med ett annat begrepp "trovärdighet". En gren härstammar från Thomas Schelling "The Strategy of Conflict" (1963). Bakgrunden är terrorbalansen och den amerikanska strategi som i fackkretsar förkortades MAD (Mutual Assured Destruction). Strategin gick ut på att USA skulle bemöta ett sovjetiskt angrepp på Västeuropa med en kärnvapenattack. Detta antogs leda till en okontrollerad eskalation av kärnvapenkriget. Ett sådant hot är svårt att göra trovärdigt. Hotet har ungefär samma logik som att försöka skrämma bort inbrottstjuvar i grannens hus genom att hota spränga sig själv i luften.

Liknande problem kan uppkomma vid överenskommelser. Ett stort byggföretag kontrakterar en hantverkare som utför en installation. Efter att installationen har utförts kan byggföretaget underlåta att betala eftersom hantverkaren inte gärna kan göra sitt arbete ogjort. Kan byggföretaget med trovärdighet utfästa sig att betala?

Det finns två slags analyser av varför utfästelser förknippas med trovärdighetsglapp. En analys (Coleman 1990, Persson 1987) går ut på att det finns ett tidsglapp mellan utfästelse och verkställan. Generellt sett har många utfästelser, dvs löften och hot, formen:

A utfäster sig vid t_1 [A ser till att p vid t_2] (där t_1 föregår t_2)

A kan ändra sina preferenser mellan t_1 och t_2. Odysseus skulle efter att ha hört sirenernas sång ha övermannats av sin längtan att spendera sin återstående tid i deras sällskap. I fallet MAD uteslöt USA:s preferenser både vid t_1 och vid t_2 att man skulle komma att utlösa ett storskaligt kärnvapenkrig.

En andra typ av analys (Schelling 1963, Elster 1984) vill förstå trovärdighetsglappet som ett kontrollglapp. Jag tror detta är riktigt. Mer generellt kan man säga att trovärdighetsproblemen för utfästelser beror av följande underliggande form:

A utfäster sig i s_1 [Någon ser till att p vid s_2] (där s_1 och s_2 är olika situationer)

Ofta finns komplexa antaganden bakom en utfästelses trovärdighet. Ibland vet man inte exakt vem som utför den handling man åtar sig. Om en städfirma åtar sig ett uppdrag, åtar man sig bara att se till att någon utför städningen. Trovärdighetsglapp kan då uppkomma genom att fallissemang hos någon annan än den som gjort utfästelsen.

Tidsglapp kan vålla glapp i trovärdighet. Men det beror då av att agenten som gör utfästelsen själv kan komma in vid t_2 samt att han inte vid t_1 kan binda besluten vid t_2. Själva tidsglappet är en konsekvens. Trovärdighetsglapp uppkommer genom oförmåga att binda instanser som kommer mellan utfästelse och verkställighet.

6. Tillförlitlighet och trovärdighet av utfästelser

Vi har diskuterat trovärdighet hos informationskedjor. Man kan använda en symmetrisk begreppsapparat för att tala om trovärdighet hos vad jag vill kalla *utfästelsekedjor*. Bevisvärdeteorin och mer generellt, trovärdighetsteorin, handlar om sådan evidens som kan "binda" en person till ett faktum. Trovärdighet av löften eller hot handlar om att en person skall "binda sig" till en framtida händelse. Vi kan tala om tillförlitlighet och trovärdighet enligt nedan:

DEF: Säg att det föreligger en utfästelse om q, U(q). En utfästelsekedja bakom U(q) sägs fungera "tillförlitligt" om och endast om i varje tillstånd där *denna* utfästelsekedja föreligger, föreligger också q.

DEF: Låt A vara en person eller institution som i situationen s utfärdar utfästelsen q som mottas av B. A:s utfästelse av q i s, U(q)

skall sägas vara "trovärdig" för B om och endast om B har evidens för att det finns en tillförlitligt fungerande utfästelsekedja bakom U(q).

Som framgår finns en stark symmetri mellan trovärdighet/tillförlitlighet hos information och hos utfästelser. Det finns en skillnad i tempus, men av skäl jag anfört tycks tidsskillnaden vara irrelevant.

7. Institutionella villkor för informations- och beslutshantering

Symmetrin mellan de två teorierna om trovärdighet ger verktyg för att analysera institutioner. Halldén (1980) anger en del praktiska tumregler för nyhetshantering som kan återföras till matematiska egenskaper hos sammansatt evidens. Multiplikationseffekter talar för att man bör ha korta informationskedjor med hög evidens för att varje länk korrekt överför information. Additionseffekter talar för att man hamnar i en bättre situation om man har oberoende informationskedjor.

Motsvarande tumregler gäller för utfästelser. Mängden av led mellan beslutsfattare och verkställare i en byråkratisk pyramid tenderar att dra ner trovärdigheten av utfästelser. Om det mellan den centrala myndigheten och verkställandet finns motsträviga instanser ute i provinsen, kan den totala sannolikheten för att utfästelsekedjan leder till verkställan inte överstiga sannolikheten för att den svagaste länken kommer att agera lojalt. Ett företag kan öka trovärdigheten av sina utfästelser genom att förlita sig på underleverantörer som är oberoende med avseende på råvaror och avtal med arbetskraften.

Trovärdighet beror inte av den evidens man själv har utan av den evidens mottagaren har. I "4.15 från Paddington" låter Agatha Christie sin hjältinna Mrs McGillicuddy dåsa till i kupén när tåget lämnar stationen. Ett annat tåg kör parallellt med hennes tåg, en gardin flyger upp och Mrs McGillicuddy ser en man strypa en kvinna på det andra tåget. Hon tillkallar konduktören. När denne kommer in i kupén ser han en gammal dam som i upprivet tillstånd berättar om ett mord hon sett. Konduktören noterar att på sätet där den gamla damen dåsat ligger en tidning uppslagen. Den visar en bild där en man stryper en kvinna. Konduktören är misstrogen mot damens uppgifter, i synnerhet som hon medger att hon dåsat till. Trots att Mrs McGillicuddy (och läsaren) vet att ett mord har begåtts kan hon inte överföra sin kunskap till konduktören. Utan trovärdighet blir hon en epistemisk ändhållplats.

Trovärdighet av information beror ofta av insyn i den process som är upphov till informationen. När man tillsätter en utredning om missförhållanden inom en samhällssektor, kan det vara exempelvis vara viktigt att utredningen medger insyn från dem man vill övertyga.

Trovärdighet av utfästelser beror av andras evidens. Är svensk neutralitetspolitik trovärdig? Det är möjligt att politiken är starkt trovärdig i förhållande till den evidens vi själva besitter om vår orubbliga sinnesförfattning och fasta beslutsamhet. Mer intressant är kanske frågan om neutralitetspolitiken är trovärdig i förhållande till den evidens som ryska bedömare kan besitta. Dels kan de rimligen inte veta hur orubbliga vi är i våra själar. Dels kan det hända att ryska principer för att vikta evidens mer ser till yttre faktorer: den svenska neutralitetspolitiken under andra världskriget, långvarigt svenskt samarbete med NATO på olika nivåer, ökade svenska åtaganden i Baltikum, minskade försvarsanslag och inrikespolitisk oenighet huruvida svenskt försvar har adekvat styrka för att säkra utrikespolitiska mål i en kärvare utrikespolitisk miljö.

Det är lättare att vara trovärdig inför sig själv än inför andra. En analys av vilken evidens andra använder för att bedöma ens trovärdighet kan ha relevans för beslut och öka den egna förståelsen av omgivningens reaktioner.

8. Vad är poängen med trovärdighet?

Vittnespsykologi och källkritik handlar om vad slags utsagor som får utgöra underlag för slutledningar. Detta ger en ledtråd till poängen med trovärdighetsbegreppet. Jag är böjd att tro att trovärdighet ofta är ett redskap för grovsållning av det brus av information och erbjudanden som man möter i samhället. Det som saknar trovärdighet hamnar under den nivå där signalerna har betydelse för vad vi sätter lit till.

Vad är det en person har förlorat när han förlorat sin trovärdighet? Fabeln om pojken och vargen vill varna oss för att förbruka vår trovärdighet. Den som förbrukat sin trovärdighet kan inte längre informera rationella åhörare. Den affärsman eller politiker som förlorat sin trovärdighet kan inte längre träffa bindande avtal med rationella parter.

Fabeln om pojken och vargen tycks antyda att trovärdighet är en egenskap hos en person som denne genom egen förtjänst kan vinna eller förlora. Så är i allmänhet inte fallet. Trovärdighet kan man förlora genom strukturella sociala förändringar av de kedjor som

binder information och utfästelser. Om dagens finansminstrar är mindre trovärdiga än finansminstrarna på 1950-talet, så beror detta av att positionen förlorat många bindningsmöjligheter, inte av personliga brister hos dess innehavare.

Trovärdighet är inte monadisk egenskap utan ytterst en flerställig relation. Bristande trovärdighet beror alltid på relationer och endast ibland på personliga egenskaper. Den som hamnat i fel slags relationer är, eller uppfattas som, en del av det otillförlitliga brus som inte informerar och som inget åstadkommer.

Jag vill tacka Martin Edman, Nils-Eric Sahlin, Wlodek Rabinowicz och deltagarna i filosofiseminariet i Lund för synpunkter och frågor som undanröjt en del av mina tidigare grumligheter.

Litteratur

COLEMAN, J S (1990): *Foundations of Social Theory*, Cambridge, Mass, The Belknap Press.

EDMAN, M (1973): "Adding independent pieces of evidence", i *Modality, Morality, and Other Problems of Sense and Nonsense: Essays dedicated to Sören Halldén*, Lund.

EKELÖF, P O (1964): "Free Evaluation of Evidence", *Scandinavian Studies in Law.*

ELSTER, J (1984): *Ulysses and the Sirens*, Cambridge University Press.

HALLDÉN, S (1973): "Indiciemekanismer", *Tidsskrift for rettsvitenskap.*

HALLDÉN, S (1980): *Nyfikenhetens redskap*, Lund, Studentlitteratur.

MOLANDER, P (1994): *Akvedukten vid Zaghouran*, Stockholm, Atlantis.

PERSSON, T (1987): *Credibility of Macroeconomic Policy: An Introduction and a Broad Survey*, Stockholm.

PUTNAM, R D (1993): *Making Democracy Work*, Princeton University Press.

ROLF, B (1983): *Kunskapskritik och kunskapsgenetik,* Lund.

SAHLIN, N-E (1986): "How to be 100% certain 99,5% of the time", *The Journal of Philosophy.* Ursprungligen publicerad som kompendium 1983.

SCHELLING, T (1963): *The Strategy of Conflict*, Cambridge, Mass, Harvard University Press.