

TIDSKRIFT FÖR POLITISK FILOSOFI  
NR 1 2017 | ÅRGÅNG 21

*Bokförlaget* THALES

RUSSELL HARDIN, *Trust & Trustworthiness*,  
New York: Russel Sage Foundation, 2002

Trust & Trustworthiness – En bok med oförtjänt gott rykte  
NYLIGEN NÄR JAG började läsa in mig på den teoretiska sidan av  
tillitsforskningen fick jag höra av både filosofer och statsvetare att  
jag borde läsa Russell Hardins magnum opus *Trust and Trustworthiness*  
(2002). Det är en späckad bok som avhandlar så väl tillit  
(»trust») och trovärdighet (»trustworthiness») som misstro  
(»distrust») och »tilltro» (»confidence»). Dessutom diskuteras  
både hur dessa tillstånd uppkommer, varför de är bra, dåliga el-  
ler något däremellan. Boken har till dags dato 1819 citeringar på  
Google Scholar vilket gör den relevant att diskutera i en lite längre  
litteraturgranskning som denna eftersom det enligt min vetenskap  
inte har gjorts i svenskt format tidigare.

Det jag kommer att diskutera är en av de mer intressanta teserna  
som drivs i boken vilken är att medborgare inte kan ha tillit till  
storskaliga institutioner som en stat i ett modernt samhälle. Argu-  
mentet byggs upp i kapitel 1 till 3 där »tillit» och »trovärdighet»  
definieras och där en särskild variant av »tillit» försvaras (»the en-  
capsulated interest account»). Försvaret av Hardins definition av  
tillit bygger till viss del på en uttalad idé om en särskild uppsättning  
adekvansvillkor som han menar gäller för en god definition.

Jag kommer att diskutera Hardins definition och adekvansvillkor  
i tur och ordning. Min slutsats är att Hardins tes inte håller samt att  
om man är ute efter begreppslig klarhet och en välgrundad uppfatt-  
ning om vad tillit är, bör man nog läsa något annat. *Trust & Trust-*  
*worthiness* är helt enkelt *betydligt* sämre än sitt rykte, detta oavsett  
vad Robert Putnam säger i en blurb på baksidan av boken.

Tillit

HARDIN ARGUMENTERAR ALLTSÅ för att vi inte kan hysa till-

lit till storskaliga institutioner som den moderna staten. Enligt honom litar A på B med avseende C när A upplever B som trovärdig. B är trovärdig när denne »omsluter A:s intressen» för att denne vill ha en fortsatt relation till A (se till exempel Hardin 2002: 1, 3, 4, 6, 11, 13, 24–27). Viktigt här är att det är just A:s intressen som ska ha inkorporerats och inte de intressen A representerar i kraft av sin grupptillhörighet. Till exempel, för att jag som svensk medborgare ska lita på en tjänsteperson i en svensk statlig myndighet krävs det att jag tror att denna tjänsteperson har inkapslat just *mina* intressen och inte *svenska medborgares* intressen generellt.

Allt detta leder till att vi inte kan ha tillit till (större) statliga organisationer. Få medborgare har kunskap nog, enligt Hardin, att tro att de enskilda personerna i staten handlar just med deras intresse för ögonen, det vill säga att B inkorporerat just A:s intressen (Hardin 2002: 153, 156).<sup>1</sup> Medborgare kan i bästa fall ha *tilltro* till staten vilket bygger på en upplevelse av att staten levererar vad den ska.<sup>2</sup> Denna tilltro baseras i sin tur på induktiva resonemang från tidigare erfarenheter (Hardin 2002: 156–159).

Ett problem med Hardins idé om tillit är att den verkar vara bristfällig och oförenlig med hur många av oss tänker kring detta med tillit. Jag brukar till exempel ofta arbeta på ett kafé nära min bostad. Eftersom jag bor mitt i staden och kaféet tillhör en större kedja känner jag sällan igen de som sitter där. När jag arbetar behöver jag ibland gå på toaletten och då brukar jag lämna datorn vid min plats. Ibland frågar jag någon om de kan passa den medan jag är borta. När jag kommer tillbaka har den (hittills) alltid varit kvar. Men tänk nu om jag kom tillbaka och såg att min dator var borta och personen som hade i »uppdrag» att passa den sa att hon struntat i det eftersom hon vid lite närmare eftertanke tyckt att det var lite jobbigt att behöva ha koll på den. I detta läge verkar det inte särskilt konstigt om jag känner mig arg och besviken och att jag till och med säger till personen »men jag litade ju på dig!».<sup>3</sup> Poängen är att vi verkar tänka oss att vi kan hysa tillit till personer som vi inte känner och inte har någon relation till samt att det är naturligt att reagera med reaktiva attityder när någon bryter mot vår tillit, vilket är oförenligt med Hardins idé.

Det kan vara värt att notera lite extra att Hardins definition inte alls lägger någon vikt vid den viktiga skillnad som finns mellan att *förlita sig* (»rely») på någon eller något och att ha *tillit* till någon. När vi litar på någon är vi alltså benägna att reagera med *ilska* och känna oss svikna när någon bryter mot vår *tillit* (se till exempel Bailer 1986 och Holden 1994). Däremot känner vi oss endast *besvikna* när någon eller något sviker oss när vi *förlitat* oss på denne eller denna. Om vi tror att någon annan hade inkapslat våra intressen för att denne ville ha en pågående relation till oss och detta senare visade sig inte vara fallet, skulle det inte vara särskilt märkligt om vi blev just besvikna snarare än arga. Detta är lätt att se om vi ändrar mitt kaféexempel ovan så att personen som vaktar min dator är någon som också brukar arbeta på kaféet och som jag har en ytlig men ändå någon sorts relation till. Vi brukar växla några ord och passa varandras datorer, men jag är klar över att denne enbart gör detta för att den upplever att den har något att vinna på det. Så fort den inte upplever sig ha det längre kommer den att avsluta sitt samarbete med mig. I detta fall skulle jag antagligen bli just besviken den dag då min dator vara borta och den som var satt att vakta den likaså. Den moraliska ilskan från det första exemplet skulle alltså inte vara lika närvarande i detta fall trots att vi här och inte i original exemplet, enligt Hardins definition, har att göra med ett brott mot vår tillit.<sup>4</sup> Att vi inte skulle reagera med reaktiva attityder när någon bryter mot vår tillit som den definieras av Hardin talar emot hans definition.<sup>5</sup>

Ett svar skulle kanske kunna vara att det jag just beskrev är en instans av kvasi-tillit (Hardin 2002: 61) och inte tillit. Så skulle det såklart kunna vara men om vi antar att det finns goda skäl att tro att vi har en benägenhet att reagera med reaktiva attityder vid ett brott mot vår tillit så har vi nog skäl att tro det omvända, att Hardins definition är en av kvasi-tillit och inte något annat.

Nu finns det dock en oklarhet rörande vad Hardin egentligen diskuterar. Han diskuterar till exempel »tillit» och att »ha skäl till tillit» på ett högst utbytbar sätt vilket är problematiskt men som också kan rädda delar av hans resonemang. Det är problematiskt

att använda »tillit» och »skäl till tillit» på ett utbytbar sätt av det enkla skälet att de inte är utbytbara. Bara för att någon presenterar vad ett *gott skäl* för att lita på någon skulle kunna vara så presenterar den inte någon definition av vad tillit *är* för någonting. Jämför till exempel med när vi tänker att vi har goda skäl att vara lyckliga. Det kanske är för att vi har nära vänner, ett bra jobb och en underbar familj. Att vi under dessa omständigheter har skäl att vara lyckliga ger oss inte en förklaring av vad »lycka» är, precis som vi inte får en definition av vad »tillit» är när någon säger att vi har skäl att lita på någon.<sup>6</sup> Däremot kan det förstås vara så att jag har ett bra skäl att lita på någon om jag tror att de omsluter mina intressen. Då har jag skäl att tro att B har egenintresseskal att C vilka brukar vara hyggligt reliabla när det kommer till att förutse mänskligt handlande.<sup>7</sup> (Dock kan detta såklart också vara en källa till misstro. Givet att jag vet, som ovan, att personen kommer att svika mig så fort denne inte omsluter mina intressen av skälet att den vill ha en fortsatt relation till mig så kan man tänka sig att jag skulle vara orolig för att den dagen skulle komma utan att jag märkte det vilket i sin tur skulle kunna leda till misstro).

Om vi antar att vi har goda skäl att lita på någon när denne omsluter våra intressen skulle vi kanske kunna kalla detta för »rationell tillit». Problemet är att vi då återigen inte vet vad tillit är för någonting. Det är ungefär som att säga att autentisk lycka är sådana lyckoupplevelser som kommer ur trosföreställningar som är sanna utan att tidigare säga något om vad »lycka» är och ändå påstå att man gett en fullständig definition av autentisk lycka. Tillit är inte heller, enligt Hardin, grundläggande i meningen att det är odefinierbart (Hardin 2002: 56–58), vilket gör att även denna utväg är stängd.

Alltså, det är oklart om vi får en definition av tillit och i den mån vi får det är den kontraintuitiv.

Adekvansvillkoren

HITTILLS HAR JAG argumenterat för att Hardins definition av tillit bland annat är kontraintuitiv och att detta är ett argument mot

denna definition, vilket i sin tur gör att hans argument för att vi inte kan ha tillit till storskaliga statliga institutioner fallerar. Ett argument mot min invändning är att de syften Hardin har med sin begreppsutredning gör dessa semantiska intuitioner irrelevanta. Denna invändning kommer jag att diskutera här.

Enligt Hardin är resultatet av en god begreppsutredning, om jag förstått saken rätt, att man får en definition som »passar in» bland de fenomen som man vill diskutera, att det är precist och smalt nog så att man kan dra ut systematiska implikationer från definitionen inom en rad områden, och att dessa implikationer ska vara handlingar som förklaras av den definition vi är satta att analysera (Hardin 2002: 1, 7–8, 11, 54, 87–88). Avslutningsvis, när vi jämför olika idéers relativa förtjänster så är en idé bättre om den uppfyller de olika villkoren och samtidigt förklarar fler handlingar än dess konkurrent.

En tolkning av detta är att metoden för att åstadkomma en definition som uppfyller dessa villkor *inte* är att undersöka termernas semantiska eller mer substantiella betydelser genom olika tankeexperiment eftersom det, enligt Hardin, ofta finns flera betydelser av en och samma term och inte något bra sätt att komma fram till vilken betydelse som är den rätta (Hardin 2002: xx). Den funktion dessa exempel har är när vi ska förklara vad våra definitioner säger mer precis. Då bör vi emellertid använda oss av fylliga och verklighetsnära exempel snarare än de stiliserade fall som ofta används i moralfilosofin (Hardin 2002: xx–xxi). Vidare ondgör sig Hardin över alla som i Platonsk mening försöker förstå vad tillit *egentligen* är (Hardin 2002: xx).<sup>8</sup> En tolkning av detta är att han menar att tillit inte är en naturlig sort som till exempel H<sub>2</sub>O, utan är konstruerat av människan. Vad gäller definitioner av sociala konstruktioner kan man tänka sig att dessa inte är sanna eller falska utan att dessa snarare är mer eller mindre lämpliga beroende på vilka syften man har (se till exempel Brülde 2003).

Argumentet för Hardins definition blir då att den förklarar fler fenomen än sina konkurrenter rörande det vi vill förklara och är tydligare än dessa konkurrenter, vilket ger oss skäl att använda den-

na definition snarare än de andra.<sup>9</sup> Han argumenterar till exempel för att hans idé om tillit är central för de fall som ofta diskuteras i tillitslitteraturen: envägstillit, ömsesidig tillit som i fångarnas dilemma, och tillitsfulla interaktioner i »tjocka» förhållanden (Hardin 2002: 3, 14–23). Mitt motexempel mot Hardins definition fungerar inte eftersom denna sorts exempel helt enkelt inte är relevanta när vi bedömer hållbarheten och relevansen hos en definition.

Nu finns det åtminstone två problem med denna tolkning. För det första, om man vill mäta tillit och inte bara undersöka hur man till exempel kan lösa kollektiva dilemman så måste vi utgå ifrån en definition som har sin grund i vad vi till vardags, möjligen vid lite eftertanke, menar med tillit. Självklart måste vi antagligen precisera vad som menas och kan menas med tillit eftersom detta kan vara oklart, men detta är nog svårt att göra oberoende av exempel av olika slag som visa på implikationerna av de definitioner man finner.

Vidare verkar det vara problematiskt att bara stipulera en definition på basis av någon specifik vardagsspråklig betydelse man finner, istället för att försöka skapa en mer heltäckande definition. Ett skäl till varför vi intresserar oss för olika fenomen är för att de upplevs som viktiga på olika sätt och om vi då går ifrån den vardagsspråkliga betydelsen alltför mycket riskerar vi att definiera bort det vi anser vara viktigt.<sup>10</sup> Självklart kan vi inte alltid mäta allt, och precision, mätbarhet och tydlighet är andra viktiga adekvansvillkor. Men om vi på den definitoriska nivån nöjer oss med att välja bara den ena aspekten av vad vi menar med lycka då har vi inte ens möjlighet att undersöka de andra aspekterna vid ett annat tillfälle. Samma sak gäller förstås när vi definierar tillit. Vidare blir resultaten av ett argument, som att vi inte kan lite på storskaliga institutioner, betydligt mindre intressant om man på förhand definierar bort många av de aspekter som vi normalt tror gör att vi kan ha en sådan tillit.

För det andra är idén om olika exempels irrelevans för hållbarheten hos en definition inte förenlig med det sätt Hardin arbetar på. För att till exempel visa att »tillit» och »överlåtande» (»entrusting») inte är samma sak använder han exemplet att vi kan bli

tvingade att överlåta något till någon även om vi inte litar på dem (Hardin 2002: 65–66). När han argumenterar emot idén om att vi kan lita på andra utan grund, som när ett spädbarn litar på sina föräldrar, argumenterar han emot detta med hjälp av en rad exempel som ska få oss att tro att vi måste vara kognitivt utvecklade för att ha tillit »på riktigt» (Hardin 2002: 69–73). När han försöker visa att idén om att vi litar mer på människor när vi tror att det finns en lägre sannolikhet för att de ska utföra det vi litar på dem att utföra så säger han att vi i så fall inte skulle lita på våra föräldrar, vilket förstås i många fall upplevs som orimligt (Hardin 2002: 73–74). Detta bara för att nämna några exempel.

Hardin skulle kanske istället invända att mitt exempel inte skulle vara relevant är för att det är för abstrakt eller dylikt och därmed inte når upp till de krav som Hardins adekvansvillkor ställer. Problemet är att detta inte heller verkar vara ett hållbart argument. Mitt kaféfall verkar inte vara relevant annorlunda än de fall han själv använder. Därmed borde mitt fall vara lika framgångsrikt som de fall Hardin använder fast omvänt; det ger oss skäl att tro att hans definition av tillit inte är hållbar.

Det är inte heller ett perifert fall. Många av oss lämnar varje dag våra dyrbara ägodelar obevakade bland en massa för oss okända människor och någon idé om generaliserad tillit skulle möjligen kunna förklara varför. För egen del skulle det kanske inte kosta så mycket pengar om min dator blev stulen. Min arbetsgivare skulle köpa en ny utan att jag behövde tänka på det. Men jag skulle bli av med veckor av arbete (om det inte är sparat i något moln eller dylikt) vilket i sin tur skulle vara oerhört smärtsamt för mig att försöka återskapa. Att jag ändå tar risken att lämna datorn vid min plats istället för att ta med den, vilket inte skulle vara särskilt jobbigt, kan förklaras av att jag litar på att människor generellt inte stjälar mina saker. Eftersom detta är ett sådant vanligt beteende och det åtminstone på ytan verkar vara en produkt av tillit så som vi normalt förstår det borde en teori om vad tillit är kunna förklara detta handlande.

Avslutningsvis ska det tilläggas att Hardin vid flera tillfällen

går med på att hans teori inte är komplett (se till exempel Hardin 2002: 6, 52) vilket kan ge hans idé ett visst skydd mot kritik som den jag lanserat i detta papper. Problemet är att det är svårt att se hur han ska kunna svara på denna kritik utan att genomföra ganska substantiella förändringar av sin teori. Det räcker med andra ord inte med att komplettera, den måste göras om, vilket gör att talet om att teorin inte är komplett antagligen inte hjälper.

Slutsatsen blir att om Hardins adekvansvillkor enligt den första tolkningen varit tillräckliga, så hade hans tes om att vi inte kan lita på storskaliga statliga institutioner hållit. Men eftersom adekvansvillkoren behöver tolkas på annat sätt för att vara rimliga och förenliga med vad Hardin själv gör så faller tesen. Självklart ska man kunna vara revisionistisk i förhållande till vardagsspråket men det verkar märkligt, givet att det man gör är tänkt att vara intressant och relevant i till exempel policysammanhang, att vara revisionistisk till den grad att så många potentiellt viktiga aspekter faller bort bara för att man vill ha en tydlig och enhetlig definition. Detta utan att göra ordentliga försök att åtgärda eller svara på dessa uppenbara problem.

Avslutande kommentarer

TRUST & TRUSTWORTHINESS är sedan länge en stapelvara i tillitslitteraturen. Jag har tagit upp några problematiska aspekter av denna bok men liknande problem finns genomgående i hela verket där det största problemet är att det många gånger är svårt att vet vad Hardin egentligen diskuterar. Vi borde därför sluta rekommendera varandra att läsa denna och istället börja rekommendera andra framstående författare inom området så som till exempel Karen Jones, Katherine Hawley och Paula Hieronymi. Det dessa skriver är ett under av klarhet i jämförelse och säger det som behöver sägas med betydligt färre ord. Precis som man vill ha det när man läser något för att ha det i arbetet och inte enbart för sitt eget höga nöjes skull.

→

*Karl de Fine Licht*

## Noter

1 Det verkar dessutom högst orimligt att tro att tjänstepersoner i myndigheternas tjänst har tillräcklig kunskap eller intresse av enskilda medborgares intressen, vilket i sin tur gör att även om medborgarna hade kunskap om tjänstepersonernas motivation skulle de inte lita på dessa. En lösning skulle kanske kunna vara att se institutionerna som innehavande av ett kollektivt agentskap. Jag kommer inte att gå in på det här eftersom jag tänker mig att det finns andra större brister i Hardins resonemang.

2 Ett annat ord för tilltro som Hardin använder är »kvasi-tillit» (Hardin 2002: 156–159). Vi kan inte heller lita på människor generellt (»generaliserad tillit») eftersom vi inte känner alla dessa eller har kontakt med dem i någon relevant mening. Men inte heller detta är särskilt problematiskt eftersom generaliserad tillit till andra människor inte kan växlas ut i till exempel »socialt kapital» på det sätt man inom tillitsforskningen ofta tänker sig (Hardin 2002: 83–84). (Socialt kapital brukar betraktas som nödvändigt, eller åtminstone väldigt viktigt, för att staten ska fungera väl).

3 Det fall jag beskriver här skulle många kalla ett fall av »generaliserad tillit». Hardin argumenterar på en del ställen mot att detta i någon intressantare mening finns: om vi skulle göra frågorna mer specifika, till exempel ifall vi var villiga att låna ut hundra dollar till kompletta främlingar, skulle resultaten i dessa studier bli extremt nedslående i jämförelse med hur de ser ut idag (se till exempel Hardin 2002: 61). Jag håller med om att detta låter rimligt men det verkar också rimligt att vi faktiskt väldigt ofta gör något liknande om vi litar på människor. Vi lämnar exempelvis våra väskor utan uppsikt på tåget eller betalar för produkter vi köper på nätet innan vi får hem dem. Att dessa beteenden där vi riskerar en hel del kan förklaras med att vi »litar på människor» verkar inte vara en helt märklig tanke i sammanhanget.

4 Vi kanske får anta fler saker som att personen inte är ett lämpligt objekt för reaktiva attityder överhuvud. (Den kanske är en psykopat som inte alls förstår sig på »detta med moral»). Det är ju trots allt inte enbart så att vi klandrar människor för att de bryter mot vår tillit, vi gör det ju också för att de utför handlingar som är moraliskt fel vilket vi kan anta är fallet när någon stjälar en dator.

5 Hardin säger: »*Betrayal is of course not a failure of trust but a failure of trustworthiness*» (Hardin 2002: 28). Samma poäng kan dock göras vad gäller hans definition av trovärdighet eftersom det bara är att omsluta den andres intressen på rätt sätt.

6 Däremot ger han något som mer liknar en definition av tillit när han säger att tillit är kognitiv till sin natur och att den inbegriper ett visst mått av risktagande (Hardin 2002: 7).

7 Hardin skriver vidare när han försöker förklara vad tillit är att vi behöver inkapsla varandras intressen för att kunna ha »tillitsfulla relationer». Nu kan detta vara sant men det verkar fortfarande inte ge oss en förklaring av vad tillit är. Det verkar trots allt vara så att vi kan lita på personer som vi inte har en tillitsfull relation till. Jämför här med kaféfallet ovan. Man kan förstås vara lockad av att tänka att det enda Hardin skriver om är tillitsfulla relationer och inte tillit (se till exempel Hardin 2002: xx–xxi, 1, 87). Men då är det oerhört märkligt att han döper kapitlet som ska handla om begreppsliga frågor till tillit (»Trust») och att han väldigt ofta säger att han ska definiera vad tillit är.

8 Vid andra tillfällen talar Hardin om sitt angreppssätt som en motsats till dem som

använder sig av »fundamentistisk epistemologi» (Hardin 2002: 54f). Dessa, menar Hardin, vet vad tillit är genom sina intuitioner och eftersom dessa intuitioner är högst personliga, påverkade av en rad störande kausala faktorer och endast leder till oklara definitioner så blir dessa definitioner av tillit ej testbara, vilket i sin tur blir ett problem för samhällsvetenskapen. Jag håller självklart med Hardin om att det är ett problem att till exempel tillit diskuteras på detta sätt, men jag är också säker på att fundamentism inte nödvändigtvis dras med denna sorts problem. Många fundamentistiska filosofer, som till exempel Jeff McMahan, producerar extremt pregnanta teorier, vilka väldigt många andra anser vara rimliga och som på olika sätt är testbara.

9 Jag har utgått ifrån att Hardins projekt är rent begreppsligt vilket kan vara i konflikt med standardtolkningen som verkar vara att den är ontologisk (McLeod 2015). Hardin säger saker som ibland kan tolkas ontologiskt som till exempel att »[according to] foundationalist epistemology [...] they know it when they see it» (Hardin 2002: 54). Men denna skillnad i tolkning borde inte ha några substantiella implikationer för mina argument.

10 Inom psykologin och värdeteorin brukar man till exempel tala om lyckobegreppets två delar: den affektiva (den känslomässiga upplevelsen) och den kognitiva (hur vi värderar vårt liv på det stora hela). Tänk att vi väljer att *definiera* lycka så att vi enbart mäter den affektiva delen och inte den kognitiva delen. Då missar vi förstås en substantiell del av vad många menar med lycka. Därmed försvinner en stor del av undersökningens relevans och resultaten blir svårare att använda i exempelvis policysammanhang eftersom vi inte kan säga om det vi mäter är vad folk sätter värde på.

## Referenser

- BAIER, A. (1986) »Trust and antitrust», *Ethics*, 96(2), ss. 231–260.
- BRÜLDE, B. (2003) »Hälsa och sjukdom – En begreppslig utredning», Lund: Studentlitteratur.
- HARDIN, R. (2002) *Trust & Trustworthiness*, New York: Russel Sage Foundation.
- HOLTON, R. (1994) »Deciding to trust, coming to believe», *Australasian Journal of Philosophy*, 72(1), ss. 63–76.
- MCLEOD, C. (2015) »Trust», i: E.N. Zalta (red.) *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/fall2015/entries/trust/>