Giangiacomo Bravo
Teoriprovning och kausal slutledning

ur boken

Bengt Larsson, Ylva Ulfsdotter Eriksson & Ola Agevall (red.)

Teoretiskt hantverk

Om teoretiserande i samhällsvetenskaplig forskning

Arkiv förlag 2025

FÖRSLAG PÅ KÄLLANGIVELSE:
Bravo, Giangiacomo (2025) ”Teoriprovning och kausal slutledning”, i Bengt Larsson, Ylva Ulfsdotter Eriksson & Ola Agevall (red.), *Teoretiskt hantverk. Om teoretiserande i samhällsvetenskaplig forskning*, s. 165–186, Lund: Arkiv förlag, https://doi.org/10.13068/9789179244033.

Det här kapitlet ur en e-bok från Arkiv förlag distribueras fritt över internet genom *open access*. Titeln finns också tillgänglig i tryckt utgåva med ISBN: 978 91 7924 402 6.

Arkiv förlag · arkiv@arkiv.nu · www.arkiv.nu

E-bokutgåva (PDF) 2025
Beständig länk: https://doi.org/10.13068/9789179244033
ISBN: 978 91 7924 403 3

Teoriprovning och kausal slutledning

GIANGIACOMO BRAVO

De flesta är överens om att vetenskaplig forskning innebär att vi prövar våra teorier om samhället och det mänskliga beteendet så rigoröst som möjligt. Men hur gör man det? Hur kan vi tänka om teoritestning och kausal inferens och vilka processer är inblandade när vi omsätter detta tänkesätt i praktiska undersökningar? I detta kapitel definieras teori på ett mer restriktivt sätt än vad som kanske är vanligt inom sociologi, nämligen som en uppsättning allmänna påståenden som etablerar empiriskt testbara relationer mellan observerbara variabler, där dessa relationer behöver testas empiriskt och eventuellt bekräftas innan de blir användbara. Definitionen har valts av två huvudsakliga skäl. (1) Detta är den standardmässiga betydelsen av ordet teori inom vetenskapen, där det finns en tydlig standard om att omsätta teoretiska idéer i testbara hypoteser och testa dem innan man ens använder ordet teori i sig. En vetenskaplig teori bör därmed inte bara uppvisa stark inre konsistens utan också överensstämma med ett betydande antal empiriska observationer. (2) Endast genom att använda denna teoridefinition blir det möjligt att skapa en tydlig koppling mellan teoretiska påståenden och en rigorös testning av dem mot empiriska data (Raub m.fl. 2022), vilket är en huvudpoäng med kapitlet.

Att testa en teori innebär kort sagt att använda den för att härleda förutsägelser om (1) förekomsten eller avsaknaden av en förbindelse mellan vissa empiriskt mätbara variabler och (2) riktningen av orsakssambandet för denna förbindelse. Dessa förutsägelser, eller *hypoteser*, jämförs sedan med empiriska data som kan

stödja dem eller ej. Om data stödjer hypoteserna kan vi öka vår tilltro till teorins förmåga att förklara något i den verkliga världen. Om inte bör vi minska vår tilltro till den ursprungliga teorin och försöka att ändra den på något sätt. Om vi sedan fortfarande inte kan validera teorin empiriskt är det dags att ge upp och helt överge den (Cohen & Nagel 2002).

Även om det är omöjligt att helt avstå från statistik i allmänhet och vissa bestämda statistiska metoder i synnerhet är procedurerna som presenteras här ganska allmänna, så att man bara behöver relativt begränsade förkunskaper för att förstå dem. Målet är att steg för steg beskriva en fiktiv (och något fånig) påhittad forskning baserad på artificiellt genererade data. Valet att hitta på ett forskningsfall med relaterade data kan verka märkligt, men på så vis kan vi undvika den onödiga komplexiteten hos verkliga data och betrakta data från en neutral utgångspunkt, utan förutfattade meningar, vilket är en central aspekt av den vetenskapliga processen.

Samhällsteori och generering av testbara hypoteser

Samhällsteori använder ofta beskrivande och ibland ganska otydliga termer. Denna vaghet utgör i bästa fall ett hinder för noggrann testning och döljer i värsta fall triviala överväganden eller högst normativa (läs ideologiska) argument under invecklade och vilseledande formuleringar (Buchanan 2007; Watts 2011). Många sociologer är visserligen inte speciellt angelägna om bättre integration mellan teori och empirisk forskning, och de mest populära teorierna är ofta de som är svårast att direkt knyta till empirisk observation (Sørensen 2009). Men kritik mot otydlig och otestbar teoretisering har ändå framförts från olika håll: från analytisk sociologi (Hedström 2005; Hedström & Swedberg 1998), från *computational social science* (Keuschnigg m.fl. 2018; Watts 2014, 2017) och till och med från sociofysik (Buchanan 2007).

Förutsatt att det är viktigt att länka teorier till empiriska data är den huvudsakliga nackdelen med otydligt formulerade teorier att de förhindrar denna process. För att kunna göra en testning måste man nämligen ha något tydligt att jämföra verkligheten med. Med andra ord är det främsta kravet för att göra teorier testbara, och

därmed vetenskapliga, att tvinga dem att producera modeller som kan ge klara *prediktioner* om entydiga variabler som kan mätas empiriskt (Watts 2014).

Enligt Wilsons (Wilson & Kelling 1982) *broken windows theory* (BWT) är det till exempel så att tecken på oordning såsom trasiga fönster, skräp och graffiti skapar andra typer av oordning och slutligen kriminellt beteende. Oavsett om detta är sant eller ej (vilket är det som faktiskt ska testas) gör BTW en klar prediktion genom att uppställa en positiv korrelation mellan synliga tecken på oordning och kriminellt beteende med ett argument om den kausala riktningen. Det betyder inte att BTW är lätt att testa. De många potentiella faktorer som kan påverka sambandet mellan oordning och brott har utlöst en decennielång debatt om teorins empiriska giltighet (Maskaly & Boggess 2014), ofta driven av normativa och ideologiska ståndpunkter snarare än av vetenskaplig noggrannhet. Men huvudpoängen är att teorin *kan* testas med adekvata data och forskningsfärdigheter. Några bra experimentella tester av BTW har faktiskt genomförts under åren (Keizer m.fl. 2008). En klar förståelse av hur man gör testbara förutsägelser och hur man faktiskt testar dem bör således vara en kärnfråga för forskare som intresserar sig för att utveckla teori.

Även när en testbar prediktion görs och dess överensstämmelse med data empiriskt verifieras förblir debatten om den faktiska riktningen av orsakssambandet mellan de observerade variablerna ofta öppen. Använder man BWT som exempel igen, kan det mycket väl vara så att en plats med högre kriminalitet framstår som mer nedskräpad enbart därför att brottslingar har en tendens att nedprioritera det allmänna bästa. I så fall skulle orsakssambandet gå i motsatt riktning i förhållande till det som antas i BTW (något som är känt som *omvänd kausalitet*) men ändå vara fullt konsistent med vilken uppsättning data som helst som påvisar korrelation mellan till exempel nedskräpning och småbrott. Värre än så: kritiker av BTW argumenterar ofta för att både oordning och brott egentligen är produkten av andra, icke observerade variabler, såsom social och ekonomisk marginalisering (Maskaly & Boggess 2014). I så fall skulle den uppenbara relationen mellan oordning och brott vara en *spuriös korrelation*, det vill säga de två variablerna är korrelerade men inte kausalt relaterade.

Att testa kausala relationer är svårt med enbart statistiska metoder, och det vanliga sättet att lösa problemet inom vetenskapen är att genomföra experiment: artificiella situationer där forskare har kontroll över ingångsvariabler och observerar de motsvarande utfallen (Thye 2007). Experimentella metoder är för närvarande underanvända inom sociologi, vilket delvis också förklarar den bristande rigorositet som präglar en del av disciplinen (Gërxhani m.fl. 2022). Av praktiska eller etiska skäl finns det samtidigt tydliga begränsningar i deras tillämpningsmöjligheter. Dock är experiment inte den enda metoden för att bedöma kausalitet hos empiriska data, och mycket kan göras genom att man utnyttjar komplexiteten i relationen mellan de olika observerade variablerna (Pearl 2019). I avsnittet om kausal inferens kommer jag att diskutera fördelar och nackdelar med några befintliga alternativ till experiment.

Exologiskt mellanspel

För att illustrera processerna för teoritestning och kausal inferens kommer jag att använda data om en fiktiv situation. Användandet av fiktiva data har två skäl. (1) Jag vill ha full kunskap om och kontroll över de faktiska relationerna mellan variablerna. (2) Att använda en fiktiv situation är ett sätt att undvika oönskad påverkan av förutfattade meningar som kan förekommande vid betraktande av verkliga data. Jag kunde ha uttryckt dem i abstrakta termer, såsom Variabel X, Variabel Y och Variabel Z, men risken med det vore att läsare som inte gillar abstrakt matematik eller statistik, som faktiskt är de jag helst vill nå fram till, förlorade intresset.

Min kompromiss går ut på att skapa en liten berättelse och ett dataset för att testa våra hypoteser. Jag gör berättelsen så främmande som möjligt för att undvika okontrollerad inverkan av verkliga idéer på läsarens omdöme. Eventuella likheter med något läsaren vet om jorden och dess samhällen är således den oavsiktliga följden av min begränsade fantasi och jag är tacksam om man försöker ignorera dem. Data hänför sig till en avlägsen planet som kretsar kring en stjärna 124 ljusår bort och bara kallas vid sitt vetenskapliga namn: K2-18b. Allt är förstås påhittat utom existensen av planeten, som faktiskt har upptäckts och kanske till och med kan ha rätt förutsättningar för liv (Madhusudhan m.fl. 2020).

En grupp rymdantropologer, här kallade *exologer*, hade möjlighet att göra en kort resa till K2-18b där de upptäckte existensen av intelligenta utomjordingar: de var inte särskilt teknologiskt avancerade men hade ändå ett relativt komplext samhälle. Trots uppenbara kommunikationsproblem och begränsad vistelsetid lyckades forskarna göra några upptäckter och samla in vissa data. Tydligen identifierar sig de infödda i två typer, vilkas namn är omöjliga att uttala. Eftersom invånarna av den ena typen verkar vara lite längre än de av den andra typen, åtminstone i genomsnitt, kallar vi dem *tallies* och *shorties* (båda typerna är faktiskt ganska långa, ofta över 10 meter!). Dessutom verkar de infödda ha speciella matpreferenser: somliga av dem äter främst bambu, medan andra äter maneter. Det är oklart om matpreferenserna är kopplade till typerna, även om några av exologerna misstänkte det.

Under sin korta vistelse på K2-18b samlade forskarna in data om längd, typ och matvanor hos 1 000 infödda. De första raderna i det resulterande datasetet visas i tabell 1. Variablerna *height*, *type* och *eat* registrerar längden i meter, typen och matvanor hos varje observerad individ.

Tabell 1. Urval om åtta individer från K2-18b

height	*type*	*eat*	*tally*	*bamboo*
11,93	tally	manet	1	0
12,57	shorty	bambu	0	1
9,50	tally	manet	1	0
10,59	shorty	manet	0	0
11,48	tally	manet	1	0
9,39	shorty	bambu	0	1
9,51	shorty	bambu	0	1
10,25	shorty	manet	0	0

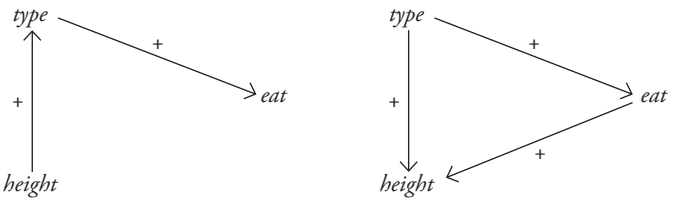
I de två sista kolumnerna i tabell 1 har jag helt enkelt omkodat variablerna *type* och *eat* till binära dummyvariabler för att förenkla några av de följande analyserna. En binär dummy är en variabel med 0 och 1 som enda möjliga värden, där 1 vanligtvis representerar förekomsten av en viss egenskap och 0 avsaknaden av densamma. I det här fallet betyder 1 i variabeln *tally* att värdet på variabeln *type* för

den individen är *tally*, medan 0 betyder att det är *shorty*. På samma sätt betyder 1 i *bamboo* att värdet på variabeln *eat* för den individen är bambu, medan 0 betyder att det är manet. Observera att dessa nya variabler exakt motsvarar de ursprungliga. Dock är den numeriska skalan matematiskt sett mer praktisk än den ursprungliga kategoriska skalan vid utförandet av vissa statistiska analyser.

Två modeller

På expeditionen till K2-18b deltog exologer med olika disciplinära inriktningar. Några av dem hade en bakgrund inom kulturantropologi och tenderade att tolka de observerade skillnaderna i empiriska observationer som socialt konstruerade. Andra hade en bakgrund inom evolutionär antropologi och tenderade att tolka de observerade skillnaderna som ett resultat av samspel mellan evolutionära anpassningar och utvecklingsmiljö hos varje individ. Dessa olika forskningsperspektiv medförde att de båda grupperna av forskare föreslog olika modeller för att förklara insamlade data (se figur 1). Observera att dessa enkla modeller inte återspeglar den faktiska komplexiteten inom antropologi. Precis som de fiktiva data som används här har de endast pedagogiska syften.

Figur 1. Modeller föreslagna av (A) kulturexologer och (B) evolutionära exologer som förklaring av data insamlade på K2-18b



A. Kulturexologernas modell *B. Evolutionära exologernas modell*

Kulturexologerna föreslog en modell enligt vilken typen är socialt konstruerad, det vill säga att barn som växer sig längre (respektive kortare) oftare betecknas som *tally* (respektive *shorty*) av de andra

infödda och antar vanor och beteenden hos den motsvarande gruppen. Dessutom postulerade forskaren att *tallies* oftare äter bambu, medan *shorties* oftare äter maneter enligt de kulturella konventionerna i deras respektive grupp (se figur 1A).

Observera att de små plussymbolerna i figuren uttrycker riktningen av den förväntade effekten, det vill säga hur värdena i den påverkade variabeln väntas förändras när värdena i de påverkande variablerna ökar: vid plus kommer de också att öka, vid minus kommer de att minska. Vad ökning (respektive minskning) betyder står klart om variabeln är numerisk, såsom *height*, men inte om den är kategorisk, såsom *type* och *eat*. I de senare fallen väljs en kategori godtyckligt som referens, och ökning betyder då helt enkelt växling från referenskategorin till den andra kategorin: från *shorty* till *tally* i *type* och från manet till bambu i *eat*. Observera att detta val också är konsekvent med omkodningen av variablerna till binära dummyvariabler *tally* och *bamboo*, där samma byte motsvarar en faktisk ökning från 0 till 1.

De evolutionära exologerna föreslog i stället en modell där genetiska egenskaper som är vanligare hos typen *tally* tenderar att få medlemmarna av denna grupp att växa sig längre. Samma egenskaper gör också att *tallies* utvecklar en preferens för att äta bambu. Å andra sidan gynnar bambu, som är näringsrikare än maneter, utvecklingen av längre kroppar. *Tallies* genetiska egenskaper har således både direkta och indirekta effekter på inföddas längd, så som visas i figur 1B.

För att underbygga sina argument tog de båda grupperna fram deskriptiv statistik. Kulturexologerna delade först upp infödingarna i två grupper beroende på om de var längre eller kortare än genomsnittet för urvalet. Sedan beräknade de andelen *tallies* och bambuätare för varje grupp. Eftersom båda siffrorna var högre för gruppen som låg över genomsnittet hävdar de att data stödjer deras modell (se tabell 2A nedan).

De evolutionära exologerna använde i stället den befintliga typvariabeln för att gruppera observationerna och beräkna genomsnittlig längd och andelen bambuätare per typ. Också de hävdar att data stödjer deras modell, även om andelen bambuätare endast är marginellt högre för *tallies* (se tabell 2B nedan).

Tabell 2. Deskriptiv statistik framtagen av (A) kulturexologer och (B) evolutionära exologer

	height-grupp	N	% *tally*	% bambuätare
A	Under medel	503	25,1	43,5
	Över medel	497	71,4	63,0

	type	N	Medel-*height*	% bambuätare
B	shorty	519	10,4	52,8
	tally	481	11,9	53,6

En livlig debatt utbryter mellan exologerna, där båda grupperna hävdar att data stödjer deras teori och där ingen vill ge efter. Problemet är att den deskriptiva statistiken i tabellerna 2A och 2B faktiskt är ganska konsistent med båda modellerna. Både kulturexologerna och de evolutionära exologerna använde strategin att dela upp observationen med utgångspunkt från deras rotvariabel (de variabler i figur 1 från vilka pilar endast går ut). Dock är statistiken ganska lika och den enda möjliga slutsatsen är att det verkar finnas någon icke slumpmässig relation mellan de tre variablerna i datasetet. Dessutom kan olika uppdelningar av data leda till något olika bilder av vad som händer. Att göra godtyckliga uppdelningar av data för att få det att se ut som om data stödjer ett visst argument är en faktisk strategi som används av många forskare, medvetet eller omedvetet. Det behöver knappast sägas att den inte leder till bra statistik eller till bra vetenskap.

Poängen här är att strukturen och tolkningen av den deskriptiva statistiken är ganska godtycklig. Deskriptiv statistik kan ge insikter om möjliga samband mellan data, men kan i de flesta fall inte användas för att testa en hypotes. Vad som behövs är mer robusta tester av modellerna, som ger tydliga indikationer om giltigheten av de föreslagna relationerna mellan variablerna utan att ge utrymme för potentiellt partisk tolkning av forskarna, vilket är hela poängen med formella metoder inom vetenskapen. Tyvärr är våra exologer bra på att skapa teorier men ganska dåliga på statistik. Vårt uppdrag blir att hjälpa dem med det genom att härleda testbara hypoteser och utföra adekvata statistiska tester för att avgöra vilken modell (om någon) som är förenlig med data.

Hypotesgenerering och testning

De två modellerna i figur 1 är bra på det sättet att de gör det lätt att härleda testbara hypoteser. Det enklaste sättet att göra detta är att bryta ner modellerna i de binära relationer som indikeras av pilarna och tydliggöra de motsvarande förutsägelserna. Till exempel antas i båda modellerna en positiv relation mellan *height* och *type*, men riktningen för orsakssambandet är olika: *height* → *type* för kulturmodellen och *type* → *height* för den evolutionära modellen. Detta resulterar i följande kontrasterande hypoteser:

HYPOTES 1 (kulturexologi): *infödda som är längre kallas oftare tallies*.

HYPOTES 2 (evolutionär exologi): *infödda av typen tally tenderar att växa sig längre*.

Om vi (för tillfället) bortser från orsakssambandets riktning är dessa två hypoteser konsekventa i det att de förutser ett positivt samband mellan *height* och *type*, det vill säga att de förutser att *tallies* i genomsnitt är längre än *shorties*. Eftersom denna relation innefattar en numerisk och en binär variabel representerar den en av de enklaste och vanligaste situationer där statistiska tester kan tillämpas. Vi ska därför börja med att samtidigt testa H1 och H2.

Kulturexologerna och de evolutionära exologerna är därtill överens om att *type* påverkar de infödda invånarnas matvanor. Närmare bestämt förväntas *tallies* enligt båda modellerna äta mer bambu än *shorties*. Detta blir vår tredje hypotes:

HYPOTES 3 (kulturexologi och evolutionär exologi): *tallies äter oftare bambu än shorties.*

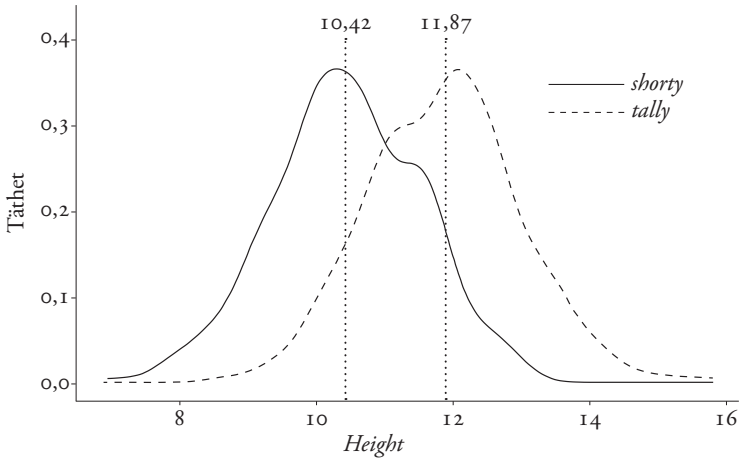
Till sist antas i kulturmodellen att det inte finns något direkt samband mellan *eat* och *height*, medan det i den evolutionära modellen föreslås att den förra variabeln positivt påverkar den senare (*eat* → *height*). Den resulterande hypotesen kan formuleras som:

HYPOTES 4 (evolutionär exologi): *infödda som oftare äter bambu tenderar att växa sig längre*.

Height *och* type *(H1 och H2)*

Båda modellerna antar ett positivt samband mellan *height* och *type*, men det betyder inte att varje enskild *tally* längre än varje enskild *shorty*, utan bara att genomsnittslängden för *tallies* större än för *shorties*. För att visuellt kontrollera detta visar figur 2 täthetsfördelningar av *height* efter *type*. En täthetsfördelning liknar ett stapeldiagram med ett mycket stort antal staplar, förutom att alla staplarna är omräknade så att arean under den resulterande kurvan alltid är lika med 1, vilket gör att grupper med olika storlekar blir lättare att jämföra.

Figur 2. Täthetsfördelningar av height *efter* type *med genomsnittslängden per grupp*



De två fördelningarna är ungefär normalfördelade (de liknar den välkända Gausskurvan) och genomsnittslängden för *tallies* ($\bar{h}^t$ = 11,87 m) verkar vara större än den för *shorties* ($\bar{h}^s$ = 10,42 m). Som förväntat finns det en stor överlappning mellan dem, men fördelningen för *tallies* verkar ligga något till höger om den för *shorties*. Vid första anblicken verkar *tallies* faktiskt vara lite längre än *shorties*.

Som alla statistiska handböcker kan upplysa om är problemet att vi inte vet om vi fick fram detta resultat bara för att vi hade tur i vår urvalsprocess eller om resultatet faktiskt gäller för hela K2-18b:s

befolkning. För att formellt kunna testa våra hypoteser behöver vi beräkna sannolikheten för att vi fick fram vårt resultat bara av en slump. Endast om denna sannolikhet är tillräckligt låg blir det möjligt att säga att data stödjer H1 och H2.

Det finns många statistiska metoder för att beräkna denna sannolikhet. Den mest grundläggande, som har den fördelen att den inte förutsätter några antaganden om datafördelningen, går ut på att ta många urval från samma population och räkna på i hur många av dem hypotesen håller. Tyvärr är en återresa till K2-18b utesluten för våra exologer, vilket innebär att metoden inte är utförbar. Dock är det möjligt att få en bra bild av hur flera urval skulle se ut genom att återskapa dem från själva data, en procedur som kallas *resampling* eller *bootstrapping*, på svenska ”återsampling”.

Det finns många återsamplingstekniker. Den enklaste, känd som *permutationstest* (se Box m.fl. 2005, kapitel 3), är lätt att förstå och kan ge en indikation om sannolikheten för att skillnaden mellan två medeltal i ett stickprov endast beror på slumpen. Här är stegen för att genomföra det:

1. beräkna skillnaden mellan de två gruppernas genomsnitt i det faktiska urvalet (i vårt fall $d_0 = \overline{h}^t - \overline{h}^s = 11{,}87 - 10{,}42 = 1{,}45$);

2. omstrukturera slumpmässigt längddata och beräkna skillnaden i det förändrade provet ($d_i = \overline{h}_i^t - \overline{h}_i^s$);

3. kontrollera om den nya skillnaden är större än den ursprungliga;

4. upprepa steg 2 och 3 ett rimligt antal gånger;

5. andelen tillfällen man finner $d_i \geq d_0$ av det totala antalet tillfällen man omstrukturerade data är sannolikheten för att man fick fram sitt resultat av en slump.

Idén bakom denna procedur är att man genom att upprepade gånger omstrukturera data kan få en bra bild av hur ett slumpmässigt utfall skulle kunna se ut med den empiriskt observerade längdfördelningen på planeten. Om det faktiska resultatet d_0 är större än de flesta av dessa slumpmässiga utfall är det osannolikt att man

bara hade tur i urvalsprocessen, vilket innebär att man slutligen kan säga att hypotesen stöds av data.

För att lära sig hur denna procedur fungerar i praktiken är det praktiskt att börja med ett mindre urval som kan behandlas med papper och penna (eller kanske ett excelark). Vi kan till exempel ta de åtta observationer som visas i tabell 1 och beräkna genomsnittslängden för *tallies* och *shorties*. I detta fall är $\overline{h}_0^t = 11{,}0$ och $\overline{h}_0^s = 10{,}5$, vilket innebär att $\overline{h}_0^t > \overline{h}_0^s$ (eller ekvivalent $d_0 = \overline{h}_0^t - \overline{h}_0^s = 0{,}5$). Vi kan nu försöka slumpmässigt omstrukturera värdena i kolumnen *height* tre gånger och beräkna skillnaden $d_i = \overline{h}_i^t - \overline{h}_i^s$ (med $i \in \{1, 2, 3\}$) för vart och ett av dessa försök. Resultaten visas i tabell 3, men jag rekommenderar att man prövar själv.

Tabell 3. Exempel på permutation av längdvärdena

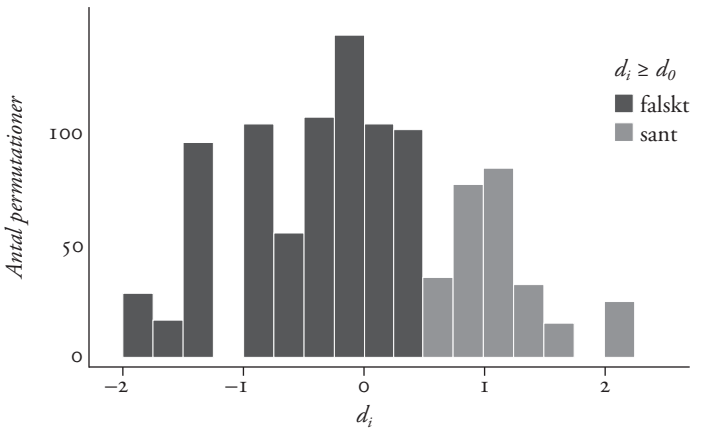
A		B		C	
height	*type*	*height*	*type*	*height*	*type*
10,25	tally	10,25	tally	12,57	tally
9,39	shorty	11,48	shorty	11,48	shorty
11,93	tally	11,93	tally	11,93	tally
12,57	shorty	9,39	shorty	10,59	shorty
9,51	tally	10,59	tally	9,50	tally
9,50	shorty	12,57	shorty	10,25	shorty
10,59	shorty	9,51	shorty	9,51	shorty
11,48	shorty	9,50	shorty	9,39	shorty
$\overline{h}_1^t - \overline{h}_1^s = -0{,}7$		$\overline{h}_2^t - \overline{h}_2^s = 0{,}4$		$\overline{h}_3^t - \overline{h}_3^s = 0{,}9$	

Endast ett av de tre resultaten i tabell 3 är större än vår ursprungliga skillnad $d_0 = 0{,}5$. Detta betyder att sannolikheten för att våra resultat är slumpmässiga är $p = \frac{1}{3} \approx 0{,}33$. Detta verkar svagt antyda att vi inte fick fram vår positiva skillnad av en slump, även om det är långt ifrån den konventionella signifikansnivån på 95 procent som innebär $p \leq 0{,}05$.

Observera att vi här bara har kontrollerat ett litet urval av alla möjliga permutationer av data. Med tanke på att vi använder åtta observationer skulle det faktiskt vara möjligt att beräkna skillnaden mellan de två genomsnitten för alla möjliga permutationer av data,

vilket är 8! = 40 320. Detta är dock varken nödvändigt eller möjligt vid ett större antal observationer. I de flesta fall räcker det att göra ett bra urval av dem. Till exempel visar figur 3 värdet av d_i för 1 000 permutationer av våra åtta observationer. Av dessa ledde 264 (de ljusgrå i figuren) till ett värde större än eller lika med d_0. Sannolikheten för att vårt resultat är en ren produkt av slumpen är därför $p = {}^{264}\!/_{1000} = 0{,}264$.

Figur 3. Histogram av värdet av $d_i = \overline{h}_i^t - \overline{h}_i^s$ som är resultatet av 1 000 permutationer av de 8 observationerna i tabell 3



Även om $p = 0{,}264$ kan verka som ett tillräckligt bra resultat (det är fortfarande nästan 75 procent chans att vår hypotes stämmer) är vetenskapen försiktig, och konventionen är att vi behöver vara minst 95 procent säkra innan vi accepterar någon hypotes (därmed $p \leq 0{,}05$). Normen i detta fall skulle vara att säga att data inte stödjer vår hypotes även om vi misstänker att det kan ligga något i den. Det är dock viktigt att förstå att inte heller den alternativa hypotesen att *tallies* är kortare än *shorties* stöds av data ($\overline{h}_0^t - \overline{h}_0^s < 0$ leder till $p = {}^{736}\!/_{1000} = 0{,}736$). Om något verkar den än mindre sannolik. Notera att det med ett urval om endast 8 fall är nästan omöjligt att testa en hypotes.

Det förra var dock det bara ett exempel och vi har faktiskt 1 000 observationer från K2-18b. Genom att utföra samma permutationstest på hela urvalet får vi det häpnadsväckande resultatet

att $d_i < d_0$ gäller för samtliga 1 000 permutationer. Sannolikheten för att vår observerade längdskillnad är en ren produkt av slumpen är därför $p < 0{,}001$. Observera att jag skriver $p < 0{,}001$ och inte $p = 0/1000 = 0{,}000$ eftersom vi genom att öka antalet permutationer tillräckligt mycket vid något tillfälle skulle få ett utfall där $d_i > d_0$, vilket inte är omöjligt, bara mycket osannolikt. Med detta sagt verkar testet vi utförde fullt ut stödja H1 och H2. Med andra ord är vi nu 99,999 procent säkra på att *tallies* verkligen i genomsnitt är längre än *shorties* i hela populationen på K2-18b.

Innan vi fortsätter med att testa de övriga hypoteserna är det viktigt att understryka att vi hittills endast har fastställt att det finns ett signifikant samband mellan *type* och *height*, medan vårt test inte kan säga något om riktningen av detta samband, det vill säga vilken variabel som är orsaken till den variation som observerats hos den andra. Vårt test stödjer därför både H1 och H2 trots deras motsatta kausala riktningar. Det var något enklare att gruppera data efter *type* och testa för skillnader i *height* eftersom det förra är en kategorisk variabel, men om vi hade gjort tvärtom skulle vi ha fått fram ett likvärdigt resultat. Med andra ord testade vi för *samband* mellan variabler, inte för *kausalitet*.

Matpreferenser (H3 och H4)

Samma procedur kan också användas för att testa hypotesen om det samband mellan *eat* och *height* (H4) som föreslagits av de evolutionära exologerna. I detta fall är den genomsnittliga längden för infödda som äter bambu $\overline{h}_b = 11{,}45$ m, medan den för manetätare är $\overline{h}_j = 10{,}75$ m, vilket ger $d_0 = 0{,}7$ m. Även i detta fall visade testet att $d_i < 0{,}7$ för alla 1 000 permutationer, vilket leder till $p < 0{,}001$. Data stödjer därför hypotesen om ett samband mellan intag av bambu och längd.

För att testa hypotesen om ett samband mellan *type* och *eat* (H3) behöver vi göra en liten justering i vårt test eftersom båda är kategoriska variabler. En av de huvudsakliga fördelarna med permutationstestet är att det inte förutsätter många antaganden om data. I praktiken behöver vi bara ändra i den statistik som används för att kunna mäta utfallet av varje permutation: i stället för att kontrollera genomsnittlig längd för *shorties* och *tallies* ska vi nu beräkna *andelen* av infödda som äter bambu i varje grupp, medan

allt annat förblir exakt detsamma. Att använda variabeln *bamboo* som endast har värdena 0 eller 1 (kom ihåg att 1 innebär en preferens för bambu) är särskilt praktiskt här eftersom andelen bambuätare i *eat* helt enkelt är medelvärdet av *bamboo*.

Andelen bambuätare bland *shorties* är $\overline{b}_s$ = 0,528 (eller 52,8 %) och bland tallies $\overline{b}_t$ = 0,536 (53,6 %) vilket ger $d_0 = \overline{b}_s - \overline{b}_t = -0{,}056$. Genom att köra permutationstestet fick vi 377 fall där d_i var större än detta tal, vilket betyder p = 0,377. Detta är inte tillräckligt för att avfärda möjligheten att vi fick fram vårt resultat av en slump, och vi kan bara dra slutsatsen att data inte stödjer (H3): både kulturexologerna och de evolutionära exologerna verkar ha fel på denna punkt.

Sammanfattningsvis har vi hittills individuellt kontrollerat de olika samband mellan variabler som föreslagits av exologerna och funnit att två av dem, nämligen *type* ↔ *height* och *eat* ↔ *height*, stöds av data, medan *type* ↔ *eat* inte stöds av data (den dubbla pilen betyder att vi är osäkra på kausalitetens riktning). Vi har dock endast testat för individuella samband, inte för hela modellerna i figur 1. Dessutom kan våra tester inte säga mycket om kausalitetens riktning mellan variablerna. Dessa frågor är verkligen mer komplicerade och behöver en noggrann diskussion.

Kausal inferens

Tre nivåer av kausal inferens

Att fastställa förekomsten av ett samband mellan variabler representerar en sorts första nivå av *kausal inferens*, det vill säga processen att dra slutsatser om kausalitet med ledning av data som samlats in från ett urval av en specifik population. Det kan ses som en nödvändig men inte tillräcklig förutsättning för att fastställa att en variabel *X* orsakar en annan variabel *Y* (Pearl 2019). Det är nödvändigt eftersom ett fastställt samband gör det möjligt både att formulera hypoteser om möjliga kausala mekanismer och att testa dem (Grimmer 2014). Dock är det inte tillräckligt, eftersom både kausalitetens riktning och den eventuella effekten av andra, icke mätta variabler (spuriös korrelation) är svåra att bedöma endast genom mätning av samband. I vårt fall kunde vi genom att individuellt testa de fyra hypoteserna som härstammar från exologernas

modeller utesluta H3 och bekräfta H4, men vilken av de två H1 och H2 som är sann kunde inte fastställas eftersom båda är förenliga med data.

Den andra nivån av kausal inferens och den traditionella vägen ut ur problemet är *intervention* (Pearl 2019; Grimmer 2014). Genom att på konstgjord väg förändra någon aspekt av verkligheten (variabel *X*) och mäta den resulterande effekten på något utfall (variabel *Y*), samtidigt som allt annat hålls konstant, kan vi logiskt dra slutsatsen att variationen i *X* orsakar de observerade förändringarna i *Y*. Konstanthållningen av allt annat uppnås i experimentella sammanhang vanligtvis genom *randomisering*, där slumpmässig tilldelning används för att skapa grupper som är statistiskt ekvivalenta, vilket innebär att de har samma fördelning av alla intressevariabler. Om man tar en tillräckligt stor grupp människor och slumpmässigt för dem till olika grupper kan man nämligen förvänta sig att varje resulterande grupp omfattar ungefär lika många kvinnor, vänsterhänta individer, vegetarianer eller vilken annan egenskap som helst, just eftersom grupptillskrivningen är rent slumpmässig. Det är kraften i randomisering!

När grupperna har bildats utförs en intervention på varje grupp och de utfall som är av intresse mäts. Om randomiseringen utfördes korrekt och grupperna var initialt ekvivalenta blir det legitimt att dra slutsatsen att eventuella observerade skillnader mellan grupperna efter interventionen är en produkt av interventionen i sig (Thye 2007). I BWT-experimentet varierade till exempel forskarna graden av oordning (nedskräpning och graffiti) runt en brevlåda belägen på en liten gata i Groningen (intervention) och mätte hur många av de förbipasserande som inte kunde motstå frestelsen att stjäla en femeurosedel i ett dåligt förseglat kuvert (utfall). Resultatet var ganska dramatiskt: 13 procent av de förbipasserande stal faktiskt fem euro när det var mer ordning, medan 27 procent stal sedeln när det var mer oordning (Keizer m.fl. 2008).

Även om resultaten från liknande studier är imponerande och experimentell forskning verkligen borde spela en större roll inom sociologin (Thye 2007; Gërxhani & Miller 2022) finns det praktiska eller etiska omständigheter som gör det omöjligt att allmänt tillämpa ett liknande förfarande. Dessutom löser vi inte problemet med att skilja mellan modellerna i figur 1 eftersom exologerna är

förhindrade att återvända till K2-18b. Detta leder oss till den tredje nivån av kausal inferens, som kallas *counterfactuals*, på svenska ”kontrafaktisk analys”, och kan ses som en begreppslig integration av samband och intervention (Pearl 2019). Av utrymmesskäl kan jag inte diskutera de logiska grundvalarna för kontrafaktisk analys (Morgan & Winship 2014). För våra syften är det tillräckligt att uppfatta den som ett slags tankeexperiment om effekterna av olika potentiella tillstånd hos en orsaksvariabel *X* på det uppmätta tillståndet hos en utfallsvariabel *Y*. Ett stort antal sådana tankeexperiment kan genereras på basis av olika antaganden om kausalitet, och resultaten av dem kan kontrolleras mot empiriska data. Med andra ord ska man först bygga upp en uppsättning möjliga modeller som omfattar väl avgränsade kausala samband mellan variablerna och alla deras möjliga tillstånd. Sedan använder man dessa modeller för att göra förutsägelser som kan jämföras med empiriska data i syfte att avgöra vilken av de kausala modellerna som med störst sannolikhet är sann med tanke på den tillgängliga empiriska evidensen.

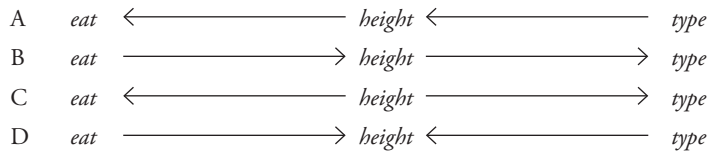
Kausala grafer

Kontrafaktisk analys kan vara både teoretiskt och praktiskt komplicerad. Användningen av *directed acyclic graphs* (DAG), på svenska ”riktade acykliska grafer”, kan göra detta till något hanterbart i praktiken (Breen 2022). Själva begreppet DAG låter lite farligt men är i sin kärna bara en metod för att visualisera och analysera kausala relationer mellan variabler. Samhällsvetenskapliga forskare använder dem faktiskt ofta utan att känna till namnet (som i figur 1), även om de innebär mer formalisering än vad som är vanligt inom samhällsvetenskap.

DAG är riktade nätverk där varje nod representerar en variabel, medan förbindelser representerar de antagna kausala relationerna. Till exempel innebär $X \rightarrow Y$ att variabeln *X* orsakar en effekt på *Y*, som i figur 1. Den främsta begränsningen hos DAG är att de måste vara acykliska, det vill säga det får inte finnas någon loop i nätverket. Ett nätverk som $X \rightarrow Y \rightarrow Z \rightarrow X$ omfattar en loop, eftersom variabeln *X* indirekt påverkar *Z* (genom *Y*) och återigen påverkas av *Z*. Ett sådant nätverk är alltså inte en DAG. Begränsningen beror inte på möjligheten att visualisera nätverket utan på de antaganden som krävs för att formellt analysera det.

För att bli mer praktisk och återgå till frågan om hur *height*, *type* och *eat* påverkar varandra, kan vi först och främst notera att vi redan vet något. Vi slog fast att det inte finns något statistiskt samband mellan *type* och *eat* medan signifikanta samband fanns för *type* ↔ *height* och *eat* ↔ *height*. Den goda nyheten är att det bara finns fyra sätt att koppla samman tre noder (som representerar variabler) med direkta förbindelser (som representerar kausala relationer) som är konsistenta med de samband som empiriskt påträffades i våra data (figur 4).

Figur 4. Möjliga kausala strukturer som förbinder de tre variablerna eat, height *och* type *utifrån de två signifikanta samband som visades i analysen ovan*



För att förstå vilken av dessa potentiella DAG som återspeglar de faktiska kausala relationerna i data är det viktigt att fokusera på det resultat vi fått fram som visar att *eat* och *type* är statistiskt oberoende av varandra, det vill säga att inget samband finns mellan dem. Tänk oss nu att exologerna inte mätte *height*. Om fallen A, B och C i figur 4 var sanna skulle vi fortfarande förvänta oss att finna ett signifikant samband mellan *eat* och *type*: i fallen A och B eftersom de påverkar varandra genom förmedling av den (icke observerade) variabeln *height*, och i fallet C eftersom de båda påverkas av *height*. Det enda fallet där *eat* och *type* kan vara oberoende av varandra är D, vilket alltså är *den enda struktur i det kausala nätverket som är förenlig med data*.

Proceduren att dra slutsatser om kausala strukturer från data blir mer komplicerad med fler variabler, men ändå genomförbar. Med hjälp av smarta algoritmer, som går under den allmänna beteckningen *Bayesian network learning* (på svenska bayesianskt nätverksinlärning), kan man göra detta på rimlig tid även med ganska komplexa dataset (Pearl 2009). Vi skulle faktiskt ha kunnat tillämpa dem direkt på våra data. Resultatet av en sådan analys visas i figur 5,

som också rymmer koefficienter som uttrycker styrkan i sambandet mellan variablerna (i detta fall standardiserade koefficienter från regressionsmodellen). Kort sagt: både *eat* och *type* påverkar signifikant invånarnas *height* där den senare visar en effekt ungefär dubbelt så stark som den förra. Observera också att båda koefficienterna är positiva, vilket är konsistent med exologernas val att placera små plussymboler nära pilarna i figur 1.

Figur 5. Bayesiansk inlärning av nätverksstrukturen utifrån K2-18b:s data med standardiserade koefficienter

Med detta har vi nått slutet av vår analys och vi överlämnar gärna figur 5 till exologerna som en resultatsammanfattning. Ingen av de föreslagna modellerna kunde helt förklara data, men så är det med vetenskapen. Den ger aldrig perfekta svar utan strävar bara efter att undan för undan öka vår förståelse. Förresten, hur kan man vara så säker på att figur 5 verkligen återspeglar de riktiga relationerna i data? Det är här användningen av ett fiktivt dataset kommer in i bilden: eftersom jag själv skapade data så! Å andra sidan är det klart att också slutsatserna av vår analys är fiktiva och inte kan generaliseras till den verkliga världen.

Slutsats: data och sociala mekanismer

Uppkomsten av maskininlärningsmetoder, såsom de bayesianska nätverk som använts här, och den ökande tillgången till *big data* har fått vissa forskare att tala om *the end of theory* (Anderson 2008; Mayer-Schönberger & Cukier 2013). Enligt denna resonemangslinje är skapandet av modeller och hypoteser förlegat med tanke på de nya möjligheter som erbjuds av *big data* och maskininlärningsmetoder, som gör det möjligt att direkt utvinna befintliga mönster ur data. Uppriktigt sagt kan den ovanstående analysen mycket väl stödja denna idé, eftersom DAG i figur 5 faktiskt avslöjade de äkta kausala relationer som låg till grund för data. Samtidigt upplyser inte DAG om orsakerna till de identifierade datarelationerna (Morgan & Winship 2014, kapitel 3). Med andra

ord kan maskininlärningsmetoder finna sambanden mellan variabler och till och med kausalitetsriktningen men inte förklara *varför* dessa mönster finns. Att förstå *vad* utan att förstå *varför* anses vanligtvis otillräckligt för en fullständig vetenskaplig förklaring (Mazzocchi 2015), och detta kanske gäller i än högre grad för samhällsvetenskapen (Hedström & Ylikoski 2010; Keuschnigg m.fl. 2018; Nelimarkka 2022).

För att bättre förstå detta kan vi jämföra figur 1 och figur 5. Båda visar ett nätverk av kausala relationer som förbindelser mellan de observerade variablerna. Den största skillnaden mellan dem är att exologerna inte bara antar en viss struktur av kausala relationer utan också skapar den på basis av underliggande *mekanismer* som antas vara verksamma i den undersökta situationen: sociala konstruktioner av typen i kulturmodellen och samspelet mellan gener och miljö i den evolutionära modellen. DAG:en i figur 5 identifierar bättre än båda modellerna de äkta relationerna i data men ger inte kunskap om de underliggande mekanismerna. Dessutom opererar bayesianska nätverk med sannolikheter, liksom alla statistiska algoritmer eller maskininlärningsalgoritmer, vilket innebär att de inte alltid lyckas identifiera den äkta strukturen av kausala relationer. I vårt fall visste vi det på förhand, och därför kan vi säga att algoritmen gjorde ett bra jobb, men i mer komplexa situationer kan algoritmer leda till identifiering av olika nätverk utan en tydlig vinnare, och ofta visas märkliga förbindelser, det vill säga förbindelser som är logiskt omöjliga eller åtminstone motsäger den gängse kunskapen. I sådana fall används teori eller expertkunskap normalt för att begränsa algoritmernas beteende så att de bara visar meningsfulla resultat: en ledtråd om att teorin kanske inte är helt död än.

Det är viktigt att förstå att teoridrivna och datadrivna tillvägagångssätt är komplementära snarare än konkurrerande (Mazzocchi 2015; Nelimarkka 2022). En tillfredsställande förklaring av data kräver både belysning av den kausala strukturen hos data och testning av de mekanismer som ligger till grund för relationerna mellan empiriska variabler. Samhällsvetare kan inte bara ignorera att olika kausala strukturer grundar sig i mycket olika perspektiv på hur samhället fungerar (Hedström & Ylikoski 2010). Det är precis det som de flesta är intresserade av. Samtidigt är vanan att kontinuerligt kontrollera alla teoretiska påståenden mot empiriska data

underutvecklad inom sociologin (Sørensen 2009). Den ökade tillgången på data och metoder bör därför välkomnas och bättre integreras med de traditionella forskningsmetoderna inom disciplinen (Lazer m.fl. 2009; Keuschnigg m.fl. 2018; Nelimarkka 2022).

Litteratur

Anderson, Chris (2008) ”The End of Theory. The Data Deluge Makes the Scientific Method Obsolete”, *Wired Magazine* 23 juni.

Box, George E.P., J. Stuart Hunter & William G. Hunter (2005) *Statistics for Experimenters. Design, Innovation and Discovery*, andra upplagan. Hoboken: John Wiley.

Breen, Richard (2022) ”Causal Inference with Observational Data”, 272–286 i Klarita Gërxhani, Nan D. de Graaf & Werner Raub (red.), *Handbook of Sociological Science. Contributions to Rigorous Sociology*, Cheltenham: Edward Elgar.

Buchanan, Mark (2007) *The Social Atom. Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks Like You*. New York: Bloomsbury.

Cohen, Morris & Ernest Nagel (2002) *An Introduction to Logic and Scientific Method*. Simon Publications.

Gërxhani, Klarita, Nan D. de Graaf & Werner Raub (red.) (2022) *Handbook of Sociological Science. Contributions to Rigorous Sociology*. Cheltenham: Edward Elgar.

Gërxhani, Klarita & Luis Miller (2022) ”Experimental Sociology”, 309–323 i Klarita Gërxhani, Nan D. de Graaf & Werner Raub (red.), *Handbook of Sociological Science. Contributions to Rigorous Sociology*, Cheltenham: Edward Elgar.

Grimmer, Justin (2014) ”We are All Social Scientists Now. How Big Data, Machine Learning, and Causal Inference Work Together”, *PS: Political Science & Politics* 48(1): 80–83.

Hedström, Peter (2005) *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Hedström, Peter & Richard Swedberg (1998) *Social Mechanisms*. Cambridge: Cambridge University Press.

Hedström, Peter & Petri Ylikoski (2010) ”Causal Mechanisms in the Social Sciences”, *Annual Review.of Sociology* 36(1): 49–67.

Keizer, Kees, Siegwart Lindenberg & Linda Steg (2008) ”The Spreading of Disorder”, *Science* 322: 1681–1685.

Keuschnigg, Marc, Niclas Lovsjö & Peter Hedström (2018) ”Analytical Sociology and Computational Social Science”, *Journal of Computational Social Science* 1(1): 3–14.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Lázló Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy & Marshall Van Alstyne (2009) ”Computational Social Science”, *Science* 323(5915): 721–723.

Madhusudhan, Nikku, Matthew C. Nixon, Luis Welbanks, Aanjali A. Piette & Richard A. Booth (2020) "The Interior and Atmosphere of the Habitable-zone Exoplanet k2-18b", *The Astrophysical Journal Letters* 891(1): L7.

Maskaly, Jon & Lyndsey N. Boggess (2014) "Broken Windows Theory", i J.M. Miller (red.), *The Encyclopedia of Theoretical Criminology*. https://doi.org/10.1002/9781118517390.wbetc127.

Mayer-Schönberger, Viktor & Kenneth Cukier (2013) *Big Data. A Revolution that will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.

Mazzocchi, Fulvio (2015) "Could Big Data Be the End of Theory in Science? A Few Remarks on the Epistemology of Data-Driven Science", *EMBO reports* 16(10): 1250–1255.

Morgan, Stephen L. & Christopher Winship (2014) *Counterfactuals and Causal Inference*. New York: Cambridge University Press.

Nelimarkka, Matti (2022) *Computational Thinking and Social Science. Combining Programming, Methodologies and Fundamental Concepts*. London: Sage Publications.

Pearl, Judea (2009) *Causality*. Cambridge: Cambridge University Press.

Pearl, Judea (2019) "The Seven Tools of Causal Inference, with Reflections on Machine Learning", *Communications of the ACM* 62(3): 54–60.

Raub, Werner, Nan D. de Graaf & Klarita Gërxhani (2022) "Rigorous Sociology", 2–19 i Klarita Gërxhani, Nan D. de Graaf & Werner Raub (red.), *Handbook of Sociological Science. Contributions to Rigorous Sociology*, Cheltenham: Edward Elgar.

Sørensen, Aage B. (2009) "Statistical Models and Mechanisms of Social Processes", 369–399 i Peter Hedström & Björn Wittrock (red.), *Frontiers of Sociology*. Leiden: Brill.

Thye, Shane R. (2007) "Logical and Philosophical Foundations of Experimental Research in the Social Sciences", 57–86 i Murray J. Webster & Jane Sell (red.), *Laboratory Experiments in the Social Sciences*. Boston: Academic Press.

Watts, Duncan J. (2011) *Everything is Obvious, Once You Know the Answer. How Common Sense Fail Us*. New York: Crown Business.

Watts, Duncan J. (2014) "Common Sense and Sociological Explanations", *American Journal of Sociology* 120(2): 313–351.

Watts, Duncan J. (2017) "Should Social Science Be More Solution-Oriented?", *Nature Human Behaviour* 1(1): 0015.

Wilson, James Q. & George L. Kelling (1982) "Broken Windows. The Police and Neighborhood Safety", *Atlantic Monthly* mars: 29–38.