

Formell modellering av kravet til årsakssammenheng i norsk erstatningsrett

Sjur K Dyrkolbotn *

Sammendrag

Vi formaliserer ulike tolkninger av kravet til årsakssammenheng i norsk erstatningsrett. Med dette viser vi at den rettslige definisjonen av en årsakssammenheng kan forstås på mange ulike måter, og at ulike forståelser kan presiseres ved hjelp av enkle formelle metoder. Et hovedpoeng er å vise at formell modellering kan brukes til å kaste lys over vanskelige problemstillinger som oppstår når vi skal ta stilling til skadelige hendelser med flere samvirkende årsaker.

1 Introduksjon

I norsk rett er såkalt “adekvat årsakssammenheng” et sentralt rettslig vilkår for skadeerstatning. For at skadelidte skal få medhold i krav om erstatning fra en påstått skadevolder må han vise at skadevolderen har handlet på en måte som forårsaket skaden. Denne kausale relasjonen må dessuten være “adekvat”, i den forstand at skadefølgen må ha vært påregnelig, ikke en fjern og avledet konsekvens av skadevolder sine handlinger. I denne artikkelen skal vi se på erstatningsretten sitt årsaksbegrep, uten å ta stilling til kravet om adekvans, som i mindre grad egner seg for formell modellering.

I avsnitt 2 introduserer vi kort de to viktigste årsaksteoriene i henholdsvis rettsvitenskap og informatikk, den såkalte “betingelseslæren” (e.g., [13]) på den ene siden og “Halpern-Pearl” (HP)-teorien på den andre (e.g., [6]).¹ Et hovedpoeng er å få frem at betingelseslæren ikke gir oss noen entydig definisjon av kausalitet, mens HP-teorien gir oss mange alternative definisjoner. Hensikten med formell modellering i denne artikkelen blir dermed å kartlegge ulike måter å tolke kausalitetsbegrepet på, noe som i sin tur kan bidra med ny og viktig innsikt inn mot både rettsvitenskap og informatikk. Vi bemerker spesielt at årsaksmodellering er et veletablert felt innenfor datavitenskapen, som kan ha betydelig nytte av en tverrfaglig tilnærming.²

*Institutt for bygg, Høgskulen på Vestlandet.

This paper was presented at the NIK-2017 conference; see <http://www.nik.no/>.

¹Både HP-teorien og betingelseslæren (i ulike varianter) er del av den såkalte “kontrafaktiske” teori-tradisjone for kausalitet. Denne tilbakeføres ofte til filosofene David Hume og John Stuart Mill, men er i moderne filosofi særlig assosiert med David Lewis [11].

²Se, for eksempel, [2, 3].

I avsnitt 3 gir vi enkle formaliseringer av en rekke tolkninger av kravet til årsakssammenheng i norsk rett. Et sentralt poeng er at vi får ulike svar på årsaksspørsmålet avhengig av hvordan vi identifiserer den hendelsen vi vil finne årsakene til. Hvis vi for eksempel nøyer oss med å spørre hvorfor det oppstod en skade vil vi generelt sett få et annet svar enn hvis vi spør hvorfor det oppstod skade på en bestemt måte. Jo mer informasjon om skadeforløpet vi tar med i spørsmålsstillingen, jo flere faktorer vil vi klassifisere som årsaker.

Vi viser at denne mekanismen lar seg beskrive og analysere på enkle modeller tatt fra HP-teorien. Ulike tolkninger av hva vi egentlig ber om når vi spør etter årsakene til en gitt hendelse fører til nokså ulike formaliseringer av det rettslig årsaksbegrepet. I avsnitt 4 kartlegges ytterpunktene for hvordan dette begrepet kan forstås. Det påvises så en sammenheng mellom disse ytterpunktene: den mest liberale tolkningen vi kan forsvare ender opp med å identifisere som årsaker alle de faktorene som *kunne vært* årsaker ifølge den strengest mulige tolkningen. Vi får med dette en enkel karakterisering av tolkningsrommet for kravet til årsakssammenheng i norsk erstatningsrett. At dette rommet er stort kan forklare hvorfor vi i norsk erstatningsrett har klart oss bra med et tilsynelatende naivt årsaksbegrep, på tross av protester fra juridisk teori (e.g., [13]).

2 Kort om nødvendige betingelser og strukturelle ligninger

I norsk erstatningsrett har definisjonen av rettslig årsakssammenheng en tilsynelatende logisk form: i tråd med langvarig praksis fra Høyesterett er X å anse som en årsak til Y når X er en *nødvendig betingelse* for Y . Her er det imidlertid ikke samsvar mellom det rettslige og det formallogiske nødvendighetsbegrepet. Når man i logikk snakker om at ϕ_X er en nødvendig betingelse for ϕ_Y betyr dette vanligvis at ϕ_Y ikke kan være sann uten at også ϕ_X er sann, det vil si $\phi_Y \rightarrow \phi_X$. Generelt sett vil dette ikke karakterisere en årsakssammenheng; det er sjelden gitt at ϕ_X kommer før i tid enn ϕ_Y , og så lenge vi jobber med materiell implikasjon, som i klassisk logikk, trenger ikke ϕ_X og ϕ_Y ha noe å gjøre med hverandre i det hele tatt.

For å ta stilling til spørsmålet om årsak må vi vite noe om hvordan X og Y står til hverandre som hendelser i et system som forandrer seg over tid. En naturlig måte å modellere dette på er ved å beskrive en hendelse som en verdi gitt ved en funksjon av andre hendelser. For en mengde hendelser får vi da en mengde strukturelle ligninger, som i HP-teorien brukes til å definere kausalitet [8, 6]. Denne teorien er nokså komplisert, kanskje mer enn man skulle tro. Flere alternative definisjoner har blitt foreslått og det foreligger ikke noe konsensus om hvordan kausalitet bør defineres for dette rammeverket [6].

Den tradisjonelle tolkningen av nødvendighetskravet – den såkalte *betingeslæren* – virker derimot nokså entydig ved første øyekast: X er en årsak til Y dersom Y ikke ville ha skjedd når X tenkes borte. I det følgende skal vi referere til dette som den *naive* tolkningen; den er enkel og klar, men som vi skal se, utilstrekkelig i mange tilfeller. For å illustrere dette kan vi vise til to viktige dommer fra Høyesterett. Den første dommen, inntatt i Rt-2007-172 (Schizofreni), gjaldt et krav om erstatning etter yrkesskadeforsikringsloven § 11. Det ble av skadelidte gjort gjeldende at det forelå adekvat årsakssammenheng mellom en fallulykke og en etterfølgende alvorlig sinnslidelse (paranoid schizofreni). Førstevoterende oppsummerte i avsnitt 42 årsaksvurderingen slik: “For at arbeidsulykken skal anses som årsak til As

sinnslidelse, må ulykken ha vært en nødvendig betingelse for at A utviklet paranoid schizofreni *slik han gjorde*” (min utheving). Som vanlig ble det stilt et nødvendighetskrav, men her ble altså dette kravet relatert til *hvordan* skadelidte utviklet paranoid schizofreni. Dette skulle vise seg å bli viktig i denne saken.

Sentralt stod nemlig spørsmålet om skadelidte ville ha utviklet paranoid schizofreni uansett, siden han var disponert for dette og nettopp hadde gått gjennom et opprivende samlivsbrudd. I så fall kunne man hevde at ulykken ikke var en nødvendig betingelse, siden sinnslidelsen ville kommet selv om ulykken ikke hadde skjedd. Førstevoterende beskrev selv samlivsbruddet som en “ikke uvesentlig faktor i sykdomsbildet” i avsnitt 63, men konkluderte samtidig med at de øvrige årsaksfaktorer var en følge av arbeidsulykken. Etter en helhetsvurdering ble konklusjonen at det var “mest nærliggende å gå ut fra at A ikke ville ha utviklet schizofreni, slik han gjorde, hvis han ikke hadde blitt utsatt for de etterfølgende omstendigheter som skyldtes arbeidsulykken”.

Denne årsakssammenhengen ble riktignok ansett som så avledet at den ikke fremstod som adekvat, men det sentrale for oss er at Høyesterett anså fallulykken som en “nødvendig betingelse” for at A utviklet paranoid schizofreni. Dette viser at nødvendighetskravet ikke skal tolkes for strengt; som det fremgår i avsnitt 62 var de sakkyndige enige i at fallulykken “i seg selv ikke har hatt betydning for As sinnslidelse”, men Høyesterett kom til at fallulykken hadde ført til en endret livssituasjon som gjorde at A utviklet sinnslidelsen *slik han gjorde*. Underforstått her er at A kunne ha utviklet paranoid schizofreni uansett, men at det i så fall ville skjedd på en annen måte. Dette var nok til at nødvendighetskravet ble ansett oppfylt.

Lignende resonnering finner vi også i andre dommer, herunder Rt-1992-64 (P-pilledom II), avgitt under dissens 3-2. Spørsmålet i saken var om bruken av en bestemt type p-pille var å anse som årsak til hjerneslag. Høyesterett tok utgangspunkt i den naive tolkningen av betingelseslæren, som ble oppsummert slik av førstevoterende på side 69: “årsakskravet mellom en handling eller unnlattelse og en skade er vanligvis oppfylt dersom skaden ikke ville ha skjedd om handlingen eller unnlattelsen tenkes borte. Handlingen eller unnlattelsen er da en nødvendig betingelse for at skaden inntreffer”. Legg merke til at førstevoterende kun gir tilstrekkelige vilkår for at en handling eller en unnlattelse skal anses som en nødvendig betingelse: det utelukkes ikke at en handling eller en unnlattelse kan bli å anse som en nødvendig betingelse også i andre tilfeller.

I Rt-1992-64 var det langt fra opplagt at p-pillene alene gjorde en forskjell. De sakkyndige var uenige om p-pillene sin rolle, men det var enighet om at en hel rekke faktorer hadde medvirket til hjerneslag. Ifølge annenvoterende, som uttalte seg for mindretallet på side 87, var det da ikke grunnlag for å konkludere med at “hjernetrombosen ikke ville ha inntruffet dersom A ikke hadde brukt p-piller”. Førstevoterende slo derimot fast på side 70 at siden det var snakk om samvirkende skadeårsaker ville det “være tilstrekkelig at det foreligger sannsynlighetsovervekt for at p-pillebruken har vært en nødvendig medvirkende årsak til skaden.”

Hva menes med en “nødvendig medvirkende årsak”? Hvis førstevoterende bare tok sikte på å få frem at p-pillebruken ikke var den eneste nødvendige betingelsen for hjerneslag var det ingen grunn til å modifisere utgangspunktet om at p-pillebruken måtte være individuelt

nødvendig. Det var heller ingen grunn til å erstatte ordet “betingelse” med ordet “årsak”, som synes å forutsette en uavhengig forståelse av årsaksbegrepet. Det virker derfor rimelig å anta at førstevoterende tok sikte på å komme forbi den naive tolkningen av betingelseslæren, ved å åpne for at nødvendighetskravet også kan oppfylles ved at man ser den aktuelle faktoren i sammenheng med andre faktorer. Det synes som at førstevoterende ikke var forberedt på å konkludere med at hjerneslaget ikke ville ha inntruffet dersom p-pillebruken tenkes borte. P-pillebruken var likevel en nødvendig medvirkende årsak, om ikke alene, så i alle fall sammen med andre samvirkende faktorer.

Med en slik mer sammensatt tolkning av betingelseslæren vil vi også kunne håndtere tilfeller med konkurrerende årsaker, der en hendelse er forårsaket av flere individuelt tilstrekkelige faktorer. For eksempel, dersom både A og B forurenser en elv så mye at fisken i elven dør, kan vi si at utslippene var “nødvendige medvirkende årsaker” til fiskedøden. Situasjonen er her slik at ingen faktor er individuelt nødvendig for fiskedød; det finnes ingen faktor slik at fisken ikke ville ha dødd dersom denne faktoren tenkes borte. Det finnes derimot en *mengde* av slike faktorer, nemlig mengden bestående av utslippene fra både A og B. Disse faktorene blir dermed nødvendig medvirkende årsaker; de er begge nødvendige medlemmer av en mengde som samlet oppfyller kravet om at når faktorene fra mengden tenkes borte ville skaden ikke ha skjedd.

I juridisk teori har man sett at den naive tolkningen av betingelseslæren ikke gir akseptable resultater for hendelser med flere tilstrekkelige årsaker [1, 13]. For slike tilfeller er det antatt at man må gjøre unntak fra betingelseslæren. I rettspraksis har imidlertid behovet for slike unntak vist seg å være lite.³ Saker som Rt-1992-64 og Rt-2007-172 kan forklare hvorfor. I praksis vil man tilpasse tolkningen og anvendelsen av betingelseslæren til omstendighetene. En faktor vil nok fort bli “nødvendig” når den er “så vidt vesentlig i årsaksbildet at det er naturlig å knytte ansvar til den”, for å sitere førstevoterende i Rt-1992-64 på side 70.

Det er i alle fall på det rene at kravet til årsakssammenheng i norsk erstatningsrett ikke er entydig og formallogisk bestemt, selv om enkelte rettsvitere antyder det motsatte [10]. Dette blir enda tydeligere når vi formaliserer årsakskravet, slik vi skal gjøre i neste avsnitt. Samtidig ser vi at det er behov for å kartlegge mer presist hvordan nødvendighetskravet kan tolkes på ulike måter. Formaliseringen vi nå gir er ment som et bidrag til en slik kartlegging.

3 Formelle tolkninger av kravet til nødvendighet

Vi tar utgangspunkt i Halpern og Pearl sin måte å modellere hendelser på, ved hjelp av strukturelle ligninger [12, 8, 6]. For å forenkle rammeverket vil vi kun se på enkeltligninger, ikke systemer av slike. Dette svarer til klassen av spill i normalform: funksjoner som avhenger av en mengde variabler (spillere) og returnerer en verdi (et utfall). For å snakke om utfall skal vi bruke et utsagnslogisk språk L , definert over en mengde Π av utsagnslogiske variabler. Dermed kan vi snakke om mer enn bare hvilken gevinst de ulike spillerene sitter

³I en nyere dom fra lagmannsretten ble det imidlertid gjort et slikt unntak, under henvisning til juridisk teori, se LB-2015-65756.

igjen med. Spesielt negasjon er viktig for å analysere kausalitet. Hvis vi lurer på hvorfor utfallet *ikke* tilfredstiller egenskapen ϕ (en annen gevinstprofil, for eksempel), så kan vi spørre hvem som forårsaket $\neg\phi$.⁴

Vi lar $\mathbb{M} = \{0, 1\}^\Pi$ være mengden av alle modeller, det vil si funksjoner fra atomene Π til sannhetsverdiene $\{0, 1\}$. Når en modell $m \in \mathbb{M}$ tilfredstiller ϕ skriver vi $m \models \phi$. Tilfredstillelse er definert som i vanlig utsagnslogikk. Vi nevner spesielt at $m \models \neg\phi$ dersom $m \not\models \phi$. Når en mengde modeller $M \subseteq \mathbb{M}$ tilfredstiller ϕ skriver vi $M \models \phi$ og når alle modeller tilfredstiller ϕ skriver vi $\models \phi$. I så fall er ϕ en tautologi. Vi definerer også $[[\phi]] = \{m \in \mathbb{M} \mid m \models \phi\}$. Dette er mengden av modeller som tilfredstiller ϕ .

Vi kan nå definere spill i normalform med utfall beskrevet i logikk som en struktur $G = (V, (D_v)_{v \in V}, \sigma)$. Her er V en mengde variabler (spillere) og $(D_v)_{v \in V}$ er en endelig ikke-tom mengde av hendelser (strategier), en for hver av spillerene $v \in V$. Vi kaller $p \in \prod_{v \in V} D_v$ for en *profil*. $P = \prod_{v \in V} D_v$ denoterer mengden av alle profiler. Funksjonen $\sigma : P \rightarrow \mathbb{M}$ sender profiler til utfall, der utfall representeres som modeller (med en tilhørende mengde av formler som er sanne på denne modellen).⁵

Hvis utfallet av et spill gjør ϕ sann er vi interessert i å vite hva som forårsaket ϕ . Vi antar at hendelsesforløpet er kjent, det vil si at vi vet hvilken $p \in P$ som skjedde. Spørsmålet er med andre ord: gitt at p skjer, hvilke av hendelsene i p er årsak til at ϕ er sann på $\sigma(p)$? En metode som gir oss et svar på dette spørsmålet er en kausal teori.

Definisjon 1. La G være et vilkårlig spill i normalform. En *kausal teori* for G er en funksjon $c : P \times L \rightarrow 2^V$, som sender par av profiler og formuler til mengder av variabler.

For et enkelt eksempel, anta at Siv og Erna hver gir halvparten av en dødlig dose gift til Jonas. Dermed dør Jonas. Dette kan representeres med et spill der Siv og Erna svarer til variablene v_1 og v_2 . La oss anta at valget de står overfor er å gi enten en halv dødlig dose, en hel dose, eller ingen gift i det hele tatt. Det vil si at vi har $D_{v_1} = D_{v_2} = \{0, \frac{1}{2}, 1\}$: Siv og Erna har tre alternative strategier hver. Den aktuelle profilen er $p = (\frac{1}{2}, \frac{1}{2})$ – begge gir en halv dødlig dose. Utfallet blir da $\sigma(p)$ med $\sigma(p) \models \alpha$, der α er påstanden om at Jonas er død. Hvis vi har en kausal teori i betydningen av Definisjon 1 kan vi spørre hvem som tok livet av Jonas. Svaret vil gis ved mengden $c(p, \alpha)$, mengden av spillere som er årsaker til at α er sann på $\sigma(p)$. Spørsmålet blir nå hvordan c må defineres, for å svare best mulig til det rettslige kausalitetsbegrepet.

Den naive tolkningen

Den naive tolkningen av nødvendighetskravet tilsier at v er en nødvendig betingelse for ϕ dersom ϕ ikke ville skjedd når v sin verdi tenkes borte. Det må med andre ord finnes en alternativ verdi for v , kall den x , slik at vi kan unngå ϕ ved å endre verdien av v til x (under forutsetning av at alt annet forblir uendret). For å formalisere dette, la oss definere

⁴Dette kan være høyst relevant også i en spillteoretisk sammenheng, for eksempel der ϕ beskriver et Pareto-optimalt utfall som av en eller annen grunn ikke oppnås, for eksempel fordi det ikke er i Nash-likevekt.

⁵Funksjonen σ trenger selvsagt ikke være surjektiv; typisk vil det finnes mange modeller som ikke svarer til noe utfall av et konkret spill.

$I(p, v) = \{q \in P \mid q(i) = p(i) \text{ for alle } i \neq v\}$, for alle $p \in P$. Det vil si at $I(p, v)$ er alle mulige varianter av p vi kan finne ved å endre *bare* verdien til v . Denne notasjonen utvides til mengder $A \subseteq V$ slik: $I(p, A) = \{q \in P \mid q(i) = p(i) \text{ for alle } i \notin A\}$. Dermed er $I(p, A)$ mengden av alle profiler vi oppnår fra p ved å endre på strategier i A , mens alle andre strategier holdes fast. Med dette kan vi gi den naive tolkningen av nødvendighetskravet, som svarer til det man i den engelskspråkelige litteraturen kaller “but-for”-kausaltet [9, 6].

Definisjon 2. For alle spill G og alle $p \in P, \phi \in L(p)$ definerer vi but-for kausalitet på følgende måte:

$$B(p, \phi) = \{v \mid \sigma(p) \models \phi \text{ and } \exists q \in I(p, v) : \sigma(q) \not\models \phi\}.$$

Hvis vi anvender denne teorien på eksempelet vårt får vi det naturlige svaret: både Siv og Erna er årsaker til at Jonas dør. Ved å endre verdien av enten v_1 eller v_2 til 0 ville vi få en profil p' slik at $\sigma(p') \models \neg\alpha$, det vil si slik at Jonas ikke dør. Både v_1 og v_2 er i den forstand nødvendige for utfallet; det finnes et alternativ for begge som alene ville ha vært nok til å avverge at Jonas dør.

I HP-litteraturen er but-for-definisjonen bare et illustrerende utgangspunkt; teorien avvises nesten med en gang den introduseres [6]. For å se hvorfor teorien ikke anses brukbar kan vi utvide eksempelet vårt med å anta at også Trine gir halvparten av en dødlige dose gift til Jonas. α blir fortsatt sann, ettersom Jonas nå får enda mer gift. Men noe sentralt har endret seg. Det er ikke lenger mulig for noen av spillerene å avverge utfallet alene. To halvparter av en dødelig dose gift er nok til at Jonas dør, så ingen av de tre giftgiverene er “but-for”-årsaker av α . Tvert imot får vi nå $B(p, \alpha) = \emptyset$: ingen av spillerene blir å anse som årsaker av α ifølge Definisjon 2.

Det er lett å si seg enig med litteraturen at dette ikke er et akseptabelt utfall; det synes opplagt at Siv, Erna og Trine alle er likeverdige årsaker til at Jonas dør i et slikt tilfelle. Det er derimot ikke opplagt at denne tolkningen av nødvendighetskravet alltid samsvarer med den tolkningen vi møter i rettspraksis. Det er heller ikke opplagt at tolkningen gir den mest naturlige formaliseringen av et krav til nødvendighet. Vi har nemlig unnlatt å avklare et viktig spørsmål: hva er det vi leter etter nødvendige betingelser for?

Det intuitive svaret, som ofte tas for gitt, er at vi leter etter nødvendige betingelser for de egenskapene ved utfallet vi anser som normativt viktige. Så, for eksempel, hvis Jonas dør av en overdose gift, så er det de nødvendige betingelser for at han dør vi er ute etter. Dermed overser vi en rekke andre egenskaper ved utfallet, herunder enhver informasjon vi måtte ha om *hvordan* Jonas døde. Det er langt fra opplagt at vi bør se vekk fra denne informasjonen. Slik argumentasjonen i Rt-2007-172 også viser kan det spille en rolle at skaden inntraff *slik den gjorde*.

For eksempel, dersom bare Erna og Siv gir gift til Jonas, virker det som en kvalitativt annerledes måte for Jonas å dø på, sammenlignet med situasjonen der han også blir forgiftet av Trine. På den annen side, dersom Trine ikke gir gift, men kunne velge mellom å gi salt eller pepper til Jonas, gir det ingen mening å skille mellom disse to utfallene når vi spør hva som er årsak til at Jonas dør. Det virker rimelig å anta at Jonas dør på nøyaktig samme måte uavhengig av om han får salt eller pepper i tillegg til to halve doser dødelig gift.

Her reiser det seg spørsmål om hvor nyansert vi *bør* være når vi spør om årsaker til en gitt egenskap ved et utfall. Når skal vi legge vekt på om egenskapen ble realisert på den ene eller den andre måten? I engelskspråklig juridisk teori har man et nokså velutviklet forslag til en slik avgrensning, kjent som den såkalte NESS-teorien. I norsk rettsvitenskap har denne teorien blitt oppfattet som et *alterantiv* til betingelseslæren [13], men det er etter mitt syn mer dekkende å beskrive teorien som en konkurrerende tolkning av hva vi mener med en “nødvendig betingelse”.

NESS-tolkningen

Ifølge NESS-teorien er X en årsak til Y dersom X er en nødvendig betingelse for at en mengde hendelser skal være tilstrekkelige til å frembringe Y [14]. Her ligger det utvilsomt et nødvendighetskrav i bunn, men dette har blitt gjort mer nyansert. For spill i normalform synes det klart nok hvordan NESS-teorien bør formaliseres. Vi er ute etter variabler slik at deres verdier er nødvendige for at en mengde variabler skal være tilstrekkelige til å fremtvinge ϕ på profilen p . For eksempel, i eksempelet der Siv, Erna, og Trine alle gir en halv dødelig dose gift til Jonas vil ethvert valg av to spillere gi oss en mengde slik at deres valgte strategier er tilstrekkelige til at Jonas dør. Begge medlemmene i alle slike mengder vil imidlertid være nødvendige for mengdenes tilstrekkelighet. Dermed blir alle tre spillerne nødvendige betingelser for at en mengde er tilstrekkelig for at Jonas dør. Alle blir da årsaker ifølge NESS-teorien, slik intuisjonen vår også krever.

Mer generelt bygger NESS-teorien på en tankegang der måten ϕ skjer på i $\sigma(p)$ kan karakteriseres ved å oppgi alle de fragmentene av p som er tilstrekkelige for å tvinge frem ϕ . Med en slik forståelse av hvordan ϕ skjer i $\sigma(p)$ følger det at v er en nødvendig betingelse for ϕ i $\sigma(p)$ når v kan forhindre minst étt fragment av p fra å fremtvinge ϕ . Dette kan formaliseres på følgende måte.

Definisjon 3. For alle spill G og alle $p \in P, \phi \in L$ definerer vi $N(p, \phi)$ som følger, for alle $v \in V$:

$$v \in N(p, \phi) \Leftrightarrow \exists A \subseteq V : \begin{cases} (a) \forall q \in I(p, A) : \sigma(q) \models \phi \text{ og} \\ (b) \exists r \in I(p, A \cup \{v\}) : \sigma(r) \not\models \phi \end{cases}$$

Vi bruker notasjonen $\bar{A} = V \setminus A$ for alle $A \subseteq V$. Da kan vi lese Definisjon 3 som følger: $v \in N(p, \phi)$ hvis, og bare hvis, det finnes en A slik at \bar{A} er tilstrekkelig for ϕ og $v \in \bar{A}$ er nødvendig for at \bar{A} skal være tilstrekkelig. På denne måten matcher vår definisjon av N den uformelle definisjonen av NESS-teorien gitt i juridisk teori.

Som vist i [5] samsvarer dette også med den siste formelle definisjonen av kausalitet gitt i HP-tradisjonen, når vi begrenser oss til å se på individuelle strukturelle ligninger, ikke systemer av slike. Slik sett kan man anse hele HP-teorien, i sin siste inkarnasjon, som en generalisering av NESS-teorien. Mer sentralt for oss her er at NESS-definisjonen gir oss en alternativ tolkning av hva som egentlig menes med en “nødvendig betingelse” i norsk erstatningsrett.

Det er imidlertid store problemer med NESS-teorien, som vår formalisering lar oss identifisere på en presis måte. For eksempel, anta at Trine ikke gir noe gift, Erna gir en

halv dose, mens Siv gir en hel dose. Er Erna en årsak til at Jonas dør i dette tilfellet? Intuitivt må svaret bli ja: hun gjør det samme nå som i de andre to eksemplene der Jonas dør, og den halve dosen med dødelig gift bidrar til at døden inntreffer selv om den ikke er tilstrekkelig alene. Ifølge NESS-teorien er imidlertid Siv eneste årsak til at Jonas dør i dette tilfellet. Hennes handling er tilstrekkelig alene, slik at Erna sin handling ikke er nødvendig for at noen mengde av hendelser skal være tilstrekkelig for å tvinge frem at Jonas dør. Dette motiverer oss til å se på nok en mulig tolkning av hva betingelseslæren innebærer.

HP-tolkningen

Tilsynelatende hadde vi i NESS-teorien funnet en fornuftig tolkning av hva vi mener når vi snakker om en “nødvendig betingelse” for ϕ på p . Vi kom forbi svakhetene i den naive tolkningen ved å slå fast at måten ϕ oppstod etter p måtte karakteriseres ved mengden av fragmenter av p som var tilstrekkelige til å tvinge frem ϕ . Hvis v var et nødvendig medlem av et fragment i denne mengden, så var v å anse som en årsak av ϕ på p .

Problemene med denne tolkningen oppstår i situasjoner der v er nødvendig for et fragment som tvinger frem ϕ under bestemte forutsetninger. For eksempel, når Erna gir Jonas en halv dose gift så sørger hun for at Jonas fortsatt dør selv om Siv bare gir en halv dose. Dersom Erna ikke hadde gitt gift hadde dette blitt annerledes; da ville Jonas ha overlevd for flere alternative valg Siv kunne ha gjort. Erna sin gift er med andre ord nødvendig for at Jonas ville ha dødd selv om Siv bare hadde gitt en halv dose.

Hvis vi går videre med dette resonnementet ender vi opp med en ny tolkning av nødvendighetskravet: v er en nødvendig betingelse for ϕ på p dersom p er nødvendig for at et fragment av p skal tvinge frem ϕ under en gitt forutsetning. For å formalisere tolkningen må vi si mer presist hva vi mener med en “gitt forutsetning”. Dette gjør at vi trenger litt mer notasjon. For $p, q \in P$ og $A \subseteq V$ lar vi $I(p, q, A)$ denotere den unike profilen r slik at

$$r(i) = \begin{cases} q(i) & \text{dersom } i \in X \\ p(i) & \text{ellers} \end{cases}$$

$I(p, q, X)$ er med andre ord som q for alle variabler i X og som p for alle andre variabler. Vi kan nå gi en definisjon av kausalitet basert på tolkningen nevnt tidligere, der en nødvendig betingelse for ϕ etter p er en variabel med en verdi som er nødvendig for å tvinge frem p under en gitt forutsetning. Formuleringen vi gir nedenfor er basert på en idé av Joseph Halpern.⁶

Definisjon 4. For alle spill G og alle $p \in P, \phi \in L(p)$, definerer vi $H(p, \phi)$ som følger for alle $v \in V$:

$$v \in H(p, \phi) \Leftrightarrow \exists A \subseteq V : \begin{cases} (i) v \in A \text{ og} \\ (ii) \exists q \in I(p, A) : (\sigma(q) \not\models \phi \text{ og } \forall B \subset A : \sigma(I(p, q, B)) \models \phi) \end{cases}$$

⁶Privat kommunikasjon.

Ifølge denne definisjonen er $v \in H(p, \phi)$ hvis, og bare hvis, det finnes en mengde variabler A og en gitt profil q slik at v er nødvendig for at ϕ skal skje selv om verdiene til et vilkårlig utvalg av variabler fra $A \setminus \{v\}$ endres til de verdiene de har i q . Det er verdt å merke seg at definisjonen forlanger at verdiene til $V \setminus (A \setminus \{v\})$ er tilstrekkelige til å frembringe ϕ uansett hvor mange av variablene fra A som endres fra sin verdi i p til sin verdi i q . Vi ville med andre ord fått enda en tolkning av nødvendighetskravet dersom vi hadde nøyd oss med å kreve at v er nødvendig for at ϕ fortsatt ville forekommet dersom *alle* variablene fra $A \setminus \{v\}$ hadde blitt endret til sine verdier i q .

Vi skal komme tilbake til denne tolkningen senere. La oss først slå fast at Definisjon 4 gir oss det vi vil ha i vårt motiverende eksempel. Hvis vi lar $A = \{v_1, v_2\}$ med $q = (0, \frac{1}{2}) \in I(p, A)$ så får vi $\sigma(q) \models \neg\alpha$ samtidig som vi har $\sigma(I(p, q, B)) \models \alpha$ for alle $B \subset A$. Under forutsetning av at Siv gir en halv dose gift, så er Erna sin dose både tilstrekkelig og nødvendig for at Jonas dør. Som vist i [5] samsvarer Definisjon 4 med de to tidligere definisjonene av kausalitet i HP-tradisjonen [7, 8], forutsatt at vi kun ser på enkeltligninger.⁷ Det finnes imidlertid eksempler på at også denne definisjonen blir for snever i sin tolkning av nødvendighetskravet. Et slikt eksempel presenteres i [4]. Formulerer vi dette som en variant av eksempelet vi har sett på tidligere svarer det til en situasjon der Jonas får både en dødelig dose gift og en dødelig overdose av en motgift som ikke virker når det gis for mye av den. La oss anta at Siv gir gift mens Erna og Trine gir en dose motgift hver. Vi antar at Siv kunne unnlatt å gi gift, mens Erna og Trine kunne unnlatt å gi motgift. Dersom Jonas hadde fått gift og nøyaktig én dose motgift ville han overlevd. For de andre tilfellene der Jonas får gift vil han dø. På tross av dette vil Siv sin dose med gift *ikke* bli ansett som en årsak til at Jonas dør. Leseren kan selv sjekke at alle de tre definisjonene vi har sett på så langt er enige om at det er Erna og Trine som er årsak til at Jonas dør i dette tilfellet, mens Siv er uten ansvar. Dette virker feil, noe som indikerer at vi trenger en enda mer vidtfavnende definisjon av kausalitet.

Mulighetstolkningen

Vi skal nå formalisere en tolkning der v er nødvendig for ϕ når det finnes et konkret alternativ for variablene i A slik at v er nødvendig for ϕ når *alle* variablene i A blir tilordnet sin alternative verdi. Interessant nok svarer dette til en tolkning der vi identifiserer måten ϕ skjer på p med mengden av intervensjoner på p som forhindrer at ϕ skjer. Formaliserer vi dette får vi, for alle p, ϕ , mengden $M(p, \phi) = \{q \mid \forall A \subseteq V : \forall r \in I(p, A) : \sigma(r) \models \neg\phi \Leftarrow \sigma(I(q, r, A)) \models \neg\phi\}$. La oss kalle dette for *mekanismen* som produserer ϕ på p . Den består av alle profiler q slik at enhver intervensjon som forhindrer ϕ på q også forhindrer ϕ på p .

For eksempel, anta at q og p begge er profiler der Jonas får en dødelig dose gift av Siv. Forskjellen mellom p og q er at på p får Jonas salt av Erna og Trine, mens på q får han pepper av disse to. Da får vi $M(p, \alpha) = M(q, \alpha)$. Det som skal til for å forhindre at Jonas dør er nøyaktig det samme uavhengig av om han får salt eller pepper av Erna og Trine. Hvis derimot Siv er den eneste som gir en hel dose gift på p mens både Siv og Erna gir hele doser gift på q får vi $M(q, \phi) \subset M(p, \phi)$: det er “enklere” å redde Jonas på p enn på q .

⁷Når vi ser på systemer av ligninger blir disse tre definisjonene forskjellige og videre utforskning av Definisjon 4 i denne sammenhengen har selvstendig interesse.

Vi kan nå definere nok en tolkning av nødvendighetskravet, der v er en nødvendig betingelse for ϕ på p dersom v er en nødvendig betingelse for at $p \in M(p, \phi)$. Med andre ord, v er nødvendig for ϕ dersom en annen verdi for v ville føre til en profil der ϕ kunne ha vært unngått på en annen måte enn i p .

Definisjon 5. For alle spill G og alle p, ϕ definerer vi $S(p, \phi) = \{v \mid \sigma(p) \models \phi \text{ og } \exists q \in I(p, v) : q \notin M(p, \phi)\}$

Legg merke til at definisjonen av $S(p, \phi)$ har nøyaktig samme form som Definisjon 2. Forskjellen ligger i vår forståelse av hva vi ser etter nødvendige betingelser for. Vi har gått fra å se etter nødvendige betingelser for at utfallet tilfredstiller α til å se etter nødvendige betingelser for at α produseres av mekanismen $M(p, \phi)$. I det følgende skal vi vise hvordan disse to tolkningene av nødvendighetskravet kan relateres til hverandre.

4 Tolkningsrommets ytterpunkter

Det viser seg at $S(p, \phi)$ inneholder nøyaktig de variablene i V som har en verdi på p som er en *mulig* but-for årsak av α . Forskjellen mellom tolkningen av nødvendighet i de to definisjonene er forskjellen mellom det som nødvendigvis er nødvendig og det som muligens er nødvendig.

Påstand 1. For alle spill G og alle p, ϕ har vi $S(p, \phi) = \{v \mid \exists q \in I(p, V \setminus \{v\}) : v \in B(q, \phi)\}$

Bevis. \subseteq) Anta at $v \in S(p, \phi)$. Da har vi en $q \in I(p, v)$ slik at $q \notin M(p, \phi)$. Dette betyr at det finnes en $A \subseteq V$ og en $r \in I(p, A)$ slik at (1) $\sigma(r) \models \phi$ og $\sigma(I(q, r, A)) \models \neg\phi$. Siden $q \in I(p, v)$ følger det at $I(q, r, A) \in I(r, v)$. Legg også merke til at $r(v) = p(v)$, siden $r(v) \neq p(v)$ ville medført $v \in A$ slik at $r = I(q, r, A)$ og (1) ikke ville ha vært mulig. Dermed får vi $v \in B(r, \phi)$ med $r \in I(p, V \setminus \{v\})$, slik at påstanden følger.

\supseteq) Anta at det finnes en $q \in I(p, V \setminus \{v\})$ slik at $v \in B(q, \phi)$. Dermed har vi enten $v \in B(p, \phi)$ eller $q \notin M(p, \phi)$. Hvis $q \notin M(p, \phi)$ er vi ferdige, så anta $v \in B(p, \phi)$. Da finnes det $r \in I(p, v)$ slik at $\sigma(r) \models \neg\phi$. Det følger at $r \notin M(p, \phi)$, så påstanden holder. \square

For å illustrere hvor vidtfavnende S er kan vi anta at Jonas får en dødelig dose arsenikk av Siv og halvparten av en dødelig dose blåsyre av Erna. Vi antar at Trine ikke gir noen gift. Her vil S kategorisere Erna sin blåsyre som en årsak til at Jonas dør, forutsatt at det også finnes noen andre (ikke nødvendigvis Siv) som *kunne* ha gitt en halv dose blåsyre. Ved å gi blåsyre forhindrer Erna da at Jonas overlever i den varianten av p der Siv ikke gir arsenikk men det blir gitt en halv dose til av blåsyre. Erna anses dermed som en årsak. Dette er neppe en forsvarlig konklusjon; det var arsenikk som tok livet av Jonas, ikke blåsyre.

På den andre siden, hvis forskjellen mellom å dø av arsenikk og å dø av blåsyre er noe vi kan gjenkjenne og vil legge vekt på, bør vi også ta dette inn vår formelle analyse. Vi bør da ikke spørre hvem som er årsak til α , men hvem som er årsak til α' , der α' representerer påstanden om at Jonas døde av en overdose arsenikk. Hvis α'' er påstanden om at Jonas døde av en overdose blåsyre vil utfallet som sådan kunne beskrives mer nøyaktig på denne måten: $\sigma(p) \models \alpha \wedge \alpha' \wedge \neg\alpha''$. Erna sin blåsyre er ikke en gang en mulig årsak til α' , så i konteksten

der $\neg\alpha''$ også er sann har vi grunnlag for å konkludere med at Erna ikke er en årsak til at Jonas dør på p . Erna er ikke en medvirkende årsak til at Jonas dør *på den måten* han gjør i p .

Slik vil S kunne brukes uten å gi oss for mange årsaker, forutsatt at vi stiller de rette spørsmålene til modellen og tar med all informasjonen vi har om det utfallet vi er interessert i å analysere. Samtidig bør vi være oppmerksomme på at S nødvendiggjør større grad av nyansering i modellene enn de andre kausale teoriene. Slik sett vil vi raskere kunne komme til et akseptabelt resultat dersom vi bruker en mer grovkornet modell sammen med en mer restriktiv definisjon av kausalitet.

Det er neppe grunnlag for å hevde at det finnes en “korrekt” teori blant de vi har gått gjennom så langt. Det vi derimot kan hevde er at vi har gitt en kartlegging av ulike tolkningsmuligheter, som alle kan relateres til den rettslige definisjonen der en årsak er en “nødvendig betingelse” for en egenskap ved verden. I teoriene B og S har vi også identifisert to naturlige ytterpunkter som svarer til den henholdsvis mest skeptiske og minst skeptiske tolkningen vi med rimelighet kan legge til grunn. Det er lett å vise at teoriene vi har sett på i denne artikkelen er relatert som følger, for alle G og alle p, ϕ :

$$B(p, \phi) \subseteq N(p, \phi) \subseteq H(p, \phi) \subseteq S(p, \phi)$$

Denne systematikken indikerer at formell modellering av kausalitet kan være nyttig. Samtidig er det av interesse å spørre etter nye resonneringsstrategier, som lar oss utforske rommet av tolkningsmuligheter på en systematisk måte. Det er også mulig at vi kan finne nye teorier, som er bedre enn de vi har vist her. Definisjonene vi gir vil likevel aldri kunne bli bedre enn modellene de anvendes på. Her ligger også de formelle metodene sin begrensning; det de forteller oss om virkeligheten avhenger av hvordan vi velger å modellere denne. Det subjektive og skjønnsmessige blir vi ikke kvitt, men det kan bli mye tydeligere hva vi mener, samtidig som vi får bedre fram i dagen alle de antagelsene vi har måttet gjøre for å komme til vår konklusjon.

5 Konklusjon

I denne artikkelen har vi vist at kravet til årsakssammenheng i norsk erstatningsrett kan tolkes på forskjellige måter, avhengig av hvordan spørsmålet om det foreligger årsakssammenheng utformes. Videre har vi vist hvordan ulike tolkninger kan formaliseres i et enkelt fragment av HP-teorien, på en måte som tydeliggjør forskjellene mellom dem. Vi har også påvist naturlige ytterpunkter i tolkningsrommet og relatert noen kjente kausalitetsteorier til hverandre etter hvordan de ligger i dette tolkningsrommet.

I fremtidig arbeid vil vi se nærmere på hvordan formelle representasjoner av ulike tolkningsalternativer kan brukes for å besvare spørsmål om årsaker for komplekse hendelser der mange faktorer spiller inn. Her er det spesielt beregningskompleksiteten til ulike definisjoner som må studeres nærmere, også for generaliseringer til (bestemte typer) mengder av strukturelle ligninger. I tillegg vil vi utforske muligheten for å formulere nye tolkninger og resonneringsstrategier for å analysere kausalitet, inspirert av våre formelle definisjoner av nødvendighetskravet i erstatningsretten.

Referanser

- [1] Johs. Andenæs. Konkurrerende skadeårsaker. *Tidsskrift for rettsvitenskap*, pages 241–298, 1941.
- [2] H. Chockler and J. Y. Halpern. Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, 20:93–115, 2004.
- [3] Hana Chockler, Joseph Y. Halpern, and Orna Kupferman. What causes a system to satisfy a specification? *ACM Trans. Comput. Log.*, 9(3):20:1–20:26, 2008.
- [4] Sjur K Dyrkolbotn. On preemption and overdetermination in formal theories of causality. In Alex Groce and Stefan Leue, editors, Proceedings 2nd International Workshop on *Causal Reasoning for Embedded and safety-critical Systems Technologies*, Uppsala, Sweden, 29th April 2017, volume 259 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–15. Open Publishing Association, 2017.
- [5] Sjur K Dyrkolbotn and Truls Pedersen. Actual causality in normal-form games. Manuscript.
- [6] Joseph Y Halpern. *Actual Causality*. MIT Press, 2016.
- [7] Joseph Y. Halpern and Judea Pearl. Causes and explanations: A structural-model approach: Part 1: Causes. In Jack S. Breese and Daphne Koller, editors, *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, University of Washington, Seattle, Washington, USA, August 2-5, 2001*, pages 194–202. Morgan Kaufmann, 2001.
- [8] Joseph Y. Halpern and Judea Pearl. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science*, 56(4):843–887, 2005.
- [9] H. L. A. Hart and Tony Honoré. *Causation in the Law*. Oxford University Press UK, 1959.
- [10] Morten Kjelland. Årsakssammenheng og bevis. In *Personskadeoppgjør*, pages 51–133. BI Bank og Finans, 2015.
- [11] D. Lewis. Causation. *Journal of Philosophy*, 70:556–567, 1973.
- [12] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [13] Magne Strandberg. Effektivitet og kausalitet. *Retfærd*, (1):79–101, 2015.
- [14] Richard W. Wright. Causation, responsibility, risk, probability, naked statistics, and proof: pruning the bramble bush by clarifying the concepts. *Iowa Law Review*, 73(5):1001–1077, 1988.