Mokslo filosofija

DEVYNIŲ KONTRAFAKTINIŲ PRIEŽASTINGUMO TEORIJŲ PALYGINIMAS PASITELKIANT SEPTYNIS PAVYZDŽIUS

Vytautas Grenda
Vilniaus Gedimino technikos universiteto
Filosofijos ir politologijos katedra
Saulėtekio al. 11, LT-10223 Vilnius
Tel. (8 5) 274 48 67
El. paštas: Vytautas.Grenda@hi.vgtu.lt

Straipsnyje lyginamos ir vertinamos devynios per pastarąjį dešimtmetį pasirodžiusios kontrafaktinės vienetinio priežastingumo teorijos, kurias visas sukūrė arba inspiravo D. Lewisas, J. Y. Halpernas ir J. Pearlas, o savo tekstuose išdėstė šeši kiti autoriai. Parodomi kai kurie literatūroje dar neaprašyti šių teorijų skirtumai. Įrodinėjama, kad dauguma šių teorijų intuityviai panašius pavyzdžius nagrinėja skirtingai, ir šiuo požiūriu primeta perskyras, kurių buitinės priežastingumo sampratos požiūriu ne tik nėra, bet ir neturėtų būti.

Pagrindiniai žodžiai: *vienetinis priežastingumas, kontrafaktiniai sąlyginiai teiginiai, „atsarginės" priežastys, persąlygojimas.*

Vienetinis priežastingumas yra dviejų konkrečių įvykių priežastinis ryšys. Kontrafaktinės vienetinio priežastingumo teorijos yra tokios teorijos, kurios remiasi įžvalga, kad jeigu koks nors įvykis *a* yra įvykio *b* priežastis, tai tada yra teisingas kontrafaktinis sąlyginis teiginys (KST) „jei nebūtų įvykęs *a*, tai nebūtų įvykęs ir *b*" (jei nebūtų priežasties, nebūtų ir jos padarinio). Jei ši įžvalga suformuluota be jokių papildomų sąlygų, tai ji yra klaidinga: jei įvykį *b* sukėlė kelios pakankamos priežastys, arba jeigu tuo atveju, kai nebūtų įvykęs *a*, įvykį *b* būtų sukėlusi kita, „atsarginė" priežastis, tada yra klaidinga sakyti, kad jei nebūtų įvykęs *a*, tai nebūtų įvykęs ir *b*.

Tačiau tai nereiškia, kad kontrafaktinės teorijos iš anksto pasmerktos žlugti. Pagrindinę šių teorijų įžvalgą galima mėginti apginti papildant ją įvairiomis sąlygomis: pavyzdžiui, jei greta priežasties *a* tam tikru atveju įvyko dar viena pakankama įvykio *b* priežastis *c*, tai tada atrodo teisinga, kad *b* nebūtų įvykęs tuo atveju, jei nebūtų įvykęs nei įvykis *a*, nei įvykis *c*, tačiau būtų, jei būtų įvykęs tik *a*, bet ne *c*.

Šiame straipsnyje lyginamos ir kritiškai vertinamos devynios per pastarąjį dešimtmetį pasirodžiusios teorijos, kurios mėgina tiksliai suformuluoti šią pagrindinę idėją (galbūt netgi būtų galima sakyti, kad tai tos pačios teorijos devyni skirtingi variantai).

Jos yra išdėstytos penkiuose šešių autorių tekstuose (Glymour & Wimberly 2007; Hall 2007; Hitchcock 2001; Menzies 2004; Woodward 2003). Dalis šių autorių savo pristatomoms teorijoms nepritaria: vieni aprašo jas tik kaip kritikos objektą (Hall 2007: 115-116; Menzies 2004: 820-823), kiti pabrėžia, kad jos nėra galutinės ir teisingos (Glymour & Wimberly 2007: 65; Woodward 2003: 85-86).

Plėtojant šiame straipsnyje nagrinėjamo tipo teorijas daugiausia nusipelnė J. Y. Halpernas ir J. Pearlas. Vienas iš čia cituojamų autorių (Menzies 2004: 820–823) priskiria vieną iš savo pristatomų teorijų Halpernui ir Pearlui (nesisavina jos autorystės). Kiti du nors ir teigia esą savo pristatomų teorijų autoriai, tačiau nurodo, kad jos panašios į Halperno ir Pearlo teoriją (Hitchcock 2001: 274, 289; Woodward 2003: 382, išn. 41). Tačiau galima numanyti, kad šiuo metu Halpernas ir Pearlas nė vienai iš čia lyginamų teorijų nepritartų, nes yra pasiūlę kitą, sudėtingesnę, kurią laiko pranašesne už visas šio tipo teorijas (Halpern & Pearl 2005: 845, 854, išn. 10, 855, 881–882)[1].

Iš gausybės pastaraisiais dešimtmečiais pasirodžiusių kontrafaktinių vienetinio priežastingumo teorijų šiame tyrime buvo pasirinktos tokios, kurias patogu palyginti, nes jos visos (1) turi panašią struktūrą, (2) yra suformuluotos pakankamai tiksliai ir (3) yra palyginti paprastos, dėl to jas įmanoma trumpai pristatyti.

Literatūroje yra ir daugiau teorijų, kurias galima priskirti tam pačiam tipui, bet jos arba yra daug sudėtingesnės (Glymour & Wimberly 2007: 62–63; Halpern & Pearl 2005; Pearl 2000: 309–329), arba reliatyvizuoja priežastingumą tam tikrų papildomų pragmatinių faktorių atžvilgiu (Hall 2007: 127–133; Menzies 2004: 826–832), dėl to jų dėl vietos stokos čia neapibūdinsime ir nepalyginsime.

Šio straipsnio tikslas yra pasitelkiant kuo mažiau kuo paprastesnių pavyzdžių atskleisti dalį teorijų skirtumų, kurie lig šiol nebuvo aprašyti literatūroje, ir įvertinti jas intuicijų apie šiuos pavyzdžius atžvilgiu. Tam bus pasitelkiami septyni pavyzdžiai.

Faktas, kad dauguma straipsnyje aprašytų teorijų skirtumų cituojamoje literatūroje nėra užfiksuoti (nors naudojami pavyzdžiai nėra sudėtingi), rodo, kad šioje srityje nuveikta palyginti nedaug. Nė viename iš cituojamų tekstų nėra palygintos daugiau negu trys iš čia lyginamų teorijų.

Vienetinio priežastingumo teorijos kūrimas yra svarbus uždavinys ne tik dėl to, kad ji būtina norint geriau suprasti ontologinį vienetinio priežastingumo ir tipų lygmens priežastingumo santykį (žr. Grenda 2006: 30 ir toliau), bet ir dėl to, kad gali būti pasitelkiama analizuojant priežastinio paaiškinimo, atsakomybės, stebėjimo ir kitas priežastines sąvokas.

Klausimas, ar šiame straipsnyje lyginamos teorijos skirtingos, – visiškai formalus. Taikant šias teorijas formalizuotam pavyzdžiui, jos visada teikia apibrėžtą atsakymą. Formalizuotuose pavyzdžiuose aprašomų įvykių, kuriuos teorija tvirtina esant priežastingai susijusius, porų aibė toliau bus vadinama teorijos *ekstensija*. Teorijos yra skirtingos, jei ir tik jei skirtingos jų ekstensijos.

Nepaisant to, ar įmanoma tokia viena priežastingumo teorija, kurią būtų galima

[1] Nagrinėjant šio straipsnio pavyzdžius Halperno ir Pearlo 2005 m. teorija niekuo nesiskirtų nuo toliau aptariamos teorijos T7.

pavadinti „teisinga“ ar „geriausia“, jau egzistuojančių teorijų skirtumų aprašymas svarbus filosofijos istorijos požiūriu, taip pat naudingas tuo, kad bent iš dalies parodo priežastingumo pavyzdžių įvairovę.

Taip pat neatsižvelgiant į tai pagrįsta kelti reikalavimą, kad teorija patikslintų egzistuojančią priežastingumo sampratą, suteiktų apibrėžtumo ten, kur jo trūksta, bet ne šią sampratą savavališkai iškreiptų. Jeigu yra tokių intuityviai panašių atvejų, kad nagrinėjant jų sutampančius aspektus teorija duoda skirtingus atsakymus (nurodo esant skirtingus priežastinius ryšius), tai aiškus ženklas, kad ji primeta perskyras, kurių buitinės priežastingumo sampratos požiūriu ne tik nėra, bet ir neturėtų būti. Būtent šiuo būdu 3 ir 4 skyreliuose bus parodyti daugumos lyginamų teorijų trūkumai.

Skirtingos šio straipsnio išvados priklauso nuo skirtingų intuicijų (šiuo atžvilgiu jų patikimumas gali būti ribotas). Išvados apie formalius teorijų skirtumus nuo jokių intuicijų nepriklauso. Išvados, kad viena ar kita teorija nagrinėjant intuityviai panašius atvejus duoda skirtingus atsakymus, yra priklausomos nuo intuicijų, kad aptariami pavyzdžiai iš tikrųjų relevantiškais atžvilgiais panašūs, tačiau nepriklausomos nuo intuicijų, kas nagrinėjant tuos pavyzdžius laikytina priežastimi ir kas ne. Nuo pastarųjų taip pat nepriklauso išvados, kuriose nurodoma, ko trūksta aptartoms teorijoms, kad jos nagrinėjant vienus ar kitus pavyzdžius teiktų tuos atsakymus, kurie tekste vadinami *intuityviai teisingais*.

Straipsnio pirmame skyrelyje supažindinama su tolesniam dėstymui būtinomis sąvokomis, antrame aprašomos visos devynios lyginamos teorijos (visi lyginimui reikalingi jų panašumai ir skirtumai pateikiami viena lentele), trečiame ir ketvirtame teorijos lyginamos pasitelkiant pavyzdžius. Galiausiai apibendrinama, kaip derėtų nagrinėti aptartus pavyzdžius, kad gautume pageidaujamus atsakymus.

1. Priežastiniai modeliai, grafai ir KST teisingumo sąlygos

Šiame skyrelyje bus pristatytos visose nagrinėjamose teorijose vartojamos pagrindinės sąvokos ir tiksliau apibrėžta nagrinėjama problema. Visą šią informaciją ne tokia glausta forma skaitytojas gali rasti J. Pearlo arba J. Woodwardo knygose (Pearl 2000: 202–207; Woodward 2003: 38 ir toliau), taip pat C. Hitchcocko straipsnyje (Hitchcock 2001).

Čia nagrinėjamos teorijos remiasi prielaida, kad atsakymas į klausimą „ar įvykis *a* buvo įvykio *b* priežastis?“ vienareikšmiškai priklauso nuo to, kokie kiti įvykiai įvyko nagrinėjamu atveju, ir nuo to, kokie įvykiai būtų įvykę tam tikrais kitais panašiais galimais atvejais. Įprastinis tikslus šios informacijos perteikimo būdas, kurį kuriant daug nuveikė ir kurį priežastingumo filosofijoje išpopuliarino J. Pearlas, naudojamas ne tik diskusijose apie vienetinį priežastingumą, bet ir svarstant įvairius kitus priežastingumo filosofijos klausimus, yra *priežastinis modelis*.

Priežastinis modelis M yra trejetas $<U, V, F>$; čia U ir V yra bendrų elementų neturinčios baigtinės netuščios kintamųjų aibės, o F – funkcijų aibė. Šios aibės atitinka tokius reikalavimus:

- Kiekvieną aibės V kintamąjį X atitinka viena ir tik viena aibės F funkcija f_X, kuri priskiria kintamajam X reikšmę

pagal visų likusių aibėms U ir V priklausančių kintamųjų (t. y. aibės $U \cup (V \setminus \{X\})$ kintamųjų) reikšmes: f_X apibrėžimo sritis yra $U \cup (V \setminus \{X\})$ kintamųjų reikšmių visų įmanomų derinių aibė, o reikšmių sritis yra X reikšmių aibė.

- Aibė F yra pakankama, kad kiekvienas aibės U kintamųjų reikšmių derinys priskirtų kiekvienam aibės V kintamajam X apibrėžtą reikšmę.

U ir V kintamieji vaizduoja įvykių įvykimą arba neįvykimą. Šiame straipsnyje bus nagrinėjami tik modeliai su dvireikšmiais kintamaisiais. Kiekvieno kintamojo reikšmė gali būti 1 arba 0. Tai, kad kintamojo X reikšmė yra 1, vaizduoja, kad tam tikras įvykis įvyksta, o tai, kad X reikšmė yra 0, vaizduoja, kad jis neįvyksta. Toliau kintamieji ir jų aibės bus žymimi didžiosiomis raidėmis, o jų reikšmės – mažosiomis: X yra kintamasis, x ir x' – jo reikšmės, $X = x$ ir $X = x'$ – įvykiai. Taip pat mažosios raidės bus naudojamos kintamųjų reikšmių deriniams vaizduoti. Pavyzdžiui, $U = u$ žymi įvykį, kurio metu aibės U kintamieji įgijo reikšmes, nurodytas derinyje u. Tos reikšmės, kurias modelio kintamieji įgyja nagrinėjamame pavyzdyje, vadinamos *faktinėmis* (angl. *actual*) reikšmėmis.

Aibės F funkcijos aprašo, kaip vieni modelio kintamieji priklauso nuo kitų jo kintamųjų. Jos vaizduoja priežastinius dėsnius, kuriems paklūsta modelyje vaizduojami įvykiai (arba dėsningus reguliarumus).

U kintamųjų (jie literatūroje vadinami egzogeniškais, angl. *exogenous*) reikšmes lemia priežastiniai veiksniai, kurie modelyje nėra aprašyti[2]. U kintamųjų reikšmių derinys u apibūdina nagrinėjamo proceso pradines sąlygas. V kintamųjų reikšmes lemia modelyje aprašyti kintamieji. Tai reiškia, kad jei U kintamųjų reikšmės bei F funkcijos yra žinomos, tada įmanoma nustatyti bet kurio aibei V priklausančio kintamojo X reikšmę. Reikšmė, kurią modelis M esant pradinėms sąlygoms u priskiria kintamajam X, yra ta reikšmė, kurią nurodo funkcija f_X.

Kiekvieną šiame straipsnyje nagrinėjamą priežastingumo teoriją galima interpretuoti kaip taisyklę, nurodančią atsakymą į klausimą, kurio forma tokia: ar įvykis $X = x$ yra įvykio $Y = y$ priežastis modelio M ir pradinių sąlygų u atžvilgiu?

Naudojantis modelyje glūdinčia informacija ir žinant pradines sąlygas, galima apskaičiuoti ne tik faktines, bet ir kontrafaktines kintamųjų reikšmes, t. y. reikšmes, kurias jie įgytų tuo atveju, jei pradinės sąlygos būtų kitokios nei yra, arba jei dalis modelio kintamųjų įgytų reikšmes, kurios yra nesuderinamos su pradinėmis sąlygomis ir aibėje F glūdinčia informacija apie dėsnius. Kintamojo Y kontrafaktinė reikšmė, kurią Y įgytų tuomet, jei X įgytų reikšmę x, apskaičiuojama pasitelkus modelio *modifikaciją*.

Priežastinio modelio $M = <U, V, F>$ *modifikacija įvykio* $X = x$ *atžvilgiu* yra priežastinis modelis $M_x = <U, V, F_x>$; čia F_x yra funkcijų aibė, sutampanti su F visų elementų atžvilgiu, išskyrus funkciją f_X, kuri pakeičiama funkcija $X = x$.

[2] Jei tartume, kad įmanomi tokie įvykiai, kurie neturi priežasties, tada kai kurių U kintamųjų reikšmės gali būti apskritai nenulemtos.

Atlikus tokią modifikaciją, galima apskaičiuoti kintamųjų kontrafaktines reikšmes, taip pat KST teisingumo reikšmes (kurios yra reliatyvios M ir u atžvilgiu).

> *Reikšmė, kurią Y įgytų tuo atveju, jei X įgytų reikšmę x, modelio M ir pradinių sąlygų u atžvilgiu* yra ta reikšmė, kurią kintamajam Y priskiria modelis M_x esant pradinėms sąlygoms u. KST, kurio forma yra „jei būtų įvykęs įvykis $X = x$, tai būtų įvykęs įvykis $Y = y$“, M ir u atžvilgiu teisingas tada ir tik tada, kai ši Y reikšmė yra y.

Analogiškai apibrėžiamos ir tų KST, kurių antecedente konstatuojamas daugiau nei vienas įvykis, teisingumo reikšmės – tuomet modelio modifikacija gaunama pakeičiant daugiau negu vieną F funkciją.

Kai KST teisingumo sąlygos yra apibrėžtos čia nurodytu būdu, juos galima interpretuoti kaip teiginius apie galimas „intervencijas“ – įvykius, kurie, jei įvyktų, paveiktų KST antecedente nurodytus nagrinėjamos priežastinės sistemos kintamuosius, tiesiogiai nepaveikdami kitų jos kintamųjų ir nepakeisdami dėsnių, kuriems ji paklūsta (žr. Woodward 2003: 94 ir toliau).

Kintamojo Y reikšmė y, kurios atžvilgiu teisingas KST „jei $X = x$, tai $Y = y$“, toliau bus vadinama „reikšme, kurią Y įgytų priskyrus X fiksuotą reikšmę x“. Jei yra teisingas toks KST, o x ir y yra kontrafaktinės reikšmės, tuomet Y vadinamas *kontrafaktiškai priklausomu* nuo X.

Priežastinį modelį ir pradines sąlygas literatūroje įprasta vaizduoti *lygčių sistemos* forma, o modelio modifikacijas – lygčių sistemos modifikacijų forma. Šios lygtys vadinamos struktūrinėmis lygtimis (angl. *structural equations*). Modelį M ir pradines sąlygas u atitinkančioje lygčių sistemoje kiekvieną U ir V kintamąjį X atitinka viena ir tik viena lygtis L_X. Lygties forma gali būti dvejopa, ir šią formą lemia tai, ar lygtis atitinka U kintamąjį, ar V kintamąjį:

1. Jei X priklauso U, tai L_X forma yra $X = x$.
2. Jei X priklauso V, tai L_X forma yra $X = f(Y_1, ..., Y_n)$.

Jei kintamasis X priklauso V ir jo lygtis L_X yra (2) tipo, tai paisoma tokio susitarimo: šios lygties dešinėje pusėje nurodyti tik tiek kintamųjų, kiek būtina X reikšmei apskaičiuoti. Pavyzdžiui, jei visoms y, z, z' teisinga, kad $f_X(y, z) = f_X(y, z')$, tai X lygties forma turi būti $f(Y)$.

Kita literatūroje naudojama vaizdinė priemonė daliai modelio lygtyse glūdinčios informacijos perteikti yra *orientuotieji grafai*. Tokį grafą sudaro aibė *viršūnių*, kurios vaizduoja U ir V kintamuosius, ir aibė *lankų*, kurie jungia viršūnes ir kurie vaizduoja modelyje aprašytus kintamųjų santykius. Kiekvieną lanką galima apibūdinti kaip viršūnių porą. Pora $\langle X, Y\rangle$ priklauso modelio M grafui tada ir tik tada, kai lygties L_X dešinėje pusėje yra kintamasis Y, t. y. jei Y reikšmė būtina norint apskaičiuoti X. Grafas vaizduojamas schema, kurioje lankai žymimi rodyklėmis, jungiančiomis viršūnes (pavyzdys – 1 pav.).

Jei grafui priklauso lankas $\langle X, Y\rangle$, viršūnė X vadinama *tėvine Y viršūne* (angl. *parent*). Tą pačią kryptį turinčių lankų seka, jungianti viršūnes X ir Y, t. y. seka $\langle\langle X, Z_1\rangle, \langle Z_1, Z_2\rangle, ..., \langle Z_n, Y\rangle\rangle$ (kur $n \geq 0$), vadinama *X ir Y jungiančiu orientuotu keliu* arba tiesiog *keliu*. Jei X ir Y jungia orientuotas

kelias, tai Y vadinama viršūnės X *palikuone* (angl. *descendant*).

Štai standartinis pavyzdys, iliustruojantis pirmiau apibūdintas sąvokas (Woodward 2003: 44).

> P1. Tarkime, kintamasis X įgyja reikšmę 1, kai aplinkoje yra deguonies, ir reikšmę 0, kai jo nėra. Y įgyja reikšmę 1, jei įvyksta trumpasis jungimas. Z įgyja reikšmę 1, jei įvyksta gaisras. Galioja toks dėsningumas: gaisras įvyksta tada ir tik tada, kai aplinkoje yra deguonies ir įvyksta trumpasis jungimas. Nagrinėjamu atveju aplinkoje yra deguonies, įvyksta trumpasis jungimas ir įvyksta gaisras.
>
> Šią situaciją galima atvaizduoti tokia lygčių sistema (ją atitinka 1 pav. grafas):

$$X = 1$$

$$Y = 1$$

$$Z = X \& Y$$

$$X \longrightarrow Z \longleftarrow Y$$

1 pav.

Šis pavyzdys įdomus tuo, kad visos toliau aptariamos teorijos sutaria dėl to, kad įvykio $Z = 1$ priežastimis čia reikia vadinti $X = 1$ ir $Y = 1$.

Kodėl, pavyzdžiui, visos teorijos teigia, kad $Y = 1$ (trumpasis jungimas) yra $Z = 1$ (gaisro) priežastis? Dėl to, kad visų šių teorijų siūlomi priežastingumo kriterijai nagrinėjant šių įvykių santykį pasidaro labai panašūs. Vienos iš jų nurodo, kad turi būti įvykdyta (1) iš toliau nurodytų sąlygų, kitos – kad (2), trečios – kad (3):

1. Jei nebūtų įvykęs $Y = 1$, tai nebūtų įvykęs $Z = 1$.
2. Jei nebūtų įvykę $X = 1$ ir $Y = 1$, tai nebūtų įvykęs $Z = 1$.
3. (1) arba (2).

Pirmiau apibrėžtas KST teisingumo sąlygas patogu performuluoti taip, kad KST teisingumą ar klaidingumą būtų galima patikrinti naudojantis vien pateiktoje lygčių sistemoje glūdinčia informacija. Reikšmė, kurią kintamajam Y priskiria modelis M esant pradinėms sąlygoms u, apskaičiuojama taip:

1. Jei lygties L_Y forma yra $Y = y$, tai Y priskiriama reikšmė y.
2. Jei L_Y forma yra $Y = f_Y(X_1, ..., X_n)$, tai pirma išsprendžiamos $X_1, ..., X_n$ lygtys, o tada naudojant gautus atsakymus Y reikšmė apskaičiuojama pagal L_Y.

Vaizdžiai tariant, lygčių sistema visada sprendžiama „iš dešinės į kairę“. Tokiu sprendimo būdu atsižvelgiama į lygtyse glūdinčią informaciją apie priežastingumo kryptį. Lygybės ženklas šiose lygtyse vaizduoja asimetrišką kintamųjų priklausomybės santykį.

Norint nustatyti KST teisingumo reikšmę, reikia išspręsti modifikuotą lygčių sistemą, kuri atitinka modifikuotą modelį. Kad nustatytume, ar teisingas KST, kurio forma yra „jei $X = x$, tai $Y = y$“, reikia modifikuoti lygčių sistemą pakeičiant joje lygtį L_X lygtimi, kurios forma yra $X = x$. Toks lygčių sistemos modifikavimas toliau bus vadinamas „fiksuotos reikšmės x priskyrimu kintamajam X“.

Jeigu KST antecedente nurodomi keli įvykiai, tuomet reikia modifikuoti kelias nagrinėjamos sistemos lygtis. Pavyzdžiui, tarkime, kad P1 pavyzdyje reikia nustatyti, ar teisingas KST „jei nebūtų įvykę įvykiai

$X = 1$ ir $Y = 1$, tai nebūtų įvykęs $Z = 1$“ („jei nebūtų buvę nei deguonies, nei trumpojo jungimo, tai nebūtų buvę gaisro“). Pagal šio KST antecedentą modifikavę turimą lygčių sistemą, gauname:

$$X = 0$$

$$Y = 0$$

$$Z = X\&Y$$

Išsprendus šią sistemą matyti, kad $Z = X\&Y = 0$. Vadinasi, nagrinėjamas KST yra teisingas.

Pirmiau išdėstyta priežastinio modelio samprata yra pakankama šio straipsnio reikmėms. Literatūroje (pvz., Halpern & Pearl 2005; Pearl 2000) dažnai aptinkami ir sudėtingesni priežastiniai modeliai: vietoje dvireikšmių kintamųjų juose gali būti naudojami daugiareikšmiai, vietoje kintamųjų deterministinių ryšių – tikimybiniai ryšiai ir kt. Į juos šiame straipsnyje nebus atsižvelgta, nes tik apsunkintų analizę.

2. Lyginamų teorijų pristatymas

Pasitelkus pirmame skyrelyje apibrėžtas sąvokas galima teigti, jog visų čia nagrinėjamų teorijų, išskyrus teoriją, kuri bus vadinama T1, bendra forma tokia:

> (KT) Įvykis $X = x$ yra įvykio $Y = y$ priežastis modelio M ir pradinių sąlygų u atžvilgiu, jei ir tik jei faktinė X reikšmė yra x, faktinė Y reikšmė yra y, modelio M grafe egzistuoja viršūnes X ir Y jungiantis kelias K ir egzistuoja tokios (tam tikrus reikalavimus atitinkančios) kintamųjų aibės F, Z ir tokie (tam tikrus reikalavimus atitinkantys) F, Z kintamųjų reikšmių deriniai f ir z, kurių atžvilgiu yra teisingas toks KST: jei $X = x'$ ir $F = f$, tai $Y = y'$ ir $Z = z$.

Čia x' ir y' yra kontrafaktinės X ir Y reikšmės. Nagrinėjamos teorijos skiriasi tuo, kaip jos apibrėžia aibes F ir Z bei derinius f ir z.

Teorijoje T1 KT formos apibrėžimas yra ne vienetinio priežastingumo apibrėžimas, bet priežastinės priklausomybės (angl. *causal dependence*) santykio apibrėžimas. $X = x$ šioje teorijoje yra $Y = y$ priežastis, jei ir tik jei juos sieja priežastinių priklausomybių grandinė. Kai M grafe viršūnė X yra viršūnės Y tėvinė viršūnė ir T1 požiūriu $X = x$ yra $Y = y$ priežastis, tai priežastinės priklausomybės santykis ir vienetinio priežastingumo santykis sutampa.

Kiti teorijų skirtumai nurodyti 1 lentelėje.

Visose teorijose, išskyrus T1 ir T8, aibė F apibrėžiama taip, kad jos apibrėžimą gali atitikti (pagal tai, kaip atrodo grafas) daugiau nei viena aibė. Tada $X = x$ laikomas $Y = y$ priežastimi, jei KT formos apibrėžime nurodytos sąlygos tenkinamos bent su viena aibe, atitinkančia F apibrėžimą. Pavyzdžiui, teorijoje T4 aibė F apibrėžiama kelio K, grafe jungiančio kintamuosius X ir Y, atžvilgiu. Todėl jei egzistuoja daugiau nei vienas toks kelias, tada teorijoje nurodytą F apibrėžimą atitinka daugiau nei viena aibė.

Teorijose T5–T9 reikšmių derinys f, kaip ir aibė F, apibrėžiamas taip, kad jo apibrėžimą gali atitikti daugiau nei vienas derinys. $X = x$ laikomas $Y = y$ priežastimi, jei KT formos apibrėžime nurodytos sąlygos tenkinamos bent su vienu f apibrėžimą atitinkančiu deriniu.

Istorinė informacija: teorijos T2 ir T6 aprašytos C. Hitchcocko (Hitchcock 2001:

1 lentelė. Teorijų T1–T9 skirtumai pagal formą KT. T1 eilutė atspindi priežastinės priklausomybės, kitos eilutės – vienetinio priežastingumo apibrėžimus. Faktinės X, Y, K reikšmės pažymėtos x, y, k. K\{X} yra aibė, kuriai priklauso visi K kintamieji, išskyrus X; k_x yra jos elementų faktinės reikšmės. E yra visų egzogeniškų kintamųjų aibė

T1	*F* tuščia aibė	*f*	*Z*, *z* aibė *Z* tuščia
T2	visi kintamieji, nepriklausantys *K*, bet priklausantys kokiam nors kitam *X* ir *Y* jungiančiam keliui	faktinės reikšmės	kaip T1
T3	bet kokia *K* nepriklausančių kintamųjų aibė (gali būti tuščia); t. y. bet koks $(U \cup V) \backslash K$ poaibis	kaip T2	kaip T1
T4	visos *Y* tėvinės viršūnės, nepriklausančios *K*	kaip T2	kaip T1
T5	kaip T4	reikšmių derinys, kuriam teisinga „jei $F = f$ ir $X = x$, tai $Y = y$“	kaip T1
T6	kaip T3	reikšmių derinys, kuriam teisinga „jei $F = f$, tai $K\backslash\{X\}$ reikšmės būtų k_x“	kaip T1
T7	kaip T3	reikšmių derinys, kuriam teisinga „jei $F = f$ ir $X = x$, tai $K = k$“	kaip T1
T8	*E*	reikšmių derinys, kuriam teisinga „jei $F = f$, tai $K = k$“	$Z = K\backslash\{X, Y\}$; *z* – kontrafaktinės reikšmės
T9	bet kokios *X* palikuonių, nepriklausančių *K*, aibės (ji gali būti tuščia) ir *E* sąjunga	kaip T8 (*E* kintamiesiems), kaip T2 (likusiems kintamiesiems)	kaip T8

286–287, 290), T8 ir T9 – C. Glymouro ir F. Wimberly'o straipsniuose (Glymour & Wimberly 2007: 53–58), T3 – P. Menzieso (Menzies 2004: 820–823) ir N. Hallo (Hall 2007: 115–116) straipsniuose (Menziesas priskiria T3 J. Y. Halpernui ir J. Pearlui, kaip šaltinį nurodydamas šių dviejų autorių neskelbtą 2001 m. rankraštį). Šiame N. Hallo straipsnyje aprašyta ir T7 (Hall 2007: 116). N. Hallas tvirtina, kad tai esanti Halperno ir Pearlo teorijos (Halpern & Pearl 2005) supaprastinta versija. Teorijos T4 ir T5 pateiktos J. Woodwardo knygoje (Woodward 2003: 77, 84).

T1 yra D. Lewiso 1973 m. kontrafaktinės teorijos (Lewis 1986 (1973)) atitikmuo, suformuluotas pasitelkus pirmame skyrelyje išdėstytą priežastinių modelių sampratą ir priėmus teorijoms T2–T9 bendras prielaidas. Šis atitikmuo aptariamas C. Hitchcocko straipsnyje (Hitchcock 2001), kur T1 lyginama su T2, nors eksplicitiškos T1 formuluotės jame nėra.

3. „Atsarginės“ priežastys ir persąlygojimas (paprasčiausi atvejai)

Dabar pateiksime du standartinius pavyzdžius, kurie jau yra daugybę kartų aprašyti

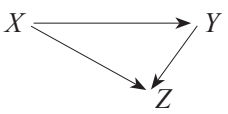
literatūroje ir kurie parodo tiek nagrinėjamų teorijų skirtumų, kad jas galima suskirstyti į keturias grupes.

> P2 (Hitchcock 2001: 276). Du žudikai – mokytojas ir mokinys – suplanuoja nužudymą. Suplanuota, kad jei mokinys atėjus numatytam laikui nepaleistų kulkos į auką, tada kulką paleistų mokytojas. Viskas vyksta pagal planą, mokinys atėjus laikui iššauna ir auka žūva. Jokių mokytojo veiksmų neprisireikia. Ar mokinio šūvis yra aukos žūties priežastis? Intuityviai atrodytų, kad taip; tačiau KST „jei mokinys nebūtų šovęs, auka nebūtų žuvusi“ klaidingas. Tą patį padarinį būtų sukėlusi „atsarginė“ priežastis (mokytojo šūvis).
>
> Šią situaciją galima modeliuoti kintamaisiais X (mokinio šūvis), Y (mokytojo šūvis), Z (aukos žūtis). Šį modelį ir pavyzdyje aprašytas pradines sąlygas galima perteikti 2 pav. grafu ir tokia lygčių sistema:

$$X = 1$$

$$Y = \sim X$$

$$Z = X \vee Y$$



2 pav.

T1 apie P2 teigia, kad $X = 1$ nebuvo $Z = 1$ priežastis, nes T1 požiūriu šis priežastinis ryšys egzistuotų tada ir tik tada, jei būtų teisinga viena iš dviejų: (1) „jei $X = 0$, tai $Z = 0$“; (2) „jei $X = 0$, tai $Y = 1$“ ir „jei $Y = 1$, tai $Z = 0$“. (1) atveju priežastingumas sutaptų su priežastine priklausomybe, (2) atveju ne. Tačiau (1) ir (2) teiginiai klaidingi.

T2 teigia, kad $X = 1$ buvo $Z = 1$ priežastis, nes egzistuoja kelias $\langle X, Z\rangle$, kurį nagrinėjant T2 nurodo alternatyviam keliui $\langle X, Y, Z\rangle$ priklausančiam kintamajam Y priskirti fiksuotą reikšmę 0 (faktinę reikšmę). Kitaip tariant, nagrinėjant $\langle X, Z\rangle$ teorija T2, kitaip nei T1, teigia, kad $X = 1$ buvo $Z = 1$ priežastis, jei ir tik jei teisinga „jei $Y = 0$ ir $X = 0$, tai $Z = 0$“ (jei nebūtų šovęs nei vienas, nei kitas žudikas, auka būtų likusi gyva). Tai teisingas KST. Šį T1 ir T2 skirtumą aprašo C. Hitchcockas (Hitchcock 2001: 283-289).

T3–T7, T9 pavyzdį P2 nagrinėja iš esmės taip pat kaip ir T2, tik kai kurios iš jų greta to KST, kurį tikrina T2, tikrina ir kai kuriuos kitus KST.

Kadangi vienintelis P2 egzogeniškas kintamasis yra X ir, priskyrus jam fiksuotą reikšmę 1, visų kelio $\langle X, Z\rangle$ kintamųjų reikšmės išlieka tokios pat kaip ir faktinės jų reikšmės, o priskyrus X reikšmę 0, jos neišliktų tokios pat (juk pats X pasikeistų), tai T8 tikrindama $X = 1$ ir $Z = 1$ priežastinį ryšį tikrina tik (klaidingą) KST „jei $X = 0$, tai $Z = 0$“. Dėl to T8 teigia, kad $X = 1$ nebuvo $Z = 1$ priežastis[3].

> P3 (Woodward 2003: 82). Du stovyklautojai miške numeta po smilkstančią cigaretę. Gaisrui sukelti būtų pakakę vien pirmosios arba vien antrosios cigaretės.

[3] Įdomu, kad jei pradinės sąlygos pavyzdyje P2 būtų $X = 0$, tai T1 teigtų, kad $X = 0$ buvo $Z = 1$ priežastis (nes KST „jei $X = 1$, tai $Y = 0$“ ir „jei $Y = 0$, tai $Z = 0$“ būtų teisingi), o T2–T9 tai neigtų (bet visų devynių teorijų požiūriu $X = 0$ būtų $Y = 1$ priežastis, o $Y = 1$ būtų $Z = 1$ priežastis). Ar tai liudija T1, ar T2–T9 naudai? Tegul skaitytojas sprendžia pats.

Ar tai, kad pirmasis stovyklautojas numetė cigaretę, buvo gaisro priežastis?

Tarkime, $X = 1$ yra įvykis „cigaretę numeta pirmasis stovyklautojas“, $Y = 1$ – „cigaretę numeta antrasis stovyklautojas“, $Z = 1$ – gaisro įvykis. Grafas pateikiamas 3 pav. Lygčių sistema:

$$X = 1$$

$$Y = 1$$

$$Z = X \vee Y$$

$$X \longrightarrow Z \longleftarrow Y$$

3 pav.

Intuityviai atrodo teisinga sakyti, kad gaisras ($Z = 1$) turėjo dvi priežastis – $X = 1$ ir $Y = 1$, iš kurių kiekviena atskirai būtų buvusi pakankama tam pačiam padariniui sukelti. Tokia situacija literatūroje vadinama persąlygojimu (angl. *overdetermination*). T1–T4 čia teigia, kad $X = 1$ nebuvo $Z = 1$ priežastis. T1 ir T2 prieina šią išvadą tikrindamos KST „jei $X = 0$, tai $Z = 0$“, T4 – tikrindama „jei $X = 0$ ir $Y = 1$, tai $Z = 0$“, T3 – tikrindama abu šiuos KST. Abu jie klaidingi.

T5–T9 nagrinėdamos P3 prieina intuityviai teisingą atsakymą, kad $X = 1$ buvo $Z = 1$ priežastis, nes tikrina KST „jei $X = 0$ ir $Y = 0$, tai $Z = 0$“ (ir taip pat kai kuriuos kitus KST, bet šiam atsakymui prieiti pakanka šio vieno). T5–T9 požiūriu tikrinant, ar $X = 1$ buvo $Z = 1$ priežastis, fiksuotos reikšmės 0 priskyrimas kintamajam Y yra leistinas, nes nepakeičia kelio $\langle X, Z \rangle$ kintamųjų reikšmių: yra teisingas KST „jei $Y = 0$, tai $X = 1$ ir $Z = 1$“.

Lygindamas T3 ir T7, šį pavyzdį aprašo N. Hallas (Hall 2007: 116-117); lygindamas T1, T2 ir T6 – C. Hitchcockas (Hitchcock 2001: 289–290), lygindamas T4 ir T5 – J. Woodwardas (Woodward 2003: 82–84).

P2 ir P3 aptarimas rodo, kad teorijos T1–T9 gali būti suskirstytos į keturias grupes (žr. 2 lentelę).

4. Sudėtingesni „atsarginių“ priežasčių ir persąlygojimo atvejai

Toliau bus aprašyti pavyzdžiai, intuityviai panašūs į P2 ir P3, tačiau atskleidžiantys papildomus T1–T9 skirtumus. Šie pavyzdžiai cituojamoje literatūroje nėra aprašyti. Tiesa, galima aptikti vieną pavyzdį, *panašų* į P4 (Glymour & Wimberly 2007: 45, pavyzdžio numeris 3.2), ir vieną, panašų į P7 (Pearl 2000: 207). Tačiau ten tų pavyzdžių pagrindu neatliekami tokie palyginimai kaip čia.

P4. Tarkime, kad viskas vyksta kaip pavyzdyje P3: du stovyklautojai miške numeta po smilkstančią cigaretę ($X = Y = 1$), kyla gaisras ($Z = 1$) ir gaisrui būtų pakakę vienos cigaretės. Šį aprašymą papildykime nauja aplinkybe: dėl miško gaisro labai nukenčia pamiškės

2 lentelė. Ar pavyzdžiuose P2 ir P3 įvykis X = 1 buvo įvykio Z = 1 priežastis? T1–T9 atsakymai (T – taip, N – ne)

	T1	T2	T3	T4	T5	T6	T7	T8	T9
P2	N	T	T	T	T	T	T	N	T
P3	N	N	N	N	T	T	T	T	T

gyventojo N medinė pirtelė (įvykis $W = 1$). Ar tai, kad pirmasis stovyklautojas numetė cigaretę, buvo priežastis pirtelei nukentėti (žr. 4 pav.)?

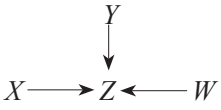
Lygčių sistema:

$$X = 1$$

$$Y = 1$$

$$Z = X \vee Y$$

$$W = Z$$



4 pav.

Šis pavyzdys iliustruoja T5 ir T6 skirtumą. Norint pamatyti, kad $X = 1$ buvo $W = 1$ priežastis, reikėtų priskirti Y fiksuotą reikšmę 0, nes KST „jei $X = 0$ ir $Y = 0$, tai $W = 0$“ teisingas. Y nepriklauso keliui $\langle X, Z, W\rangle$, ir KST „jei $Y = 0$, tai $Z = 1$ ir $W = 1$“ teisingas, todėl T6 leidžia priskirti Y fiksuotą reikšmę 0. Tačiau T5 to neleidžia, nes Y nėra viršūnės W tėvinė viršūnė. Todėl T6 požiūriu $X = 1$ buvo $W = 1$ priežastis, o T5 požiūriu ne.

Atrodo arbitralu teigti (kaip kad tektų daryti sutikus su T5), kad pavyzdyje P3 įvykis $X = 1$ buvo $Z = 1$ priežastis, bet P4 įvykis $X = 1$ nebuvo $W = 1$ priežastis.

T1–T4 nagrinėdamos P4 prieina tuos pačius atsakymus kaip ir T5, bet ne dėl tų pačių aplinkybių kaip T5, o dėl tų, kurios buvo aprašytos nagrinėjant P3. T1–T4 požiūriu tarp P3 ir P4 esminio skirtumo nėra.

T7, T8 ir T9 pavyzdį P4 nagrinėja taip pat arba panašiai kaip T6.

P5. Tarkime, kad viskas vyksta kaip pavyzdyje P2: mokinys žudikas iššauna ($X = 1$) ir auka žūva ($Z = 1$). Jei mokinys nebūtų šovęs, vietoje jo būtų šovęs mokytojas (būtų teisinga $Y = 1$) ir auka vis tiek būtų žuvusi. Papildykime šį pasakojimą nauju įvykiu: prašmatniomis aukos laidotuvėmis, įvykusiomis kitą dieną (kintamasis W). Ar mokinio šūvis buvo laidotuvių ($W = 1$) priežastis (žr. 5 pav.)?

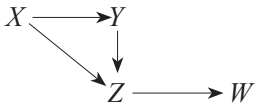
Lygčių sistema:

$$X = 1$$

$$Y = \sim X$$

$$Z = X \vee Y$$

$$W = Z$$



5 pav.

Šis pavyzdys rodo T2 ir T4 skirtumą. T2 teigia, kad $X = 1$ buvo $W = 1$ priežastis, o T4 teigia, kad nebuvo. Mat kintamasis Y, kuriam čia reikėtų priskirti fiksuotą reikšmę 0, pavyzdžio grafe nėra tėvinė W viršūnė ir dėl to T4 apsiriboja KST „jei $X = 0$, tai $W = 0$“ tikrinimu. Panašiai kaip ir T5 pavyzdyje P4, teorija T4 čia atrodo ganėtinai arbitraliai. Jeigu mokinio šūvis buvo aukos žūties priežastis, tai jis buvo ir laidotuvių priežastis, o T4 tai neigia.

T2 šį pavyzdį nagrinėja iš esmės taip pat kaip ir P2.

Šį pavyzdį T5 nagrinėja lygiai kaip ir T4, o T3, T6, T7 ir T9 – panašiai kaip ir T2. T1 ir T8 pavyzdį P5 irgi nagrinėja iš esmės tokiu

pat būdu kaip ir P2, dėl to abi teigia, kad $X = 1$ nebuvo $W = 1$ priežastis.

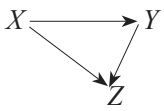
P6. Viskas vyksta kaip pavyzdyje P3: du stovyklautojai miške numeta po smilkstančią cigaretę ($X = Y = 1$) ir kyla gaisras ($Z = 1$). Papildykime tai tokia informacija: abi cigaretės numetamos tyčia ir antrasis stovyklautojas ją numeta sekdamas pirmojo pavyzdžiu. Jis cigaretę numeta *dėl to*, kad ją numetė pirmasis stovyklautojas (žr. 6 pav.). Ar antrojo stovyklautojo veiksmas buvo gaisro priežastis?

Lygčių sistema:

$$X = 1$$

$$Y = X$$

$$Z = X \vee Y$$



6 pav.

Būtų arbitralu sakyti, kad P3 įvykis $Y = 1$ buvo $Z = 1$ priežastis, o P6 ne. P6 iliustruoja T5 ir T6 skirtumą, bet kitokį skirtumą negu P4. Norint gauti intuityviai teisingą atsakymą, čia reikėtų X priskirti fiksuotą reikšmę 0, kuri yra kontrafaktinė ($Y = 1$ ir $Z = 1$ priežastinį ryšį parodo KST „jei $Y = 0$ ir $X = 0$, tai $Z = 0$“). Tačiau KST „jei $X = 0$, tai $Y = 1$“ ir „jei $X = 0$, tai $Z = 1$“ klaidingi. Mėginimas pakeisti X reikšmę pakeičia Y ir Z reikšmes, o T6 derinio *f* apibrėžime (žr. 2 skyrelį, formą KT) reikalauja, kad fiksuotos reikšmės priskyrimas kintamajam X nepakeistų nė vieno keliui $<Y, Z>$ priklausančio kintamojo, išskyrus Y, reikšmės – t. y. kad nepakeistų Z reikšmės. Dėl to T6 tikrindama $Y = 1$ ir $Z = 1$ priežastinį ryšį atsižvelgia tik į (klaidingus) KST „jei $Y = 0$, tai $Z = 0$“ ir „jei $Y = 0$ ir $X = 1$, tai $Z = 0$“: pamėgina ir apskritai nepriskirti X fiksuotos reikšmės, ir priskirti X fiksuotą reikšmę 1. T5 derinio *f* apibrėžime leidžia priskirti X tik tokią fiksuotą reikšmę, kad priskyrus šią reikšmę X ir tada priskyrus Y fiksuotą reikšmę 1 (faktinę reikšmę), Z reikšmė būtų faktinė. Kadangi KST „jei $X = 0$ ir $Y = 1$, tai $Z = 1$“ teisingas, tai reikšmės 0 priskyrimas X šį reikalavimą atitinka. Dėl to T5, kitaip nei T6, nagrinėdama P6 tikrina KST „jei $Y = 0$ ir $X = 0$, tai $Z = 0$“ ir gauna atsakymą, kad $Y = 1$ buvo $Z = 1$ priežastis (taip pat ji dar išmėgina (klaidingą) KST „jei $Y = 0$ ir $X = 1$, tai $Z = 0$“).

T7 šį pavyzdį nagrinėja lygiai kaip ir T5.

Nagrinėjant šį pavyzdį, T5, T6 ir T7 įdomu palyginti su T8 arba T9. T8 ir T9 nurodo priskirti X tokią fiksuotą reikšmę, kad ją priskyrus liktų nepakeistas *nė vienas* keliui $<Y, Z>$ priklausantis kintamasis. Šis reikalavimas yra dar griežtesnis nei atitinkamas T6 reikalavimas, ir jis duoda tokį pat rezultatą – tikrinant $Y = 1$ ir $Z = 1$ priežastinį ryšį atsižvelgiama tik į (klaidingą) KST „jei $Y = 0$ ir $X = 1$, tai $Z = 0$“, ir padaroma išvada, kad $Y = 1$ nebuvo $Z = 1$ priežastis.

T1–T4 nagrinėja P6 panašiai kaip P3, dėl to neigia, kad $Y = 1$ buvo $Z = 1$ priežastis.

Nagrinėjant P4, P5 ir P6 teorijos T1–T9 gali būti suskirstytos į šešias grupes (žr. 3 lentelę).

T5 ir T7 nagrinėjant P6 atrodo pranašesnės už visas kitas teorijas. Visgi tai, kad T5 čia duoda intuityviai teisingą atsakymą, yra atsitiktinumas: ji veikia tik dėl to, kad X

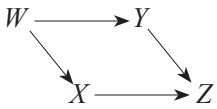
3 lentelė. Ar pavyzdžiuose P4, P5 įvykis X = 1 buvo W = 1 priežastis ir ar P6 įvykis Y = 1 buvo Z = 1 priežastis? T1–T9 atsakymai

	T1	T2	T3	T4	T5	T6	T7	T8	T9
P4	N	N	N	N	N	T	T	T	T
P5	N	T	T	N	N	T	T	N	T
P6	N	N	N	N	T	N	T	N	N

čia yra *Z* tėvinė viršūnė, o tai nėra esminė pavyzdžio ypatybė (panašiai kaip pavyzdyje P5 viršūnė *Y* nėra *W* tėvinė viršūnė, ir tai nepadaro P5 pernelyg skirtingo nuo P2).

P7. Du stovyklavimo mėgėjai (ankstesnių pavyzdžių herojai) vėl ruošiasi į mišką. Mokytojas jiems sako: „numeskite miške po smilkstančią cigaretę“. Stovyklavimo mėgėjai paklūsta, numeta po cigaretę, miške kyla gaisras. Ar mokytojo paliepimas (įvykis $W = 1$) buvo gaisro (įvykis $Z = 1$) priežastis (žr. 7 pav.)?
Lygčių sistema:

$$W = 1$$
$$Y = W$$
$$X = W$$
$$Z = X \vee Y$$



7 pav.

Šis pavyzdys rodo, kad T2 nėra ekvivalentiška T3 ir kad (kitaip nei atrodytų atsižvelgiant tik į P1-P6) tarp T1 ir T2 ekstensijų nėra subordinacijos santykio. Vienos teorijos, tarp kurios ekstensijos ir bet kurios iš likusių teorijų ekstensijų būtų subordinacijos santykis, apskritai nėra. Kadangi teisingas KST „jei $W = 0$, tai $Z = 0$“, tai T1 teigia, kad $W = 1$ buvo $Z = 1$ priežastis (priežastingumas ir priežastinė priklausomybė čia sutampa). Nagrinėjant kelią $\langle W, X, Z\rangle$ teorija T2 nurodo priskirti *Y* fiksuotą reikšmę 1 (faktinę reikšmę), nes *Y* yra alternatyvaus *W* ir *Z* jungiančio kelio – $\langle W, Y, Z\rangle$ – tarpinis kintamasis. Todėl nagrinėdama $\langle W, X, Z\rangle$ T2 tikrina (klaidingą) KST „jei $W = 0$ ir $Y = 1$, tai $Z = 0$“. Dėl analogiškų aplinkybių nagrinėjant $\langle W, Y, Z\rangle$ teorija T2 tikrina (irgi klaidingą) KST „jei $W = 0$ ir $X = 1$, tai $Z = 0$“. Taigi T2 teigia, kad $W = 1$ nebuvo $Z = 1$ priežastis, ir tai (kartu su P2 ir P5) rodo, kad tarp T1 ir T2 ekstensijų yra sankirtos santykis[4].

T3 nagrinėjant $\langle W, X, Z\rangle$ tikrina du KST: „jei $W = 0$ ir $Y = 1$, tai $Z = 0$“ ir „jei $W = 0$, tai $Z = 0$“. Šiuo atveju fiksuotos reikšmės priskyrimas $\langle W, X, Z\rangle$ nepriklausantiems kintamiesiems norint gauti intuityviai teisingą atsakymą apskritai nebūtinas: reikėtų arba jokiam kintamajam nepriskirti fiksuotos reikšmės, arba priskirti *Y* fiksuotą reikšmę 0 (kontrafaktinę reikšmę). T3

[4] Kadangi tuo atveju, jei pradinės sąlygos pavyzdyje P2 būtų $X = 0$, tada T1 apie šį pavyzdį teigtų, kad $X = 0$ buvo $Z = 1$ priežastis, o T2–T9 tai neigtų, tai tarp T1 ekstensijos ir kiekvienos iš teorijų T2–T9 ekstensijų yra sankirtos santykis.

4 lentelė. Ar pavyzdyje P7 įvykiai W = 1 ir X = 1 yra Z = 1 priežastys?

	T1	T2	T3	T4	T5	T6	T7	T8	T9
P7 ($W = 1$)	T	N	T	N	T	T	T	T	T
P7 ($X = 1$)	N	N	N	N	T	T	T	N	N

leidžia nepriskirti nė vienam kintamajam fiksuotos reikšmės, ir todėl nagrinėjant P7 ji gauna intuityviai teisingą atsakymą, kad $W = 1$ buvo $Z = 1$ priežastis.

T4 šį pavyzdį nagrinėja kaip ir T2. T5, T6 ir T7 nagrinėdamos <*W*, *X*, *Z*> tikrina KST „jei $W = 0$ ir $Y = 1$, tai $Z = 0$“ ir „jei $W = 0$ ir $Y = 0$, tai $Z = 0$“, nes nei fiksuotos reikšmės 1, nei fiksuotos reikšmės 0 priskyrimas kintamajam *Y* nepakeičia *Z* reikšmės (be to, T6 ir T7 dar tikrina „jei $W = 0$, tai $Z = 0$“). Antrojo iš šių KST pakanka $W = 1$ ir $Z = 1$ priežastiniam ryšiui atskleisti. T8 šį pavyzdį nagrinėja lygiai kaip ir T1. Panašiai daro ir T9.

Pavyzdyje P7 nagrinėjant $X = 1$ ir $Z = 1$ santykį išryškėja T6 ir T9 skirtumai. T6 jį nagrinėjant tikrina KST „jei $X = 0$ ir $Y = 0$, tai $Z = 0$“, kurio čia pakanka priežastiniam ryšiui atskleisti (ir kitus KST). T9 čia yra priversta priskirti egzogeniškam kintamajam *W* fiksuotą reikšmę 1 (reikšmė 0 pakeistų ir *X*, ir *Z*). Priskirti fiksuotą reikšmę *Y* ši teorija negali, nes *Y* nėra viršūnės *X* palikuonė. Taigi T9 čia tikrina tik (klaidingą) KST „jei $X = 0$, tai $Z = 0$“ ir gauna intuityviai klaidingą atsakymą, kad $X = 1$ nebuvo $Z = 1$ priežastis. T1 ir T8 šį atvejį nagrinėja lygiai kaip ir T9, T5 ir T7 – panašiai kaip ir T6. T2–T4 jį nagrinėja labai panašiai kaip P3, dėl to visos irgi teigia, kad $X = 1$ nebuvo $Z = 1$ priežastis (iš tiesų, $X = 1$ ir $Z = 1$ santykis pavyzdžiuose P3 ir P7 intuityviai panašus).

Kodėl T2–T4 (ir T1) šį pavyzdį ir P3 nagrinėja panašiai, o T8–T9 – skirtingai? Dėl to, kad pavyzdyje P3 kintamasis *Y* buvo egzogeniškas, o šiuo atveju ne. Tai rodo, kad nagrinėjant P3 teorijos T8 ir T9 teikė intuityviai teisingą atsakymą atsitiktinai (atsižvelgdamos į neesmines pavyzdžio ypatybes).

Apžvelgti pavyzdžiai rodo, kad T1–T9 ekstensijos skirtingos (žr. 2–4 lenteles).

5. Ko trūksta teorijoms T1–T8, kad jos tiktų pavyzdžiams P2– P7?

Kaip reikėtų nagrinėti trečio ir ketvirto skyrelio pavyzdžius, kad gautume tuos atsakymus, kurie čia buvo pavadinti intuityviai teisingais? Čia bus pasiūlyta taisyklė, kuri leistų tai padaryti. Šis pasiūlymas skirtas tik P2–P7 ir į juos labai panašiems pavyzdžiams (P1 reikalautų kitokios analizės, be to, visos teorijos jį nagrinėja vienodai gerai, dėl to į jį nebus atsižvelgta). Gali būti, jog nagrinėjant kitus, čia neaprašytus pavyzdžius jis būtų nepritaikomas arba netgi duotų nepageidaujamą rezultatą.

P2–P6 ir P7 (pastarajame pavyzdyje – nagrinėjant $X = 1$ ir $Z = 1$ santykį) priežasties kintamojo kontrafaktinę priklausomybę nuo padarinio kintamojo „užgožia“ kokia nors kita pakankama priežastis: nors yra teisinga, kad koks nors įvykis $X = 1$ yra kito įvykio $Y = 1$ priežastis, KST „jei $X = 0$, tai $Y = 0$“ yra klaidingas, nes tuo atveju, kai įvyksta $X = 0$, padarinį $Y = 1$ sukelia kažko-

kia kita priežastis $Z = 1$, arba kelios tokios priežastys (kaip rodo P4, Z gali nebūti Y tėvinė viršūnė; kaip rodo P2, Z gali būti X palikuonė; kaip rodo P3, Z gali ja ir nebūti; kaip rodo P2, Z gali nebūti egzogeniškas kintamasis). Vadinasi, norint pamatyti, kad Y kontrafaktiškai priklauso nuo X, reikia priskirti Z fiksuotą reikšmę 0 (kaip rodo P3, ši reikšmė gali būti kontrafaktinė, kaip rodo P2, ji gali būti ir faktinė). Jei Z bus priskirta fiksuota reikšmė 1, šios priklausomybės nematysime. Kartais gali būti taip, kad jei $Z = 1$ įvyktų, tai $Z = 1$ būtų $Y = 1$ priežastimi, tačiau yra teisingas KST „jei $X = 0$, tai $Z = 0$“; tuomet irgi galima, tačiau nebūtina priskirti Z fiksuotos reikšmės 0 (žr. $W = 1$ ir $Z = 1$ santykio pavyzdyje P7 aptarimą).

Tikslesnė ir šiek tiek bendresnė šių idėjų formuluotė tokia:

> *Rekomenduojamas fiksuotų reikšmių priskyrimas P2–P7 tipo pavyzdžiuose*. Jeigu (1) nagrinėjamo pavyzdžio grafe viršūnė Y yra viršūnės X palikuonė, (2) faktinė X ir Y reikšmė yra 1 ir (3) kiekvienam X ir Y jungiančiam keliui K teisinga, kad kiekvieno kintamojo Z, kuris priklauso $K\backslash\{X\}$, lygties L_Z forma yra $Z = \sim W$ arba $Z = W_1 \vee \ldots \vee W_n$, tai $X = 1$ yra $Y = 1$ priežastis, jei ir tik jei egzistuoja toks X ir Y jungiantis kelias K, kad yra teisingas KST „jei $X = 0$ ir $F = f$, tai $Y = 0$“, kur F yra aibė visų kintamųjų, kurie (a) nepriklauso $K\backslash\{X\}$, ir (b) yra tų aibės $K\backslash\{X\}$ kintamųjų, kurių lygties forma yra $Z = W_1 \vee \ldots \vee W_n$, tėvinės viršūnės, o f yra reikšmių derinys, kuriame visiems F kintamiesiems priskiriama fiksuota reikšmė 0.

T7 yra vienintelė iš aptartų teorijų, kuri nagrinėjant P2–P7 atitinka šią rekomendaciją.

Išvados

1. Formali išvada: kaip rodo išnagrinėti pavyzdžiai P1–P7, teorijų T1–T9 ekstensijos skirtingos. Nėra tokios vienos teorijos, tarp kurios ekstensijos ir bet kurios iš likusių teorijų ekstensijų būtų subordinacijos santykis.
2. Teorijų įvertinimas: jeigu pavyzdžiai P4, P6 ir P7 (pastarasis – $X = 1$ ir $Z = 1$ santykio požiūriu) laikomi intuityviai panašiais į P3, o pavyzdys P5 – intuityviai panašiu į P2, tada teorijos T2–T6, T8 ir T9 neatitinka reikalavimo, kad intuityviai panašūs pavyzdžiai nebūtų nagrinėjami skirtingai, t. y. reikalavimo, kad ten, kur jie sutampa, teorija nurodytų esant tuos pačius priežastinius ryšius. T1 ir T7 nagrinėjant šiuos pavyzdžius šį reikalavimą atitinka.
3. Tam, kad nagrinėjant P2–P7 (ir į juos gana panašius pavyzdžius) būtų gaunami tie atsakymai, kurie tekste buvo pavadinti intuityviai teisingais, reikia vadovautis taisykle „Rekomenduojamas fiksuotų reikšmių priskyrimas P2–P7 tipo pavyzdžiuose“, kuri pateikta penktame skyrelyje[5].

[5] Autorius dėkoja Viliui Dranseikai ir Ievai Vasilionytei už vertingus patarimus.

LITERATŪRA

Glymour, C. & Wimberly, F. 2007. „Actual Causes and Thought Experiments“, in J.K. Campbell, M. O'Rourke, H. Silverstein (ed.). *Causation and Explanation*. Cambridge, Mass.: MIT Press, 43–67.

Grenda, V. 2006. „Kas yra nehiumistinė priežastingumo teorija?“, *Problemos* 69: 27–38.

Hall, N. 2007. „Structural Equations and Causation“, *Philosophical Studies* 132 (1): 109–136.

Halpern, J.Y. & Pearl, J. 2005. „Causes and Explanations: A Structural-Model Approach. Part I: Causes“, *The British Journal for the Philosophy of Science* 56 (4): 843–887.

Hitchcock, C. 2001. „The Intransitivity of Causation Revealed in Equations and Graphs“, *The Journal of Philosophy* 98 (6): 273–299.

Lewis, D. 1986 (1973). „Causation“, in *Philosophical Papers,* Volume II. Oxford: Oxford University Press, 159–172.

Menzies, P. 2004. „Causal Models, Token Causation, and Processes“, *Philosophy of Science* 71 (5): 820–832.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

COMPARISON OF NINE COUNTERFACTUAL THEORIES OF CAUSATION BY THE USE OF SEVEN EXAMPLES

Vytautas Grenda

S u m m a r y

The article compares and evaluates nine last-decade counterfactual theories of singular causation, which were either created or inspired by David Lewis, Joseph Y. Halpern and Judea Pearl and presented in the texts of six other authors. Some differences between those theories that have not yet been described in literature are shown in the article. It is argued that the majority of those theories analyze intuitively similar examples in different ways. In that respect, those theories impose distinctions which, according to the folk theory of causation, are (and should be) absent.

Keywords: singular causation, counterfactuals, backup causes, overdetermination.

Įteikta 2009 06 30