



Università degli Studi di Padova

DIPARTIMENTO DI FILOSOFIA, SOCIOLOGIA, PEDAGOGIA E PSICOLOGIA APPLICATA
Corso di Laurea Magistrale in Scienze Filosofiche

TESI DI LAUREA MAGISTRALE

Modelli Causali e Causa Effettiva

Relatore:
Ch.mo Prof. Vittorio Morato

Candidato:
Alvaro Gaiotti
Matricola:
2029252

*Felice chi ha potuto conoscere la causa delle cose, e si mise sotto i piedi ogni
paura e il destino inesorabile e il fragore dell'Acheronte avido.*

Virgilio – Georgiche II, vv. 490-493

*There is only one constant, one universal. It is the only real truth. Causality.
Action, reaction. Cause and effect.*

Merovingio – Matrix Reloaded

 INDICE

Indice	v
Acronimi	vi
Ringraziamenti	vii
Introduzione	ix
 Preliminari	
1 Modelli causali	3
1.1 La cornice metodologica	3
1.2 Interventi	12
1.3 Una semantica dei controfattuali	22
2 La causa effettiva	33
2.1 La proposta di Lewis e le sue mancanze	33
2.2 La causa effettiva secondo Halpern e Pearl	36
2.3 Conclusione	45
 Discussione	
3 Il problema dell'isomorfismo e l'arte di modellare	49
3.1 Il problema	50
3.2 Le soluzioni	53
3.3 Modelli adatti e l'arte di modellarli	60
3.4 Sul Problema dell'Isomorfismo e sul Problema del Disaccordo	64
4 Una natura, diverse funzioni: sul pluralismo causale	79
4.1 L'approccio funzionale	80
4.2 Alcuni concetti causali	83
4.3 Una conclusione	89
5 Una nuova definizione di causa effettiva	91
5.1 Una nuova causa effettiva	91
5.2 Alcuni esempi	93
Conclusione	97
 Bibliografia	 101

ACRONIMI

CIP	Causal Interpretation Problem
CMC	Causal Markov Condition
HP	Halpern-Pearl
INUS	Insufficient, but Necessary part of an Unnecessary but Sufficient condition
NESS	Necessary Element of a Sufficient Set
PI	Problema dell'Isomorfismo
PD	Problema del Disaccordo
SEM	Structural Equations Model
SCM	Structural Causal Model
VVZ	virus varicella-zoster

RINGRAZIAMENTI

Non è semplice fare un conciso elenco di tutti coloro che hanno, in qualche maniera, chi più direttamente, chi più indirettamente, partecipato al raggiungimento del traguardo che mi accingo a valicare. Condensare in poche righe un percorso che è iniziato più di dieci anni fa non sembra possibile. Dedicherò una menzione diretta a coloro che, nel presente momento in cui scrivo, sento più vicini.

Innanzitutto, sono profondamente grato per il supporto dei miei genitori in questi anni e per il loro incoraggiamento a seguire la mia passione e intraprendere questo percorso di studi. Ringrazio anche mia nonna, Luciana, per la passione che mi ha trasmesso per le discipline umanistiche: senza di lei il mio interesse per la filosofia non sarebbe mai potuto nascere.

Un ringraziamento speciale va a mio fratello, sempre capace di strapparmi un sorriso, che ha stimolato alcune delle curiosità, in particolare matematiche e modellistiche, che hanno portato a questa tesi.

Vorrei ringraziare, inoltre, Alessandro, Carlo e Matteo, che mi hanno sostenuto durante la difficile stesura di questa tesi e sono sempre stati pronti a spronarmi nei momenti di maggior abbandono e sconforto.

In ogni caso, la vita è molto più che studio, lavoro e “sudate carte”. Grazie, quindi, anche ad Alex, Francesco, Giorgia, Giovanni, Giulio, Orso, Sara, Silvia e Valentina per la preziosa compagnia durante gli anni universitari e per avervi aggiunto un pizzico di goliardia.

Infine, per il Professor Vittorio Morato: non avrei mai potuto immaginare che il seguire un corso di Filosofia del Linguaggio durante il secondo anno di studi universitari avrebbe portato a tutto questo. Oltre ad aver inaugurato il mio interesse per la causalità e, quindi, essere in parte il fautore di questa tesi, mi ha accompagnato nel mio percorso universitario sin dalla stesura della tesi triennale. Non avrei potuto trovare relatore più comprensivo, gentile e disponibile. Un sentito ringraziamento.

INTRODUZIONE

We live in exciting times. By ‘we’ I mean philosophers studying the nature of causation. The past decade or so has witnessed a flurry of philosophical activity aimed at cracking this nut, and, surprisingly, real progress has been made.

— Christopher Hitchcock, «The Intransitivity of Causation Revealed in Equations and Graphs»

Comprendere il mondo è, forse, il desiderio umano che più ci caratterizza come tali. Sin da quando l’umanità è stata in grado di produrre, nella propria mente, qualcosa di simile ad un concetto, si è sempre posta una domanda fondamentale: perché le cose stanno così? Qual’è la causa di tutto ciò? La filosofia stessa nasce dalla meraviglia, ed essa “specula attorno [...] alle cause”¹.

L’embrione di questa tesi comincia a formarsi durante il corso di Filosofia della Scienza tenuto dai proff. Giaretta, Morato e Spolaore. Il fulcro del corso ruotava attorno al notevole *Philosophy of Epidemiology* di Broadbent². Uno dei fondamentali problemi che Broadbent affrontava, in questo testo, era il cosiddetto *Causal Interpretation Problem (CIP)*: questo consiste nel come interpretare la *carica causale* di una misura di associazione. Venivano affrontate varie soluzioni e, una di queste, riponeva nei modelli causali la speranza di una soluzione.

Avevo sempre ritenuto, ingenuamente, che la fisica fosse tutto quanto fosse necessario per spiegare la causalità: trasferimenti di energia, di lavoro, quantità che si conservano. Dopotutto, è questo il nostro primo approccio teorico alla causalità: le equazioni della fisica che impariamo già in giovane età. Ora mi rendevo conto che quanto avevo sempre dato per scontato era in realtà estremamente complesso e articolato, una matassa terribile di cui non si trovava il bandolo. Spesso è questo il primo passo che compie la filosofia: gettare uno sguardo su ciò che è talmente banale da passare inosservato. E scoprire che banale non è.

Le mie domande, a questo punto, si erano d’improvviso moltiplicate ed ero curioso di comprendere come il formalismo grafico e matematico dei modelli causali potesse anche solo immaginare di vantarsi di poterne risolvere una grande quantità. Quindi, mi addentrarai nei meandri di *Causality: Models, Reasoning, and Inference* di Pearl.

Questa tesi testimonia un percorso di ricerca iniziato con la lettura di questo testo. Lungo questo percorso, ogni risultato ha aperto a nuove domande, ogni nuova definizione nascondeva in sé il germe di un nuovo controesempio che la inficiava. La letteratura riguardante i modelli causali è ampia; il dibattito, in quanto recente, infuocato. Dopo aver esplorato questo campo in lungo ed in largo, catturato dalla meravigliosa potenza e formalità di questo apparato, ho provato a fornire anche un piccolissimo contributo originale. Questa tesi è una testimonianza anche di quest’ultimo mio sforzo. Procediamo con una breve presentazione dei capitoli e delle problematiche che verranno discusse all’interno di ciascuno.

Il **Capitolo 1** presenta la cornice metodologica e l’apparato teorico e formale dei modelli causali. I modelli causali sono, fondamentalmente, un’applicazione dei *Structural Equations Models (SEMs)* alle relazioni di causa ed effetto. In questo capitolo presento uno dei formalismi possibili con cui caratterizzarli, utilizzato in particolare da Halpern³, in quanto più comprensibile ad un occhio

¹ Aristotele, *Metafisica*, trad. da G. Reale, Bompiani, 2000:982b-983a.

² A. Broadbent, *Philosophy of Epidemiology*, Palgrave Macmillan UK, London 2013, DOI: [10.1057/9781137315601](https://doi.org/10.1057/9781137315601).

³ J. Y. Halpern e J. Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations», *The British Journal for the Philosophy of Science*, vol. 56, n. 4 (2005), pp. 889-911; J. Y. Halpern, «Defaults and Normality in Causal Structures», in *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, KR’08, AAAI Press, Sydney, Australia 2008, pp. 198-208; J. Y. Halpern e C. Hitchcock, «Graded causation and defaults», *The British Journal for the Philosophy of Science*, vol. 66, n. 2 (2015), pp. 413-457; J. Y. Halpern, «A Modification

filosofico. Dopo questo lavoro fondazionale, analizzo la problematica nozione di intervento, un nodo fondamentale per le teorie basate sui modelli causali, in quanto è su questa nozione⁴ che si decide della circolarità (presente, assente; buona o viziosa) della teoria. Procedo, successivamente, con l'analisi di un'interessante proposta di Hiddleston⁵: una semantica dei controfattuali basata sui modelli causali. Indirizzo, infine, qualche critica a questa proposta e propongo degli emendamenti originali in grado di migliorarla; seppur non sufficienti a riscattarla completamente.

Completata l'opera fondazionale, nel **Capitolo 2** presento l'evoluzione delle definizioni di causa effettiva della famiglia *Halpern-Pearl* (**HP**). La presentazione verrà corredata da controesempi via via più complessi, al fine di evidenziare la necessità di una raffinazione costante della proposta **HP** al fine di concordare con le nostre intuizioni.

Una raffinazione ancora oggi in corso, attraverso l'introduzione di concetti legati alla distinzione default/deviante e a misure di plausibilità⁶, utili non solo per affrontare problemi legati a strutture causali complesse, bensì a rendere conto dei differenti giudizi di causa effettiva che, talvolta, emettiamo di fronte alle medesime strutture causali e controfattuali. Il **Capitolo 3** si concentrerà proprio su questa problematica: il cosiddetto *Problema dell'Isomorfismo* (**PI**). Questo problema è un *puzzle* tanto interessante quanto problematico: i modelli causali assumono, fondamentalmente, che l'unica caratteristica rilevante per la formulazione di giudizi di causa effettiva sia la struttura intrinseca della situazione che andiamo a modellare; ma se, a parità di struttura, i giudizi causali sono differenti, questa assunzione sembra cadere, e con essa tutto l'edificio. Questo rappresenta una problematica estremamente rilevante anche per gli studi psicologici che, negli ultimi anni, provano a sostenere che la modalità con cui pensiamo ai controfattuali è molto simile al modo in cui i modelli causali rispondono a *queries* controfattuali⁷. Presenterò, nel corso del capitolo, alcune plausibili soluzioni ed, infine, nonostante alcune suggerimenti di avviso opposto da parte della letteratura psicologica⁸, sosterrò l'argomento di Blanchard e Schaffer⁹ a favore di una raffinazione della cosiddetta "arte di modellare": il **PI**, insomma, è una falso problema, in quanto sorge solo in quei casi in cui la modellazione non è stata adeguata. Non sono, quindi, le reali strutture causali della situazione ad essere isomorfe, bensì le loro rappresentazioni: l'isomorfismo è frutto di modelli che rappresentano scorrettamente una struttura causale di una situazione, rendendola identica a quella posseduta da un'altra situazione, senza che questa sia veramente identica. Proporrò, inoltre, una procedura, ispirata da Fischer¹⁰, per mettere a frutto la concettualità sviluppata per risolvere il **PI**.

Nel **Capitolo 4** presenterò uno dei miei contributi originali, ovvero la proposta di alcuni concetti causali, argomentando a favore di una posizione funzionalista e, di conseguenza, pluralista, rispetto ai concetti causali. La mia posizione, ispirata da Woodward e Danks, è che i concetti causali, ed in particolare quelli legati alla causa effettiva, svolgano una particolare funzione cognitiva; e che da questa funzione cognitiva sia scorretto prescindere in sede di analisi. Di conseguenza propongo una concettualità basata sul modo in cui ragioniamo di fronte a differenti situazioni causali, in

of the Halpern-Pearl Definition of Causality», in *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, AAAI Press, Buenos Aires, Argentina 2015, pp. 3022-3033; J. Y. Halpern, *Actual Causality*, MIT Press, 2016.

4 E, parzialmente, qualche altra.

5 E. Hiddleston, «A Causal Theory of Counterfactuals», *Noûs*, vol. 39, n. 4 (2005), pp. 632-657, DOI: [10.1111/j.0029-4624.2005.00542.x](https://doi.org/10.1111/j.0029-4624.2005.00542.x).

6 N. Friedman e J. Y. Halpern, «Plausibility Measures: A User's Guide», in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI'95*, Morgan Kaufmann Publishers Inc., Montréal, Qué, Canada 1995, pp. 175-184.

7 Si veda, solo per citarne alcuni: A. Gopnik *et al.*, «A Theory of Causal Learning in Children: Causal Maps and Bayes Nets», *Psychological Review*, vol. 111, n. 1 (2004), pp. 3-32, DOI: [10.1037/0033-295X.111.1.3](https://doi.org/10.1037/0033-295X.111.1.3); J. Knobe e B. Fraser, «Causal Judgement and Moral Judgement: Two Experiments», in, *Moral Psychology, vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, a cura di W. Sinnott-Armstrong, MIT Press, 2008, pp. 441-448; T. Gerstenberg *et al.*, «Back on track: Backtracking in counterfactual reasoning», in *Proceedings of the Annual Meeting of the Cognitive Science Society*, n. 35, 2013, vol. 35, DOI: [10.13140/2.1.1473.1209](https://doi.org/10.13140/2.1.1473.1209); E. Popa, «Getting counterfactuals right: the perspective of the causal reasoner», *Synthese*, vol. 200, n. 1 (2022), DOI: [10.1007/s11229-022-03492-6](https://doi.org/10.1007/s11229-022-03492-6).

8 C. Hitchcock e J. Knobe, «Cause and Norm», *The Journal of Philosophy*, vol. 106, n. 11 (2009), pp. 587-612.

9 T. Blanchard e J. Schaffer, «Cause without Default», in H. Beebe *et al.* (a cura di), *Making a Difference*, Oxford University Press, 2017, DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001), pp. 175-214, DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001).

10 E. Fischer, «Causation and the Problem of Disagreement», *Philos. of Sci.*, vol. 88, n. 5 (2021), pp. 773-783, DOI: [10.1086/714852](https://doi.org/10.1086/714852).

particolare: casi in cui un'azione assicura (o evita sicuramente) l'esito desiderato, afferente dominio di quelli che io definisco blocchi causali, e casi in cui, invece, sono necessarie una molteplicità di azioni al fine di evitare un esito, afferente a quella che io definisco causalità per scambio.

Infine, nel **Capitolo 5**, presenterò una definizione di causa effettiva che ho formulato grazie all'indagine svolta nel precedente capitolo. In particolare, la definizione è stata ottenuta testando il concetto di blocco causale condizionale (**Definizione 4.2.4**) su di alcuni esempi, finendo per scoprire il naturale sfociare del concetto in una definizione di causa effettiva. Questa, a mio parere, è una buona indicazione della fecondità dei concetti prodotti nel capitolo precedente.

Chiuderà il tutto una Conclusione, in cui verrà fatto un bilancio dello scritto e verranno riproposti i principali risultati ottenuti.

PRELIMINARI

MODELLI CAUSALI

Hidden in our intuition there is a notion of causation because we cannot grasp any other logic except causation. When we think about correlation we think causally: things cannot be correlated unless there is a reason for them to vary together.

— Judea Pearl

Una delle più fruttuose promesse per quanto riguarda gli sviluppi dell'Intelligenza Artificiale, dell'inferenza causale e, in generale, di formali trattazioni della causalità (e non solo¹), sono i **SEMs**. Questi, nella loro interpretazione causale, rappresentano, a mio parere, uno dei campi di ricerca più interessanti dell'ultimo ventennio, specie per la trasversalità e la multidisciplinarietà che caratterizza quest'area: si occupano, infatti, di modelli causali economisti, filosofi, scienziati sociali, epidemiologi, informatici e altri accademici con le formazioni più disparate.

Questo capitolo si occuperà di presentare questo strumento estremamente flessibile e potente, su cui si concentra ancora un fervente dibattito. Innanzitutto verrà introdotto il formalismo necessario ad una trattazione agevole dei modelli causali, che, in questa sede, coinciderà con quello sviluppato da **Halpern e Pearl**² e, successivamente adottato anche da **Halpern** nel suo *Actual Causality*³, in quanto ritenuto dotato di maggiore esplicatività e chiarezza⁴, successivamente verrà brevemente analizzata la nozione di intervento, una nozione fondamentale per la tenuta filosofica di una proposta basata sui modelli causali. Infine verrà presentata una proposta di teoria semantica dei controfattuali basata sui modelli causali, proposta da **Hiddleston**. Questa proposta verrà discussa e se ne metteranno in luce alcune problematiche. Tutto questo andrà a costituire la fondazione per affrontare il **Capitolo 2** ed i successivi capitoli.

1.1 LA CORNICE METODOLOGICA

Nonostante l'attenzione per gli approcci a modelli causali sia piuttosto recente⁵, e abbia subito una sostanziale accelerazione grazie al monumentale *Causality: Models, Reasoning, and Inference* di **Pearl**; l'utilizzo di approcci basati sulle equazioni strutturali e su modelli che descrivono relazioni

¹ Si veda: J. Schaffer, «Grounding in the image of causation», *Philosophical Studies*, vol. 173, n. 1 (2016), pp. 49-100, DOI: [10.1007/s11098-014-0438-1](https://doi.org/10.1007/s11098-014-0438-1).

² J. Y. Halpern e J. Pearl, «Causes and Explanations: A structural-model approach. Part I: Causes», *British Journal for the Philosophy of Science*, vol. 56, n. 4 (2005), pp. 843-887, DOI: [10.1093/bjps/axi147](https://doi.org/10.1093/bjps/axi147).

³ **Halpern**, *Actual Causality* cit.

⁴ Quando, per ragioni di economia, verrà usato un differente formalismo, verrà annotato.

⁵ Risalgono, infatti, al 2000 i primi segnali di un interesse per questi strumenti.

di tipo causale è nota sin dagli anni '20 e si deve in particolare a **Wright** e al suo seminale lavoro «Correlation and causation»⁶, seguito da «The Method of Path Coefficients»⁷. Il lavoro di **Wright** prende piede in particolare nelle scienze sociali⁸ ed è, ancora oggi, un approccio molto praticato e approfonditamente studiato⁹, pur nelle più mature e complete versioni offerte, tra gli altri da **Pearl**¹⁰ e da **Spirtes et al.**¹¹. Dopo questo brevissimo *excursus* storico, procediamo con la caratterizzazione dei modelli causali.

Un modello causale è la rappresentazione di una porzione di mondo descritta attraverso delle variabili e le relazioni che tra queste sussistono. Una variabile è una funzione che può assumere diversi valori in grado di rappresentare eventi, proprietà o quantità. Il valore di una variabile è l'istanziamento di una proprietà, appartenente al *range*¹² della variabile da parte di un particolare o di un individuo. La scelta di variabili non è costretta da alcuna regola, ma essa determina, naturalmente, la bontà di un modello. Del problema della scelta delle variabili e della cosiddetta “arte di fare modelli causali” ci occuperemo nel **Capitolo 3**. Generalmente, ogni variabile selezionata ha un'influenza causale su almeno un'altra variabile. Al fine di catturare le relazioni causali tra variabili e modellarle attraverso un formalismo economico, vengono usate delle *equazioni strutturali*: queste sono delle funzioni su cui sono possibili normali operazioni matematiche, che incorporano relazioni di tipo controfattuale. Si fa notare che la precisazione fatta prima riguardo ai valori concepiti come istanziati da particolari o individui è qui rilevante, in quanto le equazioni strutturali non vanno concepite come rappresentanti un effetto medio di una causa su di una popolazione, bensì come dei controfattuali che quantificano su singoli individui. Di conseguenza, le equazioni strutturali rappresentano il rapporto funzionale tra variabili a livello individuale e non a livello di risposta media ad una variazione media di una certa variabile.

Risultano evidenti, già da questo momento, due caratteristiche dei modelli causali:

- I modelli causali danno per presupposta una qualche nozione di causalità, in quanto il modello contiene variabili che hanno influenza di tipo causale tra di loro, di conseguenza non sono impiegati in trattazioni riduttive e potrebbero presentare dei rischi legati alla circolarità, specie se cerchiamo di definire un vocabolario causale in termini di un linguaggio definito su di essi. Una breve discussione di questa problematica è svolta nella **Sezione 1.2.1**.
- Le variabili che scegliamo “determinano il linguaggio usato per rappresentare la situazione”¹³, perché la relazione di causalità sussisterà tra ciò che le variabili rappresentano.

Le variabili possono essere partizionate in due gruppi, a seconda del tipo di rapporti causali (rappresentati dalle equazioni strutturali) che intrattengono: abbiamo quindi variabili *esogene* e variabili *endogene*. Le prime rappresentano variabili che potremmo definire ‘esterne’ al modello: queste determinano un contesto, sono variabili che consideriamo in un certo senso statiche, ovvero su cui

6 S. Wright, «Correlation and causation», *Journal of agricultural research*, vol. 20, n. 3 (1921).

7 S. Wright, «The Method of Path Coefficients», *The Annals of Mathematical Statistics*, vol. 5, n. 3 (1934), pp. 161-215, DOI: [10.1214/aoms/1177732676](https://doi.org/10.1214/aoms/1177732676).

8 Una panoramica, anche se un po' datata, è offerta in A. S. Goldberger, «Structural Equation Methods in the Social Sciences», *Econometrica*, vol. 40, n. 6 (1972), p. 979, DOI: [10.2307/1913851](https://doi.org/10.2307/1913851).

9 F. Russo, *Causality and Causal Modelling in the Social Sciences*, Springer Netherlands, 2009, DOI: [10.1007/978-1-4020-8817-9](https://doi.org/10.1007/978-1-4020-8817-9).

10 J. Pearl, *Causality: Models, Reasoning, and Inference*, 2. Cambridge University Press, 2009.

11 P. Spirtes et al., *Causation, Prediction, and Search*, MIT Press, 2000.

12 Questa nozione verrà formalizzata successivamente. Il *range* di una variabile è l'insieme di valori che quella variabile può assumere.

13 Halpern, *Actual Causality* cit. Traduzioni mie per tutti i passaggi dall'inglese

non possiamo intervenire, poste dall'esterno, che non hanno equazioni strutturali associate e di cui, generalmente, non siamo interessati a studiare il comportamento; le seconde rappresentano variabili interne al modello, sulle quali vogliamo intervenire e il cui comportamento e la cui variazione sono, in ultima istanza, dettate dalle variabili esogene. Ogni variabile endogena ha una ed unica equazione strutturale associata; una variabile esogena è, invece, priva di equazione di riferimento, il che implica che il suo valore è determinato dall'esterno del modello.

La trattazione seguente dei modelli causali è basata su *Actual Causality*¹⁴ di Halpern, altri formalismi isomorfi sono possibili¹⁵. Durante tutta la trattazione verrà assunto, per comodità, il determinismo, anche se una trattazione probabilistica non comporta alcuna problematica:

DEFINIZIONE 1.1.1 (MODELLO CAUSALE)

Un modello causale M è una coppia $\langle S, \mathcal{F} \rangle$, dove S è una segnatura e \mathcal{F} è un insieme di equazioni strutturali.

Un modello causale si configura, quindi, come una costruzione a molteplici livelli. Prima di tutto c'è la *segnatura*, che è, sommariamente, una descrizione della situazione che vogliamo rappresentare.

DEFINIZIONE 1.1.2 (SEGNATURA)

Una segnatura S è un tripla $\langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$ dove \mathcal{U} è l'insieme delle variabili esogene, \mathcal{V} è l'insieme delle variabili endogene e \mathcal{R} una funzione che associa ad ogni $X \in \mathcal{U} \cup \mathcal{V}$ un insieme non vuoto di valori possibili per X .

L'insieme \mathcal{U} rappresenta le *condizioni iniziali del sistema*, l'insieme \mathcal{V} raccoglie le variabili utilizzate per la descrizione dell'*evoluzione delle condizioni del sistema*: grossomodo, si tratta degli aspetti della dinamica del sistema che siamo interessati a catturare. La funzione \mathcal{R} , invece, mappa i differenti possibili valori di ogni variabile: questo fornisce, seguendo Blanchard e Schaffer, lo *spazio di contrasto delle condizioni del sistema*, ovvero le possibili situazioni alternative che possono verificarsi (anche solo a livello ideale) nel sistema¹⁶.

Vediamo ora qual è la seconda componente di un modello causale, ovvero la funzione \mathcal{F} .

DEFINIZIONE 1.1.3 (FUNZIONE \mathcal{F})

La funzione \mathcal{F} associa ad ogni variabile $Y \in \mathcal{V}$ una funzione F_Y che associa ad ogni n -upla $\times_{X \in (\mathcal{U} \cup \mathcal{V} \setminus \{Y\})} \mathcal{R}(X)$ un valore $y \in \mathcal{R}(Y)$ ¹⁷.

La funzione \mathcal{F} , così definita, mappa il valore di ogni variabile al valore di ogni altra variabile nel modello e rappresenta, grossomodo, la *dinamica del sistema*, ovvero le leggi (modificabili e non definitive) che ne governano l'evoluzione. Vedremo successivamente che gran parte dello spazio $|\mathcal{U} \cup \mathcal{V}|$ -dimensionale del dominio è, in realtà, ridondante e la funzione F_X che \mathcal{F} associa ad ogni variabile X può essere determinata da un minor numero di variabili.

È di norma abbreviare notazioni funzionali in forma di equazioni dirette, del seguente tipo $Y = X + Z + U$. Nonostante la forma sia ingannevole, l'equazione presentata non va interpretata come una normale equazione matematica, bensì come un'*equazione strutturale*: questo significa che il segno '=' non va inteso come un'uguaglianza simmetrica, bensì va interpretato come un'assegnazione

¹⁴ Ivi.

¹⁵ Si vedano in particolare: Pearl, *Causality: Models, Reasoning, and Inference* cit.; Spirtes et al., *Causation, Prediction, and Search* cit.

¹⁶ Blanchard e Schaffer, «Cause without Default» cit.

¹⁷ Il simbolo \times rappresenta il prodotto cartesiano dell'insieme.

del lato destro al lato sinistro: nella letteratura è, infatti, comune il paragone del segno '=' con il segno ':=' , che in molti linguaggi di programmazione rappresenta l'assegnazione di un valore ad una variabile. Per parafrasare Pearl¹⁸, un'equazione strutturale del tipo $Y = X + U$ vuole rappresentare un meccanismo, un processo di generazione di dati (o informazioni), su cui è possibile intervenire ma che non possiamo 'rovesciare' ($X = Y - U$): il flusso di informazione ha un solo ed unico verso. Le funzioni appartenenti a \mathcal{F} non sono, quindi, in principio, invertibili. In generale, quando verranno trattate variabili binarie, queste verranno identificate con proposizioni: di conseguenza verranno utilizzati i connettivi logici \vee, \wedge, \neg come equivalenti delle seguenti funzioni: $\min, \max, 1 - X$.

Questa imposizione di asimmetria risulta direttamente dal fatto che l'obiettivo è modellare una relazione di influenza causale¹⁹: ci permette, infatti, di caratterizzare Y come determinata dalle variabili sul lato destro dell'equazione, ma, al contempo, ci impedisce di affermare che queste variabili siano determinate dal valore di Y in maniera diretta²⁰, come diremmo che se X causa Y , Y non causa X ²¹.

In una cornice metodologica di questo tipo formulare controfattuali è estremamente semplice: dato che lo 'stato' del nostro modello è interamente determinato dai valori che le variabili hanno è sufficiente modificare il valore di queste ultime per generare tutti i controfattuali che desideriamo. Porre il valore di una variabile X a x in un modello causale $M = (\mathcal{S}, \mathcal{F})$ produce un nuovo modello causale $M_{X \leftarrow x}$. Questo modello si ottiene sostituendo la funzione F_X con la costante $X = x$, riducendo la segnatura \mathcal{S} a $\mathcal{S}_X = (\mathcal{U}, \mathcal{V} \setminus \{X\}, \mathcal{R}_{(\mathcal{U}, \mathcal{X} \setminus \{X\})})$ e riducendo \mathcal{F} a $\mathcal{F}^{X=x}$, ovvero eliminando dalla funzione \mathcal{F} le funzioni F_X e sostituendo le funzioni F_Y con $F_Y^{X=x}$, eliminando dal dominio delle funzioni X , ovvero $\times_{Z \in \mathcal{U} \cup \mathcal{V} \setminus \{Y, X\}} \mathcal{R}(Z)$ e sostituendo X con x . A questo punto, tenendo conto della seguente definizione:

DEFINIZIONE 1.1.4 (MONDO)

Un mondo w è un modello causale M in cui ogni variabile endogena, per intervento o soluzione delle equazioni, è collassata ad una costante.

è possibile far collassare le altre variabili nel modello $M_{X \leftarrow x}$ per generare i nostri controfattuali: se vogliamo sapere che valore avrebbe avuto la variabile Y se la variabile X avesse avuto il valore x , è sufficiente calcolare la funzione F_Y nel modello $M_{X \leftarrow x}$ ²². Si noti che, nel nostro caso, ovvero in cui vengano trattati solo modelli causali aciclici, dato un contesto \vec{u} , il modello collassa necessariamente ad un mondo. È, come detto poco sopra, possibile produrre un nuovo mondo attraverso l'imposizione di un valore ad una variabile.

D'ora in avanti verrà posto il requisito di minimalità alle equazioni, ovvero: la variabile designata al lato sinistro dipende in maniera diretta solo ed esclusivamente dalle variabili presenti al lato destro. Con dipendenza diretta si intende una relazione definita nel seguente modo:

¹⁸ Pearl, *Causality: Models, Reasoning, and Inference* cit.

¹⁹ In generale le equazioni strutturali sono un elegante formalismo per la modellazione di molte relazioni asimmetriche: si veda, per esempio, la proposta per il concetto di *grounding* in Schaffer, «Grounding in the image of causation» cit. Questo deriva anche dalla possibilità di tradurre queste equazioni in grafici direzionati, una tematica che verrà affrontata in seguito.

²⁰ Questo non impedisce di usare la probabilità condizionata per inferire il valore di variabili del lato destro data l'osservazione del valore della variabile del lato sinistro. I modelli causali, peraltro, sono spesso utilizzati con funzione di identificazione: Pearl, *Causality: Models, Reasoning, and Inference* cit.:cap.2-3.

²¹ Vi sono alcune marginali situazioni in cui non possiamo fare questo tipo di affermazioni, e che il presente scritto lascerà in larga parte ignorate.

²² D'ora in poi quando scriveremo $F_Y(u, x)$ intenderemo il valore di F_Y calcolato nel modello $M_{X \leftarrow x, U \leftarrow u}$.

DEFINIZIONE 1.1.5 (DIPENDENZA DIRETTA)

Se $\exists(x, x', \vec{z}) : F_Y(x, \vec{z}) \neq F_Y(x', \vec{z})$ con $x \neq x'$, $x \in \mathcal{R}(X)$, $x' \in \mathcal{R}(X)$, $\vec{Z} = \mathcal{U} \cup \mathcal{V} \setminus \{X, Y\}$, $\vec{z} \in \mathcal{R}(\vec{Z})$; allora Y dipende direttamente da X ²³.

In parole, Y dipende direttamente da X se esiste almeno un valore delle variabili del modello, escluse X e Y , tale che se *congeliamo* tali variabili a quel valore, esistono almeno due valori di X tali che il passaggio di X da uno all'altro di questi valori comporta una variazione del valore di Y . Con il formalismo introdotto in precedenza: i valori di Y nei modelli $M_{X \leftarrow x}$ e $M_{X \leftarrow x'}$ sono diversi. Si noti che, come discute ampiamente Woodward²⁴, l'intuizione, peraltro naturale, che la dipendenza diretta dovrebbe richiedere variazione simultanea di X e Y , in linea con una perfetta correlazione, è scorretta: è facile immaginare la funzione $F_Y(X)$ in modo tale che mappi due valori diversi di X allo stesso valore di Y , quindi $F_Y(x_1) = F_Y(x_2) \neq F_Y(x_3)$. In termini matematici, se la funzione F_Y non è iniettiva, richiedere che la variazione lungo tutto $\mathcal{R}(X)$ corrisponda ad una variazione in Y , porterebbe a negare la relazione di dipendenza diretta. Avremo, nel prosieguo, diversi casi in cui funzioni di questo tipo saranno necessarie per modellare diverse situazioni causali (generalmente casi in cui la causalità è *multifattoriale* o poco sensibile).

Un altro dettaglio che è necessario enfatizzare è il fatto che la definizione richiede che esista una qualche combinazione di valori esterni alla variabili da congelare in cui si verifichi la simultanea variazione. Questa specificazione è necessaria per accomodare casi simili a quello presentato nell'Equazione 1.1b, in cui, se la variabile X_1 fosse congelata ad un valore $x_1 = 0$, sembrerebbe non esserci dipendenza diretta tra X_2 e Y . È quindi necessario richiedere che sia possibile una manipolazione del *setting*, per così dire, per far emergere la relazione di dipendenza diretta. Questa caratteristica sarà interessante per la trattazione nella Capitolo 4 di alcuni fenomeni particolari di causazione per prevenzione e simili, in cui vedremo che questa richiesta di variazione di *setting* verrà di fatto emendata.

Il requisito di minimalità descritto in precedenza è esplicabile in modo forse più chiaro ricorrendo ad una parafrasi della nozione di *genitori Markoviani* proposta da Pearl²⁵, che rovescia parzialmente la prospettiva:

DEFINIZIONE 1.1.6 (GENITORI MARKOVIANI)

Sia $\mathcal{V} = \{X_1, \dots, X_n\}$ un insieme ordinato di variabili. Un insieme di variabili PA_j è detto *genitore Markoviano* di X_j se PA_j è l'insieme minimale di antenati di X_j tale da rendere X_j indipendente dai suoi altri antenati.

La corretta definizione di minimalità di un'equazione per una variabile è, quindi, la seguente:

DEFINIZIONE 1.1.7 (MINIMALITÀ)

Si dice *minimale* un'equazione rispetto ad una variabile se essa presenta, al lato destro, solamente i genitori Markoviani della variabile in questione.²⁶

²³ Questa definizione è fondamentalmente identica a quella di *effetto diretto* data da Pearl, con la differenza che l'*effetto diretto* è la misura della relazione di (in)dipendenza (*ivi*).

²⁴ J. Woodward, *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, 2004, DOI: [10.1093/0195155270.001.0001](https://doi.org/10.1093/0195155270.001.0001).

²⁵ Pearl, *Causality: Models, Reasoning, and Inference* cit.: p. 36.

²⁶ Non ci sono problemi ad estendere la definizione in termini probabilistici in questo caso, ed è, anzi, la presentazione che propone Pearl in Pearl, *Causality: Models, Reasoning, and Inference* cit.; A. Balke e J. Pearl, «Counterfactuals and Policy Analysis in Structural Models», in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, Morgan Kaufmann Publishers Inc., Montréal, Qué, Canada 1995, pp. 11-18.

Riportiamo un esempio proposto da Woodward e Hitchcock²⁷ al fine di chiarire la maglia larga della definizione a riguardo dell'esistenza di almeno un valore della variabile X che generi un cambiamento nella variabile Y , e non che vi sia un *continuum* di variazione tra le due (ovvero che vi sia perfetta correlazione): immaginiamo un interruttore, in grado di ruotare di 180° , collegato ad una lampadina. Se l'interruttore è ruotato rispetto all'asse di meno di 90° , allora la lampadina non si accende, superati i 90° si accende. Abbiamo qui un'evidente relazione causale tra la posizione dell'interruttore e la lampadina, ma l'equazione strutturale per lo stato della lampadina (accesa o spenta) è la seguente:

$$L = \begin{cases} 1 & \text{se } I \geq 90^\circ \\ 0 & \text{se } I < 90^\circ \end{cases}$$

dove L rappresenta lo stato della lampadina e I rappresenta la posizione dell'interruttore. È evidente che non tutte le variazioni all'interno del range $\mathcal{R}(I)$ portino ad una modificazione dello stato di L .

Equipaggiati con la nozione di dipendenza diretta, possiamo ora introdurre un formalismo grafico per rappresentare in maniera qualitativa (ma non quantitativa) le informazioni causali che le equazioni strutturali codificano. Si disegni un grafo attraverso la seguente procedura:

1. Si prendano le variabili appartenenti all'insieme $\mathcal{U} \cup \mathcal{V}$ e si disegni un nodo per ognuna di queste variabili.
2. Si tracci, tra tutte le coppie di variabili tra cui sussiste la relazione di dipendenza diretta, un segmento direzionato, con una freccia terminale a rappresentarne la direzione, tale che la freccia terminale punti alla variabile dipendente.

Abbiamo ora un grafo che rappresenta le relazioni di dipendenza causale (diretta) tra le variabili del modello e, di conseguenza, anche quelle di indipendenza: una freccia non presente implica, infatti, che vi sia una relazione di indipendenza diretta tra le due variabili. Halpern²⁸ indica questo tipo di rappresentazione con la dicitura *network causale*²⁹. Caratteristica di questo grafo è la capacità di fornirci una rappresentazione qualitativa (seppur in un senso debole) dei rapporti tra le variabili in gioco, ma non di darci indicazioni a livello quantitativo. Inoltre, la granularità dell'informazione è piuttosto elevata, in quanto i due seguenti rapporti funzionali vengono rappresentati nel medesimo modo all'interno di un *network causale*:

$$Y = \alpha X_1 + \beta X_2 \quad (1.1a)$$

$$Y = X_1 \cdot X_2 \quad (1.1b)$$

Si osservi l'Equazione 1.1a: il rapporto, essendo meramente additivo, è radicalmente differente da quello presente nell'Equazione 1.1b, moltiplicativo. Se una delle due variabili risulta nulla nell'Equazione 1.1a, la variabile Y seguirà, parametricamente, le variazioni della variabile non nulla;

²⁷ J. Woodward e C. Hitchcock, «Explanatory Generalizations, Part I: A Counterfactual Account», *Noûs*, vol. 37, n. 1 (2003), pp. 1-24, DOI: 10.1111/1468-0068.00426.

²⁸ Halpern, *Actual Causality* cit.

²⁹ Si noti che con questa dicitura si intende un grafo che rappresenta relazioni di tipo causale. Lo stesso formalismo grafico può essere utilizzato per rappresentare compiti di compressione, pianificazione, ecc., ed ha un uso diffuso in informatica e altre discipline.

nell'Equazione 1.1b, invece, il valore nullo di una delle due variabili implica la nullità della variabile Y in toto. Per questo l'informazione ottenibile dai *network causali* è *debolmente* qualitativa. Si noti, inoltre, che ogni modello genera un unico grafo, ma che modelli con la stessa cardinalità di variabili e le stesse relazioni di parentela producono *network causali* identici: le Equazioni 1.1a e 1.1b producono, infatti, il medesimo *network causale*.

Si noti che, dato che le variabili esogene $U: U \in \mathcal{U}$ non dipendono da alcuna variabile per definizione, di conseguenza non avranno alcuna freccia che punta ad esse³⁰; le variabili endogene $X: X \in \mathcal{V}$ avranno tutte almeno una freccia che punta ad esse, proprietà che deriva loro dal fatto di dipendere almeno da una variabile esogena. Nonostante questo, spesso verranno omesse le variabili esogene.

Osserviamo, in Figura 1.1 due rappresentazioni possibili del seguente modello causale:

$$\begin{aligned} S &= U_1 \\ F &= U_2 \\ E &= \min(S, F) \end{aligned}$$

dove U_1 e U_2 sono le variabili esogene in quanto non determinate da alcun'altra variabile.

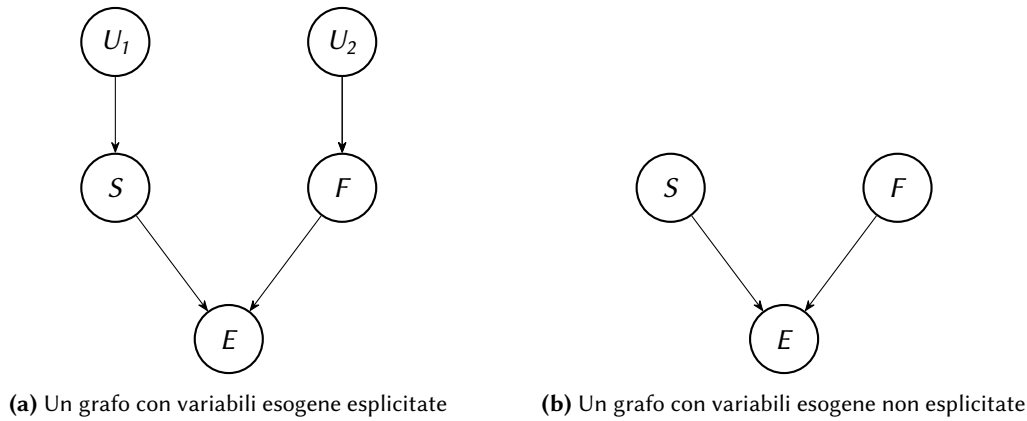


Figura 1.1: Due network causali dello stesso modello

Una particolare caratteristica di questi due grafi, che risulterà fondamentale per la presente trattazione, è il fatto di essere *aciclici*, o anche *ricorsivi*. Questa è una proprietà che deriva ai grafi dal modello di cui sono rappresentazioni. Un grafico è aciclico, intuitivamente, se, partendo da un nodo rappresentante una variabile, non è possibile tornare allo stesso nodo seguendo un percorso direzionato secondo le frecce (ovvero seguendo le frecce dalla coda alla punta). Riportiamo la definizione di modello causale aciclico formulata da Halpern³¹:

DEFINIZIONE 1.1.8 (ACICLICITÀ DI UN MODELLO CAUSALE)

Un modello causale è *aciclico* se vi è un ordine parziale \leq in \mathcal{V} tale che se $X \not\leq Y$, allora Y non è influenzata da X ³².

³⁰ In realtà non sorge alcun problema per questa trattazione se vi sono relazioni di dipendenza diretta tra variabili esogene, purché siano riconosciute e non generino correlazioni spurie; l'importante è che non vi siano relazioni di dipendenza tra variabili esogene e variabili endogene non rappresentate o non riconosciute.

³¹ *Ivi*.

³² Questa condizione equivale alla condizione che il grafo generato dal modello causale abbia un *ordine topologico*, ovvero che sia possibile ordinare i vertici del grafo in modo tale che, per ogni segmento direzionato, il vertice di partenza appaia prima rispetto al vertice finale.

Si osservino i grafi in **Figura 1.2**: nel caso del primo grafo (**Figura 1.2a**) non vi sono cicli, in quanto, partendo da una variabile qualsiasi diversa da W (che, avendo solo frecce entranti non ha cammini verso l'esterno), si sia costretti a passare per W , finendo *intrappolati* prima di poter completare il ciclo. A titolo di esempio, partendo dalla variabile X : $X \rightarrow Y \rightarrow W$, oppure $X \rightarrow Z \rightarrow W$, e in entrambi i casi il cammino risulta interrotto. Non vale lo stesso per nel caso del secondo grafo (**Figura 1.2b**), nel quale W non agisce da *collisore*³³ globale, e Z ha una relazione di dipendenza con X , permettendo l'instaurarsi del ciclo $X \rightarrow W \rightarrow Z \rightarrow X$. Si noti che non è necessaria, ne sufficiente, per un grafo, l'assenza di un *collisore* per renderlo ciclico, e viceversa.



Figura 1.2: Esempi di grafi aciclici e non aciclici

Dato che le relazioni rappresentate da un *network causale* sono, come già detto, asimmetriche, è costume utilizzare un lessico di tipo parentale quando ci si riferisce alle variabili all'interno di un grafo di questo tipo (e, per estensione, anche all'interno del modello causale). Data una variabile Y , una variabile X è un *genitore* di Y sse Y dipende direttamente da X ; viceversa, se, data una variabile Y , una variabile Z è un *figlio* di Y sse Z dipende direttamente da Y . Inoltre, se vi è un cammino direzionato da X a Y , allora X è un antenato di Y e Y è un discendente di X .

Una caratteristica peculiare dei modelli causali aciclici è che, una volta specificato una valore \vec{u} per le variabili esogene, è possibile collassare il modello causale ad un mondo. \vec{u} è chiamato da **Halpern**³⁴ un *contesto*. L'unione di un modello causale ed un contesto è definita da **Pearl** un *mondo causale*:

DEFINIZIONE 1.1.9 (MONDO CAUSALE)

Un mondo causale è una coppia (M, \vec{u}) , dove M è un modello causale e \vec{u} è una particolare realizzazione delle variabili in \mathcal{U} .³⁵

Si noti che la definizione di **Pearl** di *mondo causale* è una specificazione della **Definizione 1.1.4**, in quanto, nel caso in cui il modello non sia aciclico, un *mondo causale* non produce un mondo nel senso della **Definizione 1.1.4**. **Pearl** definisce anche la nozione di *teoria causale*:

DEFINIZIONE 1.1.10 (TEORIA CAUSALE)

Una teoria causale è un insieme di mondi causali.

Una teoria causale rappresenta i vari mondi causali generati dai diversi contesti \vec{u} . Una teoria causale è, fondamentalmente, una coppia formata da un modello causale M e un insieme di contesti $\{\vec{u}_1, \dots, \vec{u}_n\}$ applicabili al modello causale.

Concludiamo presentando un modello causale per il classico scenario del *lancio del sasso*:

³³ **Pearl** usa questo termine per definire i nodi con almeno due frecce entranti: **Pearl**, *Causality: Models, Reasoning, and Inference* cit. A titolo di esempio, la variabile W nel cammino $X \rightarrow W \leftarrow Z$ in **Figura 1.2a** è un collisore.

³⁴ **Halpern**, *Actual Causality* cit.

³⁵ **Pearl**, *Causality: Models, Reasoning, and Inference* cit.: la notazione vettoriale è stata introdotta da me per questioni di uniformità; **Pearl** usa u con il medesimo significato.

ESEMPIO 1.1.1 (Rock Throwing)

Billy e Suzy, due monellacci, cercano di colpire una bottiglia di vetro lanciando dei sassi. Suzy lancia il sasso una frazione di secondo prima di Billy e il suo sasso colpisce e manda in frantumi la bottiglia. Una frazione di secondo dopo, il sasso di Billy passa sopra i frammenti della bottiglia. ■

Un modello causale che potrebbe rappresentare questa storia è il seguente, dove si assumono tutte variabili binarie con valore 1 se l'evento si verifica, o altrimenti:

$$\begin{aligned} SL &= U_1 && \text{con } SL \text{ rappresentante il lancio di Suzy} \\ BL &= U_2 && \text{con } BL \text{ rappresentante il lancio di Billy} \\ BR &= SL \vee BL && \text{con } BR \text{ rappresentante la rottura della bottiglia} \end{aligned}$$

il grafo generato da questo modello causale è rappresentato in **Figura 1.3**, dove i nodi rappresentati le variabili esogene non sono rappresentati.

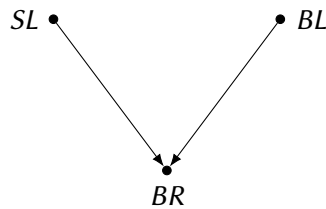


Figura 1.3: Grafo generato dal modello causale della storiella di Suzy e Billy

Un modello alternativo, che tiene conto del fatto che il lancio di Suzy avviene prima e che, una volta rotta, la bottiglia non può essere rotta nuovamente, è generabile variando le equazioni del modello nel seguente modo, il cui grafo è rappresentato in **Figura 1.4**:

$$\begin{aligned} SL &= U_1 && \text{con } SL \text{ rappresentante il lancio di Suzy} \\ BL &= U_2 && \text{con } BL \text{ rappresentante il lancio di Billy} \\ SC &= SL && \text{con } SC \text{ rappresentante il sasso di Suzy che colpisce la bottiglia} \\ BC &= BL \wedge \neg SC && \text{con } BC \text{ rappresentante il sasso di Billy che colpisce la bottiglia} \\ BR &= SC \vee BC && \text{con } BR \text{ rappresentante la rottura della bottiglia} \end{aligned}$$

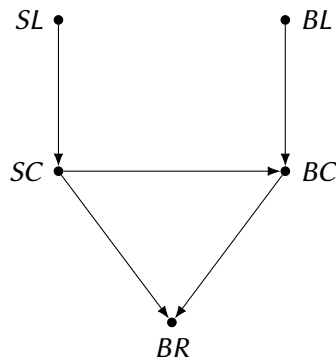


Figura 1.4: Grafo generato dal modello causale alternativo della storiella di Suzy e Billy

Questo tipo di rappresentazioni offre enormi vantaggi rispetto alla classica rappresentazione, diventata ormai una *lingua franca* tra chi si occupa di causalità, a *diagrammi neuronali*, proposti per la prima volta da Lewis³⁶. Innanzitutto possiedono una più chiara traduzione in termini immediatamente controfattuali: le frecce rappresentano non segnali stimolatori, ma dipendenze dirette che, per definizione, sono anche dipendenze controfattuali.

Un'altra problematica, che fa notare Hitchcock³⁷, è che essi non sono abbastanza *potenti*, nel senso che non sono in grado di rappresentare un gran numero di situazioni causali. Per esempio, è complicato produrre un diagramma neuronale che rappresenti la seguente relazione: $B = \neg A$. Che tipo di segnale A invia a B ? Inibitorio? Stimolatorio?

In realtà nessuno delle due tipologie di segnale. La classica soluzione che viene utilizzata è quella di proporre un diagramma con tre neuroni, di cui uno svolge ruolo stimolatorio, l'altro, ovvero il nostro A , quello inibitorio (rappresentato dalla freccia che termina con un cerchio).



Figura 1.5: Rappresentazione di $B = \neg A$ via diagramma aumentato

Un'altra soluzione, ed è quella tendenzialmente più utilizzata quando si tratta di rappresentare situazioni causali complesse, è quella di creare rappresentazioni con nuovi tipi di neuroni (con pareti più spesse, colorati ecc.), come rappresentato in Figura 1.6.



Figura 1.6: Rappresentazione di $B = \neg A$ via convenzione grafica

Questa soluzione, però, comporta un importante ostacolo: permettendo una rappresentazione *ad hoc* per ogni situazione causale, non pone alcun vincolo per quanto riguarda strutture causali possibili: non vi è più nessuna struttura causale logicamente impossibile, in quanto siamo sempre in grado di inventare nuovi neuroni. Questo pone delle problematiche per quanto riguarda la scoperta e l'analisi di nuove strutture causali, in quanto aumenta lo spazio di ricerca di queste strutture esponenzialmente.

1.2 INTERVENTI

Come abbiamo visto, i modelli causali incorporano controfattuali, e la loro valutazione appare, a prima vista, estremamente elementare, anche se in realtà vi sono alcune particolarità interessanti che sono implicate dal fatto che, quando si valuta un controfattuale all'interno di un modello causale, si interviene, di fatto, sul valore della variabile e, di conseguenza, sulle *equazioni strutturali*.

In questa sezione verrà affrontato innanzitutto il concetto di intervento, un concetto chiave per evitare rischi di circolarità in sede di definizione di nozioni causali, in quanto la nozione di

³⁶ D. Lewis, «Postscripts to 'Causation'», in *Philosophical Papers*, Oxford University Press, 1986, vol. II.

³⁷ C. Hitchcock, «What's Wrong with Neuron Diagrams?», in J. K. Campbell *et al.* (a cura di), *Causation and Explanation*, Topics in Contemporary Philosophy, Bradford Books, Cambridge, MA 2007, pp. 69-92.

intervento viene considerata, nei **SEMs**, primitiva e, quindi, se definita in termini causali, potenzialmente in grado di generare problematiche di circolarità. Vedremo che questo rischio va fortemente ridimensionato e non si rivela problematico in quanto a circolarità.

Verrà successivamente presentata una semantica³⁸ per la valutazione dei controfattuali piuttosto ricca, anche se di fatto limitata rispetto alla la classica semantica a mondi possibili proposta da Lewis³⁹.

1.2.1 Che cos'è un intervento

Una cornice metodologica basata sui modelli causali richiede, da un punto di vista filosofico, un impegno nei confronti di una teoria causale di tipo *manipolazionista*⁴⁰. La nozione di intervento, infatti, risulta fondamentale per questo tipo di cornici e, se nelle scienze sociali, econometria ecc., è concessa una nozione di intervento definita in maniera vaga, in questa sede non si può dire altrettanto.

Come precedentemente annotato, quando ci proponiamo di utilizzare i modelli causali per parlare di causalità, il rischio di circolarità è, per così dire, sempre dietro l'angolo: è necessaria una profonda attenzione alla definizione e specificazione del vocabolario e dei concetti che vogliamo utilizzare se vogliamo evitare rischi di regresso nella nostra analisi⁴¹. Pertanto, in questa sezione, analizzeremo la nozione di intervento, riconoscendone, certamente, la natura causale, ma presenteremo una revisione del concetto, proposta da Woodward, che permetta di, per così dire, deviare il *focus* della nozione di causalità che lo definisce al fine di eliminare rischi di circolarità viziosa.

Innanzitutto impostiamo la discussione presentando alcune delle problematiche che riguardano la nozione di intervento:

- Intervenire è diverso da osservare: il concetto di intervento su di una variabile non è l'equivalente di osservare la variabile prendere naturalmente il valore impostato⁴². Questo perché,

³⁸ Tra le tante possibili quella sviluppata da Briggs è probabilmente la più idonea: R. Briggs, «Interventionist counterfactuals», *Philosophical Studies*, vol. 160, n. 1 (2012), pp. 139-166, DOI: [10.1007/s11098-012-9908-5](https://doi.org/10.1007/s11098-012-9908-5). Due le principali alternative: J. Y. Halpern, «Axiomatizing Causal Reasoning», *Journal of Artificial Intelligence Research*, vol. 12 (2000), pp. 317-337, DOI: [10.1613/jair.648](https://doi.org/10.1613/jair.648) (seppur non completa come trattazione e che Briggs punta ad espandere e correggere, sacrificando la possibilità di gestire modelli non ricorsivi) e J. Zhang, «A Lewisian Logic of Causal Counterfactuals», *Minds and Machines*, vol. 23, n. 1 (2013), pp. 77-93, DOI: [10.1007/s11023-011-9261-z](https://doi.org/10.1007/s11023-011-9261-z). Nonostante questo, qui si affronterà la proposta di Hiddleston (Hiddleston, «A Causal Theory of Counterfactuals» cit.), che verrà trattata nella Sezione 1.3.1.

³⁹ D. Lewis, *Counterfactuals*, Harvard University Press, 1973.

⁴⁰ La nozione di *manipolazione* è utilizzata principalmente da Woodward (Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.). Cartwright (N. Cartwright, «Two Theorems on Invariance and Causality», *Philosophy of Science*, vol. 70, n. 1 (2003), pp. 203-224, DOI: [10.1086/367876](https://doi.org/10.1086/367876)) fa notare che la nozione è problematica in quanto a rischio di risultare troppo forte. La preoccupazione di Cartwright sembra, da come è espressa, principalmente pragmatica: requisiti troppo stretti su test di causalità, non praticabilità in *setting* sperimentali, ecc. Queste preoccupazioni esulano dall'argomento di questo scritto e resteranno, quindi, in larga parte ignorate, pur riconoscendo l'importanza di una corretta impostazione a fini pragmatici del concetto di manipolazione.

⁴¹ Questo vale anche se la proposta di cui qui si tratta non è riduttiva. Infatti, la problematica della circolarità, come estesamente riportato in letteratura (Broadbent, *Philosophy of Epidemiology* cit.; N. Hall e L. A. Paul, *Causation: A User's Guide*, Oxford University Press, 2013; Halpern, *Actual Causality* cit.; Pearl, *Causality: Models, Reasoning, and Inference* cit.; J. Pearl, «Reply to Woodward», *Economics and Philosophy*, vol. 19, n. 2 (2003), pp. 341-344, DOI: [10.1017/S0266267103001196](https://doi.org/10.1017/S0266267103001196); Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.; J. Woodward, «Critical Notice: Causality by Judea Pearl», *Economics and Philosophy*, vol. 19, n. 2 (2003), pp. 321-340, DOI: [10.1017/S0266267103001184](https://doi.org/10.1017/S0266267103001184)), è assolutamente presente anche in questo tipo di analisi e, facilmente, è in grado di inficiare proposte, altrimenti solide, di teorie riguardanti la causazione in generale, causazione effettiva o anche solo teorie epistemiche e concentrate sulla spiegazione basate sugli approcci a modelli causali. È, pertanto, necessario indirizzare questa problematica.

⁴² Questo vale nella quasi totalità dei casi. Vi sono delle eccezioni nei casi in cui l'osservazione del valore di una variabile non comporta la modifica delle distribuzioni dei valori nelle altre variabili esclusi i propri figli.

come Pearl chiarisce perfettamente⁴³, osservare una variabile ad un certo valore implica che le variabili antenate della variabile osservata possiedano certi valori, il che influisce su tutto il modello. In termini matematici, in generale $P(Y = y \mid X = x) \neq P(Y = y \mid do(X = x))$ ⁴⁴.

- Alcuni interventi sono fisicamente impossibili. Ad esempio risulta impossibile modificare la forza di attrazione gravitazionale tra la Terra e il Sole.
- Interventi *imprecisi*: quali sono i requisiti che un intervento deve avere al fine di preservare il meccanismo causale (o presunto tale) che vogliamo indagare ed evitare perturbazioni che inficino il nostro tentativo? Che tipo di manipolazione va eseguita per garantire la sicurezza delle nostre conclusioni?
- Strettamente connesso al punto precedente: è necessaria la *modularità* del sistema per potervi intervenire in maniera proficua? Per asserire la causalità è necessaria la *modularità*? Con *modularità* si intende, fondamentalmente, l'autonomia di ogni meccanismo di influenza tra una variabile e un'altra. Ad esempio, Cartwright⁴⁵ è fortemente critica riguardo al requisito di modularità.
- L'intervento, come già affermato in precedenza, è una nozione causale. Come possiamo utilizzare una nozione causale per parlare di causalità?

Le due principali analisi riguardanti la nozione di intervento state svolte da Woodward⁴⁶ e da Pearl⁴⁷, ma la letteratura, come già detto, è piuttosto vasta⁴⁸. Pearl produce un'analisi della nozione di intervento che permette la distinzione di interventi e controfattuali, attraverso l'operatore $do(\cdot)$; Woodward, invece, non sviluppa una semantica così dettagliata riguardante gli interventi, ma, in compenso, si concentra sulla questione della circolarità in maniera assai più approfondita rispetto a Pearl e propone una interessante nozione di intervento caratterizzata dal riferimento ad una seconda variabile nella definizione; inoltre coglie, fondamentalmente, il fatto che l'obiettivo di una proposta manipolazionista non è quello di espungere nozioni causali, bensì quello di mostrare come la proposta non sia viziosamente circolare, non triviale e informativa. Verranno utilizzate, in quanto filosoficamente complementari, entrambe le analisi al fine di presentare un'immagine il più ricca e precisa possibile.

Questa sottosezione si occuperà innanzitutto di chiarificare la nozione di intervento. Questa è una nozione che ritengo necessario specificare con attenzione in quanto la modalità secondo la quale questo concetto è concepito e esplicato porta a risultati anche molto differenti per quanto riguarda ciò che possiamo fare con il nostro account. Se la nostra intenzione, come lo è parzialmente quella di questo scritto, è una chiarificazione non meramente *semantica* della questione della causalità e

⁴³ Pearl, *Causality: Models, Reasoning, and Inference* cit.

⁴⁴ Si ricorre ad una equazione in termini probabilistici in quanto cattura in maniera più naturale la condizione. In termini deterministici potrebbe essere reso nel seguente modo: $\because X = x \rightarrow Y = y \therefore do(X = x) \rightarrow Y = y$.

⁴⁵ N. Cartwright, «Against Modularity, the Causal Markov Condition, and Any Link between the Two: Comments on Hausman and Woodward», *The British Journal for the Philosophy of Science*, vol. 53, n. 3 (2002), pp. 411-453, <http://www.jstor.org/stable/3541889>; Cartwright, «Two Theorems on Invariance and Causality» cit.; N. Cartwright, *Hunting Causes and Using Them. Approaches in Philosophy and Economics*, Cambridge University Press, 2007.

⁴⁶ Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.

⁴⁷ Pearl, *Causality: Models, Reasoning, and Inference* cit.

⁴⁸ Cartwright, «Two Theorems on Invariance and Causality» cit.; Cartwright, *Hunting Causes and Using Them* cit.; Spirtes et al., *Causation, Prediction, and Search* cit.; J. Woodward, «What Is a Mechanism? A Counterfactual Account», *Philosophy of Science*, vol. 69, n. S3 (2002), S366-S377, DOI: 10.1086/341859; Woodward e Hitchcock, «Explanatory Generalizations, Part I: A Counterfactual Account» cit., solo per citarne alcuni.

di cosa significhi per X causare Y , bensì strutturale, questo è un compito che non va preso alla leggera. Nella fattispecie, avremo bisogno di una nozione di intervento che ci eviti problematiche in quanto a circolarità e che ci permetta di reggere affermazioni riguardanti la relazione causale tra una variabile ed un'altra in maniera sufficientemente solida.

Questa mia convinzione deriva innanzitutto dal fatto che se si presenta una teoria *manipolazionista* senza aver chiara e definita una nozione di intervento, ci si ritrova immediatamente nei pantani di problematiche collegate all'integrità del sistema causale sul quale si opera, all'influenza reciproca tra variabili e al significato di *impostare* o *imporre* un valore ad una variabile. Molte delle critiche alla letteratura sui modelli causali si basano su questo tipo di problematiche⁴⁹. Dato che le obiezioni che naturalmente sorgerebbero da una mancata specificazione della nozione di intervento ci porterebbero mano a mano a restringere il campo della nostra teoria attraverso la definizione e l'imposizione di alcune limitazioni e la necessità di accettare alcune premesse che ne permettano un utilizzo, per così dire, sereno, ci ritroveremmo con una teoria la cui base teorica è disorganizzata e confusionaria, e che necessiterebbe una rifondazione concettuale radicale; ritengo che sia necessario agire in maniera proattiva e fondare, innanzitutto, la teoria su una base concettuale solida.

La definizione della nozione di intervento ci permette di caratterizzare questa base concettuale in maniera chiara, di presentare delle premesse necessarie alla trattazione a cui ci avviamo e di elaborare una teoria della causalità che non presenti delle dubbie assunzioni metafisiche, perlomeno per quanto riguarda la nozione centrale in questo contesto, ovvero la manipolazione.

Chiarisco fin da ora che la formulazione della nozione di intervento a cui si punta non ha l'obiettivo di determinare che tipo di intervento sia *valido* o *utile* in un setting sperimentale in cui l'obiettivo è determinare l'effetto causale che qualcosa ha su qualcos'altro, bensì quello di fornire una definizione di principio che ci permetta di non cadere in una circolarità viziosa in sede di presentazione di una teoria causale. Di conseguenza non ho intenzione di affermare che *solo* interventi che rispettano le caratteristiche proposte siano informativi in sede sperimentale, cosa che, anzi, ritengo fondamentalmente scorretta⁵⁰. Questa mia impostazione è, in realtà, leggermente in controcorrente rispetto a qualche autore (penso in particolare a **Cartwright** che predilige, invece, nozioni utili in situazioni sperimentali⁵¹).

Come abbiamo visto nella **Sezione 1.1**, e in particolare nella **Definizione 1.1.5 (Dipendenza diretta)**, è necessario, al fine di operare con concetti causali in un contesto in cui si utilizzano i **SEMs**, riferirci ad una nozione di intervento: risulta necessaria, infatti, per definire la stessa nozione preliminare di dipendenza diretta, intervenire sulle variabili, modificandone il valore, oppure mantenendolo stabile nonostante le condizioni del sistema siano variate. Le domande a cui è necessario dare risposta sono quindi le seguenti: che cos'è un intervento? Che caratteristiche deve avere una manipolazione perché possa essere definita un intervento valido? Le due domande sono strettamente connesse. Verranno analizzate, come anticipato, le posizioni di **Pearl** e **Woodward**.

⁴⁹ Si veda: **Cartwright**, *Hunting Causes and Using Them* cit. per un'estesa trattazione della problematica.

⁵⁰ Questo è un punto fermo della discussione di **Woodward**, che condivido pienamente. Ad esempio, generalmente, otteniamo informazioni di tipo causale da situazioni che sono, oltre che non sperimentali, non interventi ideali nel senso in cui verranno presentati in questo scritto. Esempi sono osservazioni di fenomeni in natura o grossolani interventi su piccola scala (si immagini il premere un interruttore per vedere se causa l'accensione di una lampadina o di un'altra).

⁵¹ N. Cartwright, *Nature's Capacities and Their Measurement*, Oxford University Press, 1994, DOI: [10.1093/0198235070.001.0001](https://doi.org/10.1093/0198235070.001.0001); **Cartwright**, *Hunting Causes and Using Them* cit.

1.2.2 La proposta di Pearl

Il problema principale di cui si occupa *Causality: Models, Reasoning, and Inference* è, fondamentalmente, quello di proporre una cornice metodologica che permetta di calcolare, numericamente, diverse *proprietà causali*, che normalmente abbiamo grossi problemi a calcolare (es. l'effetto netto di X su Y , effetto causale medio, effetto del trattamento sul trattato⁵², ecc.), e come farlo a partire anche da soli dati sperimentali⁵³. Propone, inoltre, tecniche di programmazione lineare e vari algoritmi per identificare strutture causali e produrre stime di misure che di fatto non possiamo calcolare.

Escluso il Capitolo 7 di *Causality: Models, Reasoning, and Inference*, che si occupa di controfattuali e si inserisce in una linea di riflessione differente, prettamente filosofica, e in cui Pearl, tra l'altro, dimostra che dai suoi assiomi seguono i controfattuali non-*backtracking* di Lewis, l'intento di Pearl è quello di dare allo scienziato una cornice che gli permetta di fare chiarezza su una molteplicità di misure statistiche che nascondono (o che circumnavigano) nozioni causali, prendere coscienza del fatto che alcune problematiche richiedono un punto di vista extra-statistico per essere trattate (ad esempio, il paradosso di Simpson⁵⁴), demistificare la natura dei controfattuali e, infine, calcolare o stimare effetti causali.

Se questa è la problematica di cui ci si deve occupare, allora non abbiamo grossi problemi per quanto concerne la nozione di intervento. La nostra preoccupazione è, infatti, calcolare (o stimare), le variazioni nella distribuzione di probabilità sul modello, non quella di fondare un criterio con il quale possiamo affermare che X causa Y .

La formulazione della nozione di intervento da parte di Pearl è, quindi, estremamente elementare, una caratteristica che la rende particolarmente attraente come punto di partenza per una riflessione più approfondita, ma ad un prezzo: l'assunzione di una nozione di dipendenza causale come primitiva (simile alla nozione di dipendenza diretta introdotta precedentemente seguendo Halpern, che, infatti, è uno stretto collaboratore di Pearl e ne condivide alcune posizioni di fondo). Pearl concepisce gli interventi su variabili come interventi *atomici*: chirurgiche variazioni del sistema che impongono un valore x_i ad una variabile, senza intaccare la forma funzionale del sistema e generando un semplice sottomodello M_X in cui la funzione F_X è sostituita dalla costante x_i . L'intervento è quindi il *sollevare* la variabile dal meccanismo che normalmente ne determina il valore, ed imporre un valore equivalente ad una costante, il tutto senza modificare il resto del sistema.

Questa concezione di intervento impone la necessità di modificare il *network causale* di riferimento: la variabile soggetta ad un intervento, infatti, non subirà più l'influenza delle variabili genitori e, di conseguenza, questa indipendenza acquisita attraverso l'imposizione di un valore andrà rappresentata adeguatamente nel grafo. La nozione di intervento di Pearl equivale, in termini grafici, all'eliminazione di ogni freccia entrante nella variabile. Un esempio è riportato nella Figura 1.7, dove è stato graficamente rappresentato un intervento sulla variabile C .

Pearl, e con lui molti altri, concepisce le equazioni strutturali come rappresentanti dei meccanismi che controllano il comportamento di singole variabili. Di conseguenza, un intervento su di una

⁵² L'effetto di un farmaco sul paziente che l'ha assunto, ad esempio.

⁵³ Alcuni di questi risultati, per essere validi, richiedono l'assunzione di popolazioni di studio infinite, ma questo non è un problema che tocca il nostro interesse.

⁵⁴ J. Pearl, «Comment: Understanding Simpson's Paradox», *The American Statistician*, vol. 68, n. 1 (2014), pp. 8-13, DOI: [10.1080/00031305.2014.876829](https://doi.org/10.1080/00031305.2014.876829); M. A. Hernán et al., «The Simpson's paradox unraveled», *International Journal of Epidemiology*, vol. 40, n. 3 (2011), pp. 780-785, DOI: [10.1093/ije/dy041](https://doi.org/10.1093/ije/dy041).

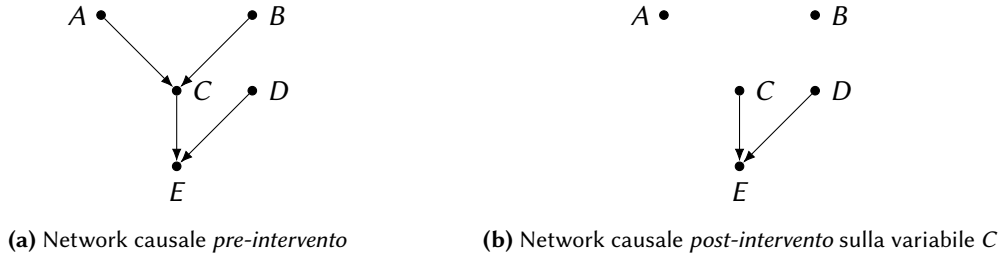


Figura 1.7: Rappresentazione grafica di un intervento secondo **Pearl**

variabile diventa una perturbazione di quel singolo meccanismo che determina il valore di una variabile, senza andare ad intaccare il resto dei meccanismi nel sistema.

Pearl tratta anche interventi che, invece di imporre un determinato valore ad una variabile, vanno ad influire sul suo comportamento funzionale, ovvero interventi che ne modificano la funzione F_X senza necessariamente fissarla ad una costante.

In questo contesto, la forma funzionale che descrive il comportamento della variabile diventa la seguente:

$$x_i = I(pa_i, f_i, u_i)$$

dove I è una funzione a tre posti tale che:

$$I = I(a, b, c) \quad \text{ogni volta che } b = f_i$$

Ora è sufficiente aggiungere una variabile genitore di X_i che rappresenti una forza che modifica la forma della funzione f_i . Questo permette una trattazione più ricca, in quanto in grado di catturare modificazioni differenti e di analizzare anche modificazioni spontanee del meccanismo.

Come già detto nel capitolo precedente, una volta compiuto questo intervento si genera un sottomodello M_X in cui la variabile X è sostituita dal valore (o dalla funzione modificata) ogni volta che appare al lato destro di una equazione. Questo genera il sottomodello utilizzato per l'analisi di controfattuali e rapporti causali. L'operazione equivalente in un *network causale* è rappresentata dall'eliminazione di ogni freccia entrante nella variabile X ⁵⁵. Questo genera un grafico in cui la variabile sulla quale abbiamo agito rimane fondamentalmente *orfana*, ovvero non viene più determinata dai propri parenti ma è determinata in maniera esterna.

La prima cosa da dire riguardo a questo approccio è che è estremamente rarefatto, nel senso che la sua principale preoccupazione è quella di gettare delle basi teoriche che permettano una trattazione fondamentalmente teorica e formale della problematica dell'intervento e che prescindano da questioni di tipo empirico (ad esempio di fattibilità)⁵⁶. Questo rende la trattazione estremamente generale ed è, a mio modo di vedere, il punto di forza dell'approccio proposto da **Pearl**, che è volto a fornirci una strumentazione teorica di fondo con cui poter iniziare a discutere determinate problematiche (di causalità nel nostro caso) in maniera chiara e non ambigua.

Inoltre, nella proposta di **Pearl**, un intervento non è soggetto a restrizioni di alcun tipo, se non quelle riguardanti i valori a cui setta la variabile su cui si interviene, che devono necessariamente appartenere al *range* di valori che la variabile può assumere. Questa generalità è stata fatta oggetto di

⁵⁵ Nel caso di una variazione funzionale come quella esposta sopra questo non è necessario.

⁵⁶ Questa impostazione è testimoniata dallo sviluppo di una logica formale per l'inferenza causale da parte di **Pearl**: il cosiddetto *do calculus*.

critiche da Nancy Cartwright, che considera la cornice metodologica di Pearl, ed in particolare la sua nozione di intervento, praticamente inutilizzabile. L'argomento principe che Cartwright propone contro Pearl è che non esistono, nella realtà, interventi veramente atomici. Riportiamo un esempio discusso dalla stessa Cartwright.

Immaginiamo un ingegnere che lavora su di un carburatore. Una molteplicità di meccanismi che valutano la performance del carburatore dipendono da una sua singola componente, ovvero la sua geometria tridimensionale. In questa condizione come è possibile produrre interventi che siano atomici, quando un intervento sulla geometria del carburatore ne modifica una varietà estremamente ampia di meccanismi, si chiede Cartwright.

Risulta innegabile che l'obiezione sembri piuttosto problematica, dato che effettivamente ci risulta difficile immaginare situazioni in cui possiamo compiere interventi così atomici tali da non perturbare lo stato di altri meccanismi che fanno parte del sistema e che influenzano la forma funzionale di altre variabili. Resta il fatto, comunque, che quello che implica l'obiezione è solo che non si possa modificare la geometria senza influire su altri valori. Pearl fa notare, ad esempio, che nulla ci impedisce di installare una valvola invece che modificare la geometria del carburatore. Un articolo di Hernán e Taubman⁵⁷ presenta una problematica simile riguardo al fatto che la modalità di intervento in questioni di *policy* spesso non può essere chirurgica (viene, nell'articolo, portato come esempio l'obesità), ma deve venire effettuata attraverso dei *proxy* che, di fatto, inquinano l'intervento.

A mio parere la problematica fondamentale di questa obiezione di Cartwright è che si pone su di un piano completamente diverso rispetto a quello discusso da Pearl: mentre Cartwright è estremamente attenta a questioni di carattere pragmatico, e il suo interesse è chiaramente orientato verso la pratica scientifica, Pearl non condivide lo stesso punto di vista. Nella fattispecie, come ho evidenziato precedentemente, il problema di Pearl è quello di fornire una *framework* il più generale e flessibile possibile, fornire una base comune di discussione, settare una *baseline* per un dibattito che altrimenti rischia di andare a perdersi su dettagli di poca importanza invece di analizzare le questioni fondamentali.

Il punto principale è che Pearl non ci vuole spiegare *come si fa un buon intervento nella pratica scientifica*, bensì vuole mostrarci *come dovremmo pensare un intervento*, che tipo di design questo deve avere, quali devono essere le nostre aspirazioni e, soprattutto, darci uno strumento per produrre delle previsioni senza dover necessariamente intervenire nel mondo reale. L'obiettivo di Pearl è:

[...] definire e fornire condizioni di identificazione per la nozione manipolazione-neutra di "effetto dell'obesità sulla morbosità" [...] la framework dei *Structural Causal Models (SCMs)* fornisce *definizioni formali e condizioni di identificazione* per ciascuno dei molteplici effetti di manipolazioni specifiche.⁵⁸

Insomma, la differenza fondamentale è che Pearl non ha come obiettivo quello di indicare modalità di approccio alla pratica scientifica; piuttosto si preoccupa di come pensare e parlare di un intervento quando siamo nella fase concettuale della perturbazione di un sistema, non quando si tratta intervenire di fatto. Mentre Cartwright guarda alla proposta di Pearl come ad uno strumento che ha scarsa ricaduta pragmatica, anche per il prerequisito di *modularità* che Pearl impone ai sistemi

⁵⁷ M. A. Hernán e S. L. Taubman, «Does obesity shorten life? The importance of well-defined interventions to answer causal questions», *Int J Obes*, vol. 32, n. S3 (2008), S8-S14, DOI: [10.1038/ijo.2008.82](https://doi.org/10.1038/ijo.2008.82).

⁵⁸ J. Pearl, «Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes», *Journal of Causal Inference*, vol. 6, n. 2 (2018), DOI: [10.1515/jci-2018-2001](https://doi.org/10.1515/jci-2018-2001): enfasi mia.

sotto studio, prerequisito che la **Cartwright** aborre e che categoricamente rifiuta, come sostenuto in diverse sedi⁵⁹, io credo che la maniera corretta di concettualizzare il contributo dell'autore sia quella di porre dell'ordine in quello che era⁶⁰ un dibattito estremamente poco multidisciplinare⁶¹, e porre in luce la fundamentalità della questione causale rispetto a quella statistica, e mostrare che abbiamo uno strumento chiaro e potente con cui discutere di una tematica che, fino a 20 anni fa, era guardata con estremo sospetto nella maggior parte degli ambienti accademici. **Pearl** vuole proporci uno strumento con cui possiamo fare inferenza causale in modo chiaro e affidabile, e attraverso il quale possiamo mettere da parte questioni riguardanti la modalità di queste inferenze, per concentrarci sulle questioni veramente importanti, ovvero quelle di modellazione: quali variabili dovrebbero essere esogene? Quali variabili dovrei utilizzare? È quella proposta una struttura causale sufficientemente dettagliata rispetto al problema che stiamo affrontando?

1.2.3 La proposta di Woodward

Differente, come vedremo, il focus di **Woodward**, che nel suo *Making Things Happen: A Theory of Causal Explanation*, propone la seguente definizione di intervento:

(IN) Il far assumere un qualche valore z_i alla variabile I è un intervento su di X rispetto a Y se e solo se I è una variabile di intervento per X rispetto ad Y e $I = z_i$ è una *causa effettiva* del valore assunto da X .⁶²

con la seguente definizione di variabile di intervento (I):

(IV)

I(1). I causa X .

I(2). I fa da interruttore per tutte le variabili che causano X . Ovvero, alcuni valori di I sono tali per cui, quando I realizza questi valori, X smette di dipendere da altre variabili che causano X e, invece, dipende solo dal valore assunto da I .

I(3). Qualsiasi percorso direzionato da I a Y passa attraverso X . Ovvero I non causa direttamente Y e non è una causa di alcuna causa di Y distinta da X escluse, ovviamente, le cause di Y , se ce ne sono, che fanno parte della catena I - X - Y ovvero, escluse (a) cause di Y che non sono effetti di X (es. variabili che sono comprese tra X e Y) e (b) qualsiasi causa di Y inclusa tra I e X e che non ha effetto su Y indipendentemente da X .

I(4). I è (statisticamente) indipendente da qualsiasi variabile Z che causa Y e che è su di un percorso direzionato diretto che non passa attraverso X .⁶³

⁵⁹ Cartwright, «Against Modularity, the Causal Markov Condition, and Any Link between the Two: Comments on Hausman and Woodward» cit.; Cartwright, *Hunting Causes and Using Them* cit.

⁶⁰ L'enorme risultato di **Pearl** è esattamente questo: aver riportato in auge e reso una questione di dibattito trasversale tra discipline i modelli causali.

⁶¹ *Causality: Models, Reasoning, and Inference* ha lunghi capitoli in cui discute della sorta di bivio in cui si sono dipartite le scienze sociali e le scienze statistico/economiche e quanto si siano addentrate nel loro percorso tanto da rendersi quasi mutualmente incomprensibili.

⁶² Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.: p. 98.

⁶³ *Ivi*: p. 98.

Ci sono, innanzitutto, due cose interessanti che vanno annotate riguardo queste definizioni. La prima è che, nella definizione di intervento, **Woodward** fa specificamente riferimento alla nozione di causa (ad esempio nella **Cond. I(1)**). La seconda è che si riferisce ad una nozione di *causa effettiva*, che per il momento possiamo caratterizzeremo nel seguente modo:

DEFINIZIONE 1.2.1 (CAUSA EFFETTIVA PRELIMINARE)

$X = x$ è una causa effettiva di $Y = y$ se:

$ACP(1)$. $(M, \vec{u}) \models (X = x, Y = y)$.

$ACP(2)$. Vi è almeno un insieme di variabili \vec{R} , con $\{R_1, \dots, R_n\} \in \vec{R}$ e $R_1 = X, R_n = Y$, tale che R_{k+1} dipende direttamente da R_k .

con la formula $(M, \vec{u}) \models (X = x)$ che denota la verità di $X = x$ nel modello causale M con setting \vec{u} ; e dove la **Cond. ACP(2)** è equivalente a richiedere che vi sia un cammino direzionato tra X e Y nel *network causale* che rappresenta il modello causale. Queste condizioni sono chiaramente un punto problematico, dato che, nuovamente, ci ritroviamo con una nozione di causa dove noi vorremmo definirla. Vedremo in seguito come **Woodward** risolve questo problema.

La nozione di intervento presentata da **Woodward** richiede l'introduzione di una variabile di intervento, che impone un determinato valore alla variabile su cui vogliamo intervenire. Questo è un punto di differenza radicale rispetto a **Pearl** e deriva da un filone di riflessione riguardo ai modelli causali differente, del quale possiamo tracciare le origini nella riflessioni di **Hausman**⁶⁴ e **Hausman e Woodward**⁶⁵.

Fondamentalmente **Pearl** pensa ai modelli causali come ad una famiglia con molti membri, di cui ogni membro è un modello causale differente in cui o una variabile è settata ad un valore, escludendo l'influenza delle sue variabili genitore, oppure è una combinazione lineare di diversi modelli in cui sono settate diverse variabili. **Hausman e Woodward** ritengono, invece, più opportuno far svolgere questa funzione ad una variabile strumentale che setta la variabile in questione.

Altra differenza radicale è quella della definizione del concetto di variabile di intervento (e di intervento) in maniera relativa ad una terza variabile. Questo è una scelta metodologica a mio parere estremamente interessante, in quanto segnala un'attenzione a questioni di tipo fondazionale che non troviamo in **Pearl**. La motivazione che giustifica questa mossa da parte di **Woodward** è legata al fatto che, quando vogliamo determinare effetti causali, è estremamente rilevante tra quali variabili stiamo cercando il legame causale.

Il punto focale della trattazione di **Woodward** è quello di proporre una nozione di causalità che non sia viziosamente circolare. Questo obiettivo viene raggiunto attraverso l'utilizzo di informazioni causali riguardanti il *contorno* del modello rispetto alle due variabili tra cui vogliamo determinare la presenza, assenza o grado di relazione causale. **Woodward** ritiene che affermare che tra X e Y vi sia una relazione causale attraverso informazioni causali riguardanti la variabile di intervento I e altre variabili Z non sia circolare, in quanto nel compiere questa affermazione non facciamo uso di alcuna informazione che riguarda la relazione tra X e Y . Questo è anche il motivo per cui non sposa la necessità di considerare interventi validi solo quelli che non *perturbano* il meccanismo di azione tra X e Y , come invece fa **Pearl**. Questo perché richiedere l'invarianza del meccanismo tra

⁶⁴ D. M. Hausman, *Causal Asymmetries*, Cambridge University Press, 1998, DOI: [10.1017/cb09780511663710](https://doi.org/10.1017/cb09780511663710).

⁶⁵ D. M. Hausman e J. Woodward, «Independence, Invariance and the Causal Markov Condition», *The British Journal for the Philosophy of Science*, vol. 50, n. 4 (1999), pp. 521-583, DOI: [10.1093/bjps/50.4.521](https://doi.org/10.1093/bjps/50.4.521).

X e Y richiede informazioni riguardo a questo meccanismo, informazioni necessariamente causali, che quindi ci riportano al problema della circolarità della definizione⁶⁶.

Woodward evita completamente questa problematica ritenendo validi anche interventi che, di fatto, intaccano il meccanismo tra le due variabili. La vera necessità di un intervento è, secondo **Woodward**, quella rappresentata dalle **Cond. I(3)** e **Cond. I(4)**, legate alle relazioni che la variabile di intervento intrattiene con il resto del sistema⁶⁷.

Questa idea non è controintuitiva come sembra: non è irragionevole, infatti, sostenere che un intervento che danneggia il meccanismo sia una causa dell'effetto che questo danneggiamento produce. Diverse volte **Woodward**, quando si tratta di discutere tematiche legate all'integrità dei meccanismi, fa riferimento alla molla come esempio paradigmatico: se la variabile è la lunghezza a cui questa molla è tesa (o compressa), il fatto di applicare una tensione troppo elevata e tale da romperla è un evento rilevante per una trattazione della causalità, nonostante vi sia una modifica radicale del meccanismo. Se ci si riflette, facciamo spesso uso di questo tipo di informazioni causali nella nostra vita quotidiana, di conseguenza escludere radicalmente casi come questo (cosa che fanno la maggior parte dei teorici, compreso **Pearl**), non sembra così giustificata.

Un esempio in cui questa *flessibilità* della nozione di intervento potrebbe essere utile è il seguente: si immagini che un montacarichi venga caricato da un operaio con un peso superiore a quello che viene riportato nelle indicazioni di sicurezza. Al contempo, il montacarichi non è stato correttamente mantenuto dal responsabile della sicurezza in loco, portando ad un danno marginale ad un meccanismo fondamentale per l'operazione in sicurezza. Il montacarichi, sollevando il carico, si rompe. Ora ciò che ci interessa per valutare la causalità dell'evento $E = \text{rottura del montacarichi}$, è esattamente un intervento che non tutela l'integrità del meccanismo causale in opera, bensì in condizioni in cui la rottura del meccanismo è in gioco.

Le due linee di pensiero portano a due differenti problematiche, in particolare si fa notare come, nel caso di **Pearl**, la problematica sia quella della possibilità della costruzione di modelli alternativi e delle loro combinazioni lineari senza intaccare il sistema; nel caso dell'approccio di **Hausman e Woodward** la problematica è, come tra l'altro discusso da **Cartwright**⁶⁸, quella di avere una garanzia della non interferenza delle diverse variabili-intervento e se il loro effetto sulla variabile di effetto si espliciti esclusivamente attraverso la variabile causa o se invece non componga con altre variabili, andando ad inficiare di fatto l'*identificazione* e la garanzia della *modularità*.

Entrando nello specifico della forma di intervento proposta da **Woodward**: esso viene rappresentato attraverso un modello causale aumentato M_I , formato dal modello causale M , a cui viene aggiunto all'insieme \mathcal{V}_M una variabile I , e viene modificata la funzione F_{X_M} nel seguente modo:

$$F_{X_{M_I}} = \begin{cases} F_{X_M}(\mathcal{V}_M \setminus X) & \text{se } I \neq i \\ I & \text{se } I = i \end{cases}$$

66 Si fa notare che **Fenton-Glynn** critica tutto questo apparato, sostenendo che **Woodward** definisce il concetto di intervento attraverso la nozione di cammino, a sua volta definita in termini di controfattuali basati sulla nozione di intervento. **Fenton-Glynn** propone l'innesto di una semantica dei controfattuali di stampo lewisiano, rigettata da **Woodward** sulla base di una presunta erroneità e vaghezza, all'interno dell'analisi della causalità fondata su modelli causali, al fine di evitare questa problematica. Si veda: L. Fenton-Glynn, «Of Miracles and Interventions», *Erkenntnis*, vol. 78, n. S1 (2013), pp. 43-64, DOI: [10.1007/s10670-013-9436-5](https://doi.org/10.1007/s10670-013-9436-5).

67 Si noti quanto sia pregevole questa posizione: quando si interviene, non è importante evitare di turbare il meccanismo, l'importante è che vi siano le variazioni tra le variabili ricercate e che non siamo noi a causarle, bensì la variabile su cui interveniamo.

68 **Cartwright**, *Hunting Causes and Using Them* cit.

con $i \in \{i_1, \dots, i_n\}$. Una rappresentazione grafica di questa modalità di intervento viene data in **Figura 1.8**.

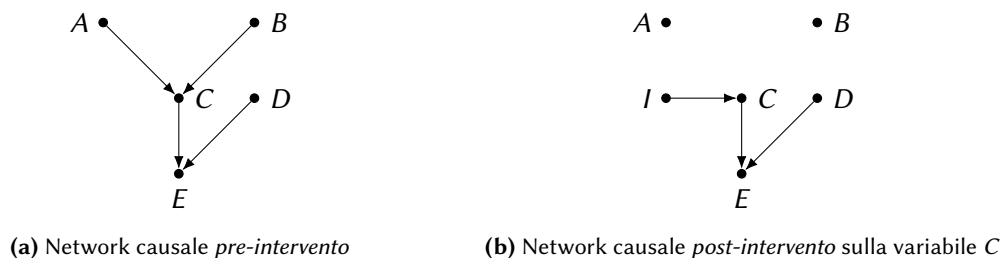


Figura 1.8: Rappresentazione grafica di un intervento secondo Woodward

La variabilità di F_{X_M} in base al valore di I è dovuta alla natura di interruttore, esplicitata alla **Cond. I(2)**, della variabile I . Un esempio di variabile interruttore è il pulsante di accensione e spegnimento di uno stereo: se questo è in posizione di accensione abbiamo un'abbondanza di potenziometri per regolare volume, bassi, alti ecc. in grado di agire sul sistema e modificare l'uscita sonora (anche portandola a zero, semplicemente ruotando al minimo il potenziometro del volume), ma questa loro capacità va perduta nel momento in cui il pulsante è in posizione di spegnimento, che comporta il non passaggio di corrente all'interno dell'apparecchio, impedendo ai potenziometri di realizzare il loro potere causale sull'uscita sonora dell'apparecchio, che rimane, in questo caso, nulla⁶⁹.

Vediamo che questa modalità di riflessione sulla causalità e sugli interventi rappresenta un parallelo teorico del prototipico esperimento galileiano, come Cartwright giustamente nota⁷⁰. L'idea di Woodward è quindi quella di proporre un criterio per gli interventi che rappresenti il modo in cui, idealmente, vorremmo condurre un esperimento per testare una determinata ipotesi causale. Il lavoro teorico è, quindi, focalizzato nel dare delle rigide condizioni che ci permettano di affermare con certezza che se queste vengono rispettate (i) le relazioni su cui stiamo indagando sono genuinamente causali; (ii) le inferenze causali che facciamo sono garantite.

Ed infatti, lo stesso Woodward afferma:

Una teoria manipolazionista plausibile [...] suggerirà una modo specifico di pensare alle inferenze [causali]: dovremmo pensare a queste inferenze come tentativi di determinare (sulla base di altri tipi di evidenza) quale sarebbe il risultato di un ipotetico esperimento progettato in maniera adeguata o di una manipolazione, senza compiere questo esperimento veramente.⁷¹

L'introduzione di complicazioni di questo tipo (es. *variabili interruttore*, funzioni definite a casi ecc.) risultano necessarie per catturare la doppia natura dell'intervento.

1.3 UNA SEMANTICA DEI CONTROFATTUALI

Stabilite le principali caratteristiche dei modelli causali nella **Sezione 1.1**, ci occupiamo ora di trattare la questione della valutazione dei controfattuali attraverso questi modelli. Le opinioni a riguardo alla

⁶⁹ Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.; Woodward e Hitchcock, «Explanatory Generalizations, Part I: A Counterfactual Account» cit.

⁷⁰ Cartwright, *Hunting Causes and Using Them* cit.

⁷¹ Woodward, *Making Things Happen: A Theory of Causal Explanation* cit.

capacità di questi approcci, intesa come ampiezza di controfattuali valutabili e di inferenze garantite. C'è chi, come Briggs⁷², ritiene che gli approcci a modelli causali siano radicalmente meno potenti di approcci alla Lewis⁷³: la semantica sviluppata sinora per questo approccio, infatti, non è in grado di valutare controfattuali con antecedenti e conseguenti di complessità elevata, e che solo una piccola classe di controfattuali siano valutabili in questa cornice, peraltro non senza fatica. Briggs, inoltre, nota che i due approcci (quello di Lewis e quello a modelli causali), nemmeno concordano riguardo alla valutazione dei controfattuali. Quella di Briggs viene considerata la semantica più completa.

Di avviso diametralmente opposto è Hiddleston⁷⁴, che, invece, ritiene gli approcci a modelli causali, e, nella fattispecie, la sua formulazione di una semantica causale per i controfattuali, decisamente superiore rispetto a quella proposta da Lewis, affermando che la sua trattazione è in grado di valutare le condizioni di verità dei controfattuali in maniera corretta anche in casi che la teoria di Lewis non è in grado di analizzare e, inoltre, è in grado di trattare anche di una classe di controfattuali piuttosto problematica per Lewis: i controfattuali cosiddetti *backtracking*. La proposta di Hiddleston verrà presentata nella sottosezione successiva, in quanto risulta piuttosto interessante.

1.3.1 Una curiosa creatura: una teoria causale dei controfattuali

Come espresso nella Sezione 1.1, gli approcci a modelli causali si fondano su delle nozioni causali prese per primitive. Un interessante lavoro di Hiddleston⁷⁵, ispirato parzialmente da Cheng⁷⁶ parte proprio da questo fatto e cerca di elaborare una teoria dei controfattuali *rovesciata*, ovvero: invece di tentare di fondare il discorso riguardo alla causalità sui controfattuali perché, dato il fatto che assumiamo un concetto implicito di causalità come primitivo quando lavoriamo con i modelli causali, non tentiamo di fondare una semantica dei controfattuali proprio sui modelli causali e la loro assunzione riguardo alla causalità? L'approccio, devo ammettere, è, personalmente, *intrigante*. Ciò che Hiddleston sostiene è che i tentativi di analisi riduttiva della causalità in termini di controfattuali siano informati da un progetto destinato al fallimento, in quanto pone le proprie priorità *al contrario*: il nostro progetto non dovrebbe essere fondare la causalità sui controfattuali, bensì fondare questi ultimi sulla prima.

Effettivamente non risulta così controintuitivo ritenere che in realtà le valutazioni controfattuali siano basate su assunzioni mentali di tipo causale: quando valutiamo un controfattuale tendenzialmente ricostruiamo la storia causale all'indietro fino al punto dove si verifica la rottura con la storia causale nel mondo attuale, effettuiamo la rottura e, dopodiché, evolviamo il sistema fino alla momento in cui il conseguente si verifica. Risulta quindi un'idea interessante quella di ribaltare la prospettiva e provare ad analizzare i controfattuali in termini di causalità.

L'approccio di Hiddleston può essere riassunto nel seguente modo: posti di fronte ad un controfattuale, è necessario ricostruire un *buon* (in un senso che andrà specificato) modello della situazione, inserire la minima *rottura* possibile al fine di rendere l'antecedente vero all'interno del modello causale, mantenendo, in linea con il requisito di minimalità, la più larga parte possibile del modello intatta, e infine evolvere le equazioni secondo il nuovo stato del sistema, ottenendo il valore di verità

⁷² Briggs, «Interventionist counterfactuals» cit.

⁷³ Lewis, *Counterfactuals* cit.

⁷⁴ Hiddleston, «A Causal Theory of Counterfactuals» cit.

⁷⁵ *Ivi*.

⁷⁶ P. W. Cheng, «From covariation to causation: A causal power theory.» *Psychological Review*, vol. 104, n. 2 (1997), pp. 367-405, DOI: [10.1037/0033-295X.104.2.367](https://doi.org/10.1037/0033-295X.104.2.367).

del controfattuale che volevamo valutare. Una interessante conseguenza di questo approccio è la capacità di valutare alcuni controfattuali *backtracking*.

I controfattuali *backtracking* sono una classe di controfattuali che contiene elementi piuttosto particolari. Si tratta, ad essere precisi, più che di una tipologia di controfattuali, di un'interpretazione di questi, come vedremo. Prima della discussione analitica, un esempio classico per chiarificare:

ESEMPIO 1.3.1 (Tuffo nel vuoto)

Vedete il vostro amico Smith sul tetto di un edificio alto decine di piani, vicino al cornicione, pronto a saltare. Tra lui e il marciapiede c'è solo il vuoto. Sul momento vi assale il panico e temete per la vita del vostro caro amico. Ma d'improvviso vi riprendete e vi rendete conto che Smith è razionale, non ha tendenze suicide e sa che saltare in quella situazione lo ucciderebbe sicuramente. Di conseguenza vi tranquillizzate. Smith, di lì a breve, si allontana dal cornicione e scende. A questo punto voi dite:

- (1) “Se Smith fosse saltato sarebbe sicuramente morto”.

Questo controfattuale è facilmente classificato come vero e, anzi, è il motivo per cui vi sentireste sollevati dal fatto che Smith non è effettivamente saltato⁷⁷.

Beth, una vostra amica che era lì con voi, quando vi sente affermare (1), dice: “Smith è un tipo razionale e certamente non ha ideazioni suicide. Sapeva bene che se fosse saltato senza nulla tra lui e il marciapiede sarebbe morto. Quindi, se fosse saltato, sicuramente ci sarebbe stata qualcosa per frenare la caduta, al fine di renderla innocua. Di conseguenza

- (2) se Smith fosse saltato, non sarebbe morto”.

Pur, intuitivamente, d'accordo con il controfattuale (1), il controfattuale (2) ci pare comunque una buona obiezione al nostro ragionamento.

Si tratta, in generale, di una classe di controfattuali piuttosto particolari. Nella fattispecie, Khoo⁷⁸, autore di quello che forse è il più esteso e dettagliato lavoro a riguardo, sostiene che questo tipo di controfattuali sorge da due fattori concomitanti:

1. Un'interpretazione *retrograda* rispetto alla classica interpretazione dei controfattuali.
2. L'affermazione preliminare di un controfattuale rovesciato (dall'effetto alla causa), che rende evidente l'interpretazione *retrograda*

Nel nostro caso, il Punto 2 è evidente: l'argomentazione preliminare di Beth cerca di trovare delle condizioni nelle quali il conseguente (il salto) è giustificato dall'antecedente (qualcosa che freni la caduta). Possiamo vedere che questo tipo di controfattuali è particolarmente curioso in quanto non sorge da un'interpretazione naturale di un controfattuale, bensì ha una sorta di carattere *giustificativo a posteriori*.

Torniamo ora alla proposta di Hiddleston. Introduciamo l'idea attraverso un esempio:

ESEMPIO 1.3.2

Jane riflette se scommettere \$1000 sull'uscita degli occhi del serpente (una coppia di uno) al prossimo lancio dei due dadi. Decide, alla fine, di non scommettere, i dadi vengono tirati ed escono

⁷⁷ Khoo (J. Khoo, «Backtracking Counterfactuals Revisited», *Mind*, vol. 126 (503 2016), pp. 841-910, DOI: [10.1093/mind/fzw005](https://doi.org/10.1093/mind/fzw005)) fa notare che una comparazione controfattuale è una proprietà fondamentale di molti verbi che designano un'emozione rispetto a fatti (es. *sorpresa*, *rimorso*, ecc.).

⁷⁸ Ivi.

effettivamente gli occhi di serpente. Jane ignora che, sotto al tavolo, ci sono dei magneti che, se attivati, impediscono che escano gli occhi di serpente. Questi magneti sono controllati da Mandrake, che ha la seguente politica: interferire con i grossi scommettitori il 5% delle volte. ■

Si consideri ora il seguente controfattuale:

- (3) Se Jane avesse scommesso sugli occhi di serpente (B) e Mandrake non avesse attivato i magneti ($\neg M$), allora l'esito del lancio sarebbe stato 2 (S).

il controfattuale

$$B \wedge \neg M \Box \rightarrow S$$

sembra essere vero⁷⁹, ma è, per molti account, problematico renderne conto.

Hiddleston, per quanto riguarda la sua teoria, utilizza modelli causali estremamente basilari, nel senso che non richiedono condizioni né necessarie né sufficienti riguardo alle dipendenze e indipendenze rappresentate all'interno del modello⁸⁰ indeterministico. Questo non comporta alcuna complicazione per la nostra discussione. Solo, si tenga presente, che ora le equazioni rappresentano relazioni indeterministiche, del seguente tipo:

$$\{X_1 = x_1, \dots, X_n = x_n\} \Rightarrow p(Y = y) = k$$

Le frecce tra variabili rappresentano, quindi, *influenze causali* positive o negative. Hiddleston, inoltre, caratterizza un modello causale con una assegnazione di valori tali per cui appartengono al range della variabile.

Caratterizza, a questo punto, la nozione di influenza positiva (o negativa) diretta nella maniera in cui noi abbiamo caratterizzato la nozione di dipendenza diretta (Definizione 1.1.5)⁸¹, ma con relazioni sempre su distribuzioni di probabilità. Definisce, ulteriormente, il concetto di *rottura causale* relativa ad un modello M_i rispetto al modello M : questa è una variazione nel valore di una variabile, nonostante i valori degli antenati (ovvero le variabili appartenenti ad una catena di influenze) siano i medesimi. Definisce, inoltre, la nozione di insieme di variabili intatte, come le variabili non soggette ad una *rottura causale* i cui antenati non sono soggetti ad una *rottura causale*⁸².

L'ultimo pezzo del puzzle è la definizione di un concetto strumentale alla valutazione dei controfattuali attraverso una sorta di criterio per misurare la distanza tra modelli, più o meno a svolgere il ruolo che nella teoria di Lewis svolge la funzione di somiglianza. Hiddleston definisce il seguente concetto:

79 Esempio ripreso da S. Barker, «Counterfactuals, probabilistic counterfactuals and causation», *Mind*, vol. 108, n. 431 (1999), pp. 427-469, DOI: [10.1093/mind/108.431.427](https://doi.org/10.1093/mind/108.431.427), citato da Hiddleston, «A Causal Theory of Counterfactuals» cit.

80 In letteratura, queste due condizioni si chiamano *Causal Markov Condition (CMC)* e *Faithfulness condition*, per una discussione più precisa si vedano: Spirtes et al., *Causation, Prediction, and Search* cit.; J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., 1988; Pearl, *Causality: Models, Reasoning, and Inference* cit.; C. Hitchcock, «Causal Models», in *The Stanford Encyclopedia of Philosophy*, a cura di E. N. Zalta, Summer 2020, Metaphysics Research Lab, Stanford University, 2020, <https://plato.stanford.edu/archives/sum2020/entries/causal-models/>.

81 Vi è una leggera differenza tra le due definizioni: quella di Hiddleston diventerà una definizione *contrastiva*, ma per i nostri fini è sufficiente la nozione preliminare non contrastiva.

82 Tutti i concetti qui descritti sono stati parafrasati e non sono state date definizioni rigorose per questioni di spazio e utilità. Si veda Hiddleston, «A Causal Theory of Counterfactuals» cit. e E. Hiddleston, «Causal Powers», *The British Journal for the Philosophy of Science*, vol. 56, n. 1 (2005), pp. 27-59, DOI: [10.1093/phisci/axi102](https://doi.org/10.1093/phisci/axi102) per definizioni precise ed una trattazione puntuale.

DEFINIZIONE 1.3.1 (MONDO Φ -MINIMALE)

Un modello M_i ed una sua rottura causale rispetto ad M sono Φ -minimali sse

(i) M_i è un Φ -modello

(ii) Tutte le variabili non discendenti di Φ nel modello sono massimamente intatte tra i Φ -modelli

(iii) La rottura causale è minimale tra i Φ -modelli

con Φ che rappresenta una proposizione atomica del tipo $X = x$, o una negazione o congiunzione di proposizioni atomiche ed un Φ -modello un modello nel quale Φ è vera. Le nozioni di minimale e massimale vanno interpretate come inclusioni di insiemi e non nel senso di cardinalità: un insieme di *rotture causali* è più minimale di un altro se è un sottoinsieme del secondo, e lo stesso vale, rovesciando la relazione, tra insiemi di variabili intatte.

Un controfattuale della forma $\Phi \square \rightarrow \Psi$ è vero in un modello M ed un contesto C sse Ψ è vero in tutti i Φ -modelli minimali M_i per i quali la rottura rispetto ad M è rilevante rispetto al contesto C .

L'aggiunta del contesto C sembra essere gratuita, ma in realtà permette di rendere conto delle diverse interpretazioni dei controfattuali, al fine di poter catturare i già nominati controfattuali *backtracking*.

Data la definizione di controfattuale vero in un modello, si tratta di dare una semantica ai controfattuali in generale. La risposta di Hiddleston è che un controfattuale è vero se è vero in un buon modello.

Al fine di valutare la bontà di un modello, Hiddleston propone una concezione di evento decisamente simile a quella proposta da Kim⁸³, seppur priva di tutte le finezze dell'account di Kim⁸⁴: un evento è un oggetto⁸⁵ che istanzia un *n-upla* di proprietà in un dato istante t o un intervallo Δt ⁸⁶. Questa è una teoria che si sposa estremamente bene con gli approcci a modelli causali in quanto le variabili all'interno del modello possono facilmente rappresentare la presenza/assenza di proprietà all'istante t da parte di un oggetto. Un buon modello M ha le seguenti caratteristiche: le proprietà rappresentate da M sono quelle possedute dagli oggetti in C , le leggi di M sono sufficientemente accurate e M è sufficientemente completo da rappresentare le relazioni causali tra gli eventi rappresentati in M . Un controfattuale sarà quindi vero in riferimento ad un insieme di eventi sse è vero in un buon modello rispetto alla serie di eventi.

Riprendiamo l'Esempio Tuffo nel vuoto, rappresentato in Figura 1.9, dove R è la variabile che rappresenta lo stato razionale o meno di Smith, P rappresenta la presenza di una protezione tra lui

83 J. Kim, «Causation, Nomic Subsumption, and the Concept of Event», *The Journal of Philosophy*, vol. 70, n. 8 (1973), p. 217, DOI: [10.2307/2025096](https://doi.org/10.2307/2025096); J. Kim, «Events as Property Exemplifications», in *Action Theory*, a cura di M. Brand e D. Walton, Springer Netherlands, 1976, pp. 159-177, DOI: [10.1007/978-94-010-9074-2_9](https://doi.org/10.1007/978-94-010-9074-2_9); J. Kim, «Events: Their Metaphysics and Semantics», *Philosophy and Phenomenological Research*, vol. 51, n. 3 (1991), p. 641, DOI: [10.2307/2107887](https://doi.org/10.2307/2107887).

84 Non sembra, ad esempio, che Hiddleston sposi una visione simile a quella di Kim per quanto riguarda la molteplicità di eventi non distinti es. Il mio starnutire, il mio starnutire alle ore 10.00 e il mio starnutire alle ore 10.00 appena uscito di casa sono, per Kim, eventi differenti seppur non distinti. Non così per Hiddleston.

85 Kim usa la più vaga nozione di sostanza, specificando però che non intende fornire alcuna precisa definizione filosofica e che userà il termine sostanza per indicare 'cose tipo tavoli, sedie, atomi, creature viventi, pezzi di cose come acqua e bronzo e simili'.

86 Non mi è chiaro se Hiddleston sposi un account degli eventi fragile a tal punto da rendere eventi avvenuti in istanti di tempo t differenti due eventi distinti, ma al fine della discussione questo è irrilevante e tratterò la concezione come non fragile entro standard ragionevoli.

e il marciapiede, S rappresenta il salto e V lo stato di salute del potenziale saltatore. Lo stato del modello è il seguente:

$$R = 1$$

$$P = 0$$

$$S = 0$$

$$V = 0$$

quello che guida le differenti interpretazioni è, secondo Hiddleston, il fatto che il contesto della

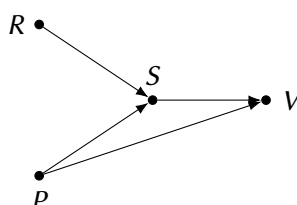


Figura 1.9: Grafo rappresentante il modello causale relativo all'Esempio Tuffo nel vuoto

prima interpretazione seleziona $R = 0$ come *rottura causale* minimale, mentre l'interpretazione *backtracking* seleziona $P = 1$. Questo tipo di interpretazione mi sembra curiosa, visto quanto discusso nella [Sezione 1.2.1](#): la variazione per giungere allo stato di salto viene effettuata prima della variabile che rappresenta il salto. Ovvero, invece di agire direttamente sulla variabile che rappresenta l'effettuazione o meno del salto, andiamo a modificare le cause del salto (la razionalità e le protezioni).

1.3.2 Problemi della teoria

La discussione dell'esempio riportato poco sopra non può essere corretta, per una serie di motivi. Il primo è che sembra scorretto modificare una variabile causalmente precedente rispetto all'antecedente del controfattuale, e questa è una prassi che la letteratura ha stabilito è bene evitare⁸⁷. Questa posizione viene anche mantenuta da ogni teorico dei modelli causali che ho potuto reperire in letteratura, salvo chi, come [Woodward](#), preferisce trattare gli interventi come modifiche su di una variabile di intervento che poi influenza la variabile che vogliamo modificare⁸⁸. Ma anche in questo caso, la differenza è radicale in quanto la variabile che manipoliamo al fine di produrre l'antecedente che ci interessa è una variabile *ad hoc*, il cui compito è esclusivamente quello di modificare i valori della variabile *target* e non è nemmeno compresa nel modello causale di partenza, bensì fa parte di un *network causale* aumentato nel quale essa vive come meccanismo apposito. Lo status delle due variabili è, quindi, estremamente differente e, per questo, mi sento di dire che la proposta di Hiddleston, in questo caso, mi sembra non in linea con la prassi seguita normalmente per effettuare interventi.

La seconda problematica che questa trattazione presenta riguarda sempre la modifica di variabili causalmente precedenti rispetto alla variabile che rappresenta l'antecedente, ma per altri motivi. Nella fattispecie, si noti che la variabile che nel secondo contesto intendiamo modificare ($P = 0$) ha, nel *network causale*, una freccia che indica il possesso di influenza causale rispetto alla variabile che

⁸⁷ Per citare il più importante: [Lewis](#), *Counterfactuals* cit.

⁸⁸ Si rimanda alla [Sezione 1.2.3](#) per la trattazione.

rappresenta il conseguente (V). Si noti che questo non è un intervento vietato da alcuna definizione che **Hiddleston** propone, ed è permessa dalla **Definizione 1.3.1**. Questa è una palese violazione della condizione alla **Cond. I(3)** proposta da **Woodward**: la variabile che stiamo modificando ha un percorso direzionato che si dirige alla variabile che rappresenta il conseguente nel modello. **Woodward** pone la suddetta clausola proprio con l'obiettivo di evitare che i nostri interventi inquinino lo stato del sistema e che non vengano introdotte influenze indipendenti dall'antecedente.

Presentiamo un esempio in cui questa violazione ci appare palesemente errata:

ESEMPIO 1.3.3

Un paziente si presenta dal medico con dei gravi sintomi riconducibili ad una nota patologia, spesso fatale. Il medico che ha il paziente in carico si appresta a iniettare un farmaco estremamente efficace, ma un collega che conosce il paziente lo ferma, facendogli notare il bracciale del paziente che indica un'allergia letale a quel farmaco. Al paziente vengono, quindi, somministrate solo cure di supporto, sperando che sopravviva. Il paziente muore a causa della patologia. ■

Il modello in **Figura 1.10** rappresenta la situazione, definita dalle seguenti equazioni:

$M = 1$	che rappresenta la presenza della malattia
$O = 1$	che rappresenta l'osservazione del bracciale
$F = M \wedge \neg O$	che rappresenta l'iniezione del farmaco
$G = \neg M \vee F$	che rappresenta la guarigione
$V = G \wedge \neg F$	che rappresenta la sopravvivenza

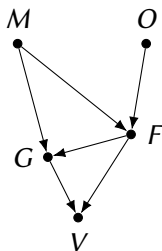


Figura 1.10

Valutiamo ora il controfattuale:

- (1) Se il paziente fosse guarito ($G = 1$), allora sarebbe sopravvissuto ($V = 1$).

La valutazione di questo controfattuale è, per quanto mi riguarda, piuttosto chiara: il controfattuale è vero senza ombra di dubbio. Vediamo come **Hiddleston** potrebbe discutere questo esempio con la sua teoria: vogliamo rendere $G = 1$, e abbiamo diversi modi per farlo. Iniziamo proponendo la *rottura causale* $M = 0$. Questo ci porta a sostenere che il paziente sarebbe sopravvissuto.

Questa trattazione risulta, come minimo, artificiosa. Quello che ci colpisce come assolutamente contrario al senso comune è che la guarigione viene imposta attraverso la non malattia. Questa è una modalità di intervento estremamente controintuitiva, ma non è questo quello che ci interessa: ciò che è rilevante è che questa modifica agisce sul conseguente (V) solo attraverso l'antecedente del controfattuale (G). Questo suggerisce che il risultato sarà quello intuitivamente ritenuto corretto,

in quanto non vi sono strade causali alternative che influenzano il conseguente senza influenzare l'antecedente, ovvero, l'influenza che la *rottura causale* ha sul conseguente è propagata esclusivamente attraverso la variabile G .

Ma passiamo, ora, alla nostra seconda opzione: porre la *rottura causale* sulla variabile F e portarla al valore di 1. Questo causerebbe la guarigione del paziente, ma sappiamo che ne causerebbe anche la morte istantanea a causa di una reazione avversa. Quindi, seguendo questa interpretazione, il controfattuale si rivelerebbe falso.

Questa è un'interpretazione palesemente scorretta, e risulta dal fatto che, come precedentemente fatto notare, l'intervento avviene su di una variabile che influenza il conseguente indipendentemente dall'antecedente. Questo provoca delle influenze sul conseguente che non dipendono direttamente dallo stato della variabile antecedente, ma da una costellazione di variabili che vengono modificate dall'intervento (in questo caso solo la variabile di intervento). Onde evitare questo tipo di problematiche, è necessario modificare la definizione di controfattuale vero, richiedendo che la *rottura causale* del modello Φ -minimale sia *pura* rispetto a Ψ , con la seguente definizione di *rottura causale pura*:

DEFINIZIONE 1.3.2 (ROTTURA CAUSALE PURA)

Una rottura causale è pura rispetto ad una proposizione Ψ sse esiste un solo cammino direzionato dalla variabile soggetta alla rottura alla variabile (o alle variabili) che rappresenta Ψ .

Ritornando alle due analisi, esse ci colpiscono per la loro stranezza ed artificiosità: la modalità di interpretazione non è intuitivamente immediata, non ci appare naturale. Ritengo che il motivo per cui queste analisi provocano questo effetto in sede di valutazione, per così dire, ci fanno *storcere il naso* è che non rispettano il modo con cui ragioniamo a riguardo e valutiamo i controfattuali.

Questo avviene perché, quando riflettiamo riguardo a queste situazioni, cerchiamo attivamente di valutarle attraverso un cambiamento minimale, seguito da una simulazione guidata dalla nostra conoscenza riguardo al dominio su cui stiamo effettuando la simulazione⁸⁹, e le discussioni di **Hiddleston** non fanno altrettanto⁹⁰. Si noti che la discussione delle due opzioni precedenti è parallela alla discussione di **Hiddleston** dell'**Esempio Tuffo nel vuoto**: vengono modificate, in base all'interpretazione, due variabili precedenti alla proposizione Φ , si modificano i valori e, infine, si valuta il controfattuale. Nessuna delle due opzioni da me discusse risulta la nostra naturale interpretazione della situazione.

Il cambiamento che cerchiamo, quanto riflettiamo su questo controfattuale, è quello di una guarigione spontanea, ovvero quando la malattia è presente e la cura non è somministrata. La stessa teoria di **Hiddleston** ci spiega il perché quelle discusse non ci sembrano interpretazioni naturali, mentre quella di una guarigione spontanea collima maggiormente con il modo in cui pensiamo alla situazione: la **Cond. (ii)** della **Definizione 1.3.1** non è rispettata. Se noi interveniamo o modifichiamo una variabile causalmente antecedente ad una delle variabili parte di Φ (generiamo una *rottura causale* in una variabile precedente, quindi), il nostro insieme di non discendenti intatti sarà sempre della forma $\vec{Z}_Y = \{X: X \notin \text{de}(\Phi)\} \setminus Y$, con Y la variabile su cui pratichiamo la *rottura*

⁸⁹ Un ottimo sondaggio della letteratura psicologica a riguardo, con un focus importante sui controfattuali *backtracking*, è: **Popa**, «*Getting counterfactuals right: the perspective of the causal reasoner*» cit.

⁹⁰ Nonostante ciò, si fa notare che alcuni esperimenti, seppur considerati non privi di problematiche, sembrano suggerire che non sia sempre questo il caso (**Gerstenberg et al.**, «*Back on track: Backtracking in counterfactual reasoning*» cit.).

causale e $de(\Phi)$ che rappresenta i discendenti delle variabili parte di Φ . Ma se noi pratichiamo la nostra *rottura causale* su Φ direttamente, avremo:

$$\vec{Z}_Y \subset \vec{Z}_\Phi$$

di conseguenza, l'insieme massimamente intatto sarà sempre quello che non comprende una variabile d'intervento diversa da Φ .

Tenendo conto di quanto detto sopra, ritengo che la vera proposta di **Hiddleston** (anche se mai esplicitamente caratterizzata in questi termini), sia la seguente:

1. Enumerare tutti i modelli in cui si verifica Φ spontaneamente, ovvero seguendo le leggi causali rappresentate dal sistema.
2. Selezionare il modello Φ -minimale tra questi.
3. Valutare il controfattuale.

Questa procedura permette di evitare parzialmente le problematiche, ma implica un esplicito ed esclusivo appoggio alla procedura *backtracking*, contro una grossa porzione di letteratura sulla causalità. Questo rende anche più debole l'affermazione di poter catturare i controfattuali *backtracking*, in quanto incorporati, in partenza, all'interno del processo di scelta del modello su cui valutare il controfattuale. Nuovamente, per certi versi, la proposta pare estremamente controintuitiva, seppure vi siano evidenze del fatto che essa mimi, da vicino, la modalità con la quale ragioniamo in quei casi in cui dobbiamo cercare le cause di uno stato controfattuale presentatoci⁹¹.

Possiamo quindi affermare che la teoria di **Hiddleston** collimi con le nostre intuizioni quando ci troviamo in situazioni di diagnosi e ricostruzione del passato, ma non quando il nostro focus sono gli effetti di uno stato controfattuale. La proposta di Hiddleston, quindi, caratterizzata come una ricerca nello spazio dei Φ -modelli al fine di rintracciare quello con la *rottura causale* minima è un interessante alternativa all'approccio standard proposto dai teorici dell'intervento.

Un'ultima problematica che ci preme presentare e che segue da questa modalità di interpretazione e valutazione dei controfattuali è l'impossibilità di valutare situazioni *non lawful*: se le leggi del modello assegnano una probabilità $p(X = x) = 0$, allora non possiamo utilizzarlo come antecedente di un controfattuale. Questo limita in qualche maniera il range di controfattuali che possiamo valutare. Pur non essendo una grave limitazione, il controfattuale $S_3 = 1 \sqcap \rightarrow L = 1$ non è valutabile nel seguente modello, rappresentato in **Figura 1.11**.

$$S_1 = A \oplus B$$

$$S_2 = A \vee B$$

$$S_3 = S_1 \wedge S_2$$

$$L = S_3$$

Infatti, se prendiamo $\Phi = \{S_3 = 1\}$, l'insieme dei Φ -modelli è vuoto⁹². Si fa notare, *en passant* che un approccio à la **Pearl** non ha problemi nella valutazione del controfattuale, che risulta vero.

⁹¹ *Ivi*.

⁹² Questo vale in un caso deterministico. La teoria, corredata con un *twist* probabilistico, potrebbe rendere la situazione trattabile, ad esempio associando una probabilità ad un malfunzionamento che permette $S_3 = 1$.

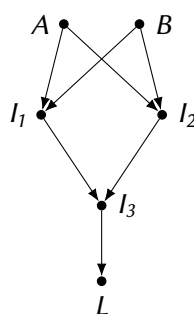


Figura 1.11

Un'obiezione a questa critica potrebbe essere formulata basandosi sull'idea che se una variabile non può assumere quel valore all'interno del modello allora essa non deve comprenderlo nel proprio *range* \mathcal{R} . Mi sento di dissentire, in quanto, quando progettiamo un modello, utilizziamo determinate variabili principalmente per due motivi:

- Rappresentano meccanismi distinti da altri e vogliamo mantenere la distinzione in quanto la riteniamo importante.
- Consideriamo la possibilità di intervenire come plausibile e, di conseguenza, ci impegniamo all'utilizzo di una variabile per rappresentare questa possibilità di intervento, anche se la variabile non può, naturalmente, assumere quel valore.

Ritengo che il secondo di questi motivi sia estremamente importante per la pratica scientifica. Dopotutto, la creazione di situazioni artificiali che non si verificano in natura è questione di ordinaria amministrazione per la scienza odierna, e il rappresentare queste possibilità risulta fondamentale. Ad esempio potremmo essere interessati a discernere l'effetto causale di una serie di interventi di ingegneria genetica su determinate caratteristiche di una pianta. Pur non essendo possibile, in natura, che una pianta erediti un gene che le fa produrre una proteina che protegge le cellule dal freddo, risulta, per noi, una possibilità concreta. Si immagini ora di voler analizzare (per quanto la questione possa essere complicata) l'effetto che un tale intervento sul genoma comporta a livello di velocità di crescita. Uno scienziato potrebbe voler valutare i diversi meccanismi causali attraverso cui questa inserzione opera.

Insomma, risulta assai semplice immaginare una serie di interventi *impossibili* a livello di modello anche nella pratica scientifica. Inoltre, se l'obiettivo è una semantica dei controfattuali, è facile pensare a controfattuali con antecedenti *fisicamente impossibili*⁹³.

Ritengo, in ogni caso, interessante notare che il lavoro di Hiddleston sia stato utilizzato, nella fattispecie, nello sviluppo di un'analisi dei condizionali standard (non controfattuali), ovvero i cosiddetti condizionali indicativi, del tipo:

1. Se non è stato Giovanni a mangiare la cioccolata, è stato qualcun altro.

La semantica sviluppata da Deng e Lee⁹⁴ fa, infatti, uso del concetto di *extrapolation* al fine di trattare questi condizionali indicativi. Questo concetto è estremamente simile nell'idea di partenza, seppur più dettagliato, all'analisi proposta da Hiddleston: l'*extrapolation* è l'imposizione di un valore

⁹³ Per citare il più classico: *se Giulio Cesare fosse vissuto ai tempi del Vietnam, avrebbe usato la catapulta*.

⁹⁴ D. Deng e K. Y. Lee, «Indicative and counterfactual conditionals: a causal-modeling semantics», *Synthese*, vol. 199, n. 1-2 (2021), pp. 3993-4014, DOI: [10.1007/s11229-020-02966-9](https://doi.org/10.1007/s11229-020-02966-9).

di una variabile in un modello causale attraverso una *risrittura* della sua storia causale. Detto in altri termini, il *focus* della *extrapolation* è l'integrità del sistema di leggi causali, ovvero il totale rispetto delle *equazioni strutturali*: se la variabile può assumere il valore $Y = y$ solo se le variabili genitore hanno il valore $\vec{Z} = \vec{z}$, allora è necessario che queste assumano quel valore, e che le variabili genitore di \vec{Z} assumano, a loro volta, i valori necessari secondo le equazioni strutturali, e così ricorsivamente⁹⁵, un approccio molto vicino ad una ricerca nello spazio dei Φ -modelli proposta da [Hiddleston](#).

⁹⁵ Per un'analisi dettagliata del concetto di *extrapolation* e per la semantica proposta si rimanda a [ivi](#).

2

LA CAUSA EFFETTIVA

Una delle grandi problematiche che hanno costituito un rompicapo per chi si occupava di modelli causali è stata quella della determinazione della causa effettiva. Con causa effettiva, in questo capitolo, denoteremo quale(i) variabile(i) ha causato il dato valore di una data variabile (o di date variabili). Questo limitato campo di ricerca, estremamente dibattuto ed in fermento, si occupa fondamentalmente di proporre una nozione teorica in grado di discriminare che valore di quale variabile abbia prodotto il tal valore dell'altra variabile nella presente istanziazione del modello causale o, per utilizzare un linguaggio più tecnico, nel mondo attuale, con la definizione di mondo data nella **Definizione 1.1.4.**

Come ho detto, il problema è considerato estremamente affascinante, oltre che di grande importanza (si pensi a casi di attribuzione di responsabilità in sede giuridica), e la quantità di letteratura reperibile a riguardo è immensa. Si sono occupati di causa effettiva **Hitchcock, Halpern, Pearl, Woodward, Halpern e Pearl, Fenton-Glynn, Glymour e Wimberly, Glymour et al.** solo per citarne alcuni.

La mia discussione sarà basata principalmente su *Actual Causality* di **Halpern**, in quanto focalizzato esclusivamente sulla tematica e relativamente recente. La discussione seguirà un andamento graduale: verranno presentate delle definizioni preliminari e degli esempi che sono in grado di metterle in difficoltà, verrà corretta la definizione e così via, sino ad una definizione definitiva che sia estensivamente corretta. Il problema della spiegazione della scelta dei criteri formali verrà lasciato in larga parte ignorato e, quindi, ci si concentrerà principalmente sulla capacità della teoria di incorporare la maggior parte dei casi in maniera corretta.

2.1 LA PROPOSTA DI LEWIS E LE SUE MANCANZE

La maggior parte degli account sviluppati per definire la causalità si ispirano alla proposta di Lewis¹. L'idea di Lewis è, sommariamente, che se tra due eventi² sussiste una relazione di dipendenza controfattuale, dove i controfattuali sono di tipo non *backtracking*, allora sussiste una relazione di dipendenza causale, dato che la dipendenza causale è l'ancestrale della dipendenza controfattuale. Di conseguenza, se $C \Box \rightarrow E$ e $\neg C \Box \rightarrow \neg E$, allora vi è relazione di dipendenza causale tra C ed E e, se nel mondo attuale si verifica $C \Box \rightarrow E$, allora C è la causa effettiva di E . Lewis ha sempre sostenuto

¹ D. Lewis, «Causation», *Journal of Philosophy*, n. 70 (1973), pp. 556-67.

² Vi sono account che propongono proprietà di eventi (L. A. Paul, «Aspect Causation», in J. Collins et al. (a cura di), *Causation and Counterfactuals*, MIT Press, 2004, pp. 205-224), o altri relata causali. Il problema dei relata causali è ampio e, per questioni di spazio, non verrà affrontato in questa sede. **Ehring** propone una buona introduzione: D. Ehring, «Causal Relata», in H. Beebe et al. (a cura di), *The Oxford Handbook of Causation*, Oxford University Press, 2010, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001), pp. 387-413, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001).

che la relazione di dipendenza causale non è una proprietà necessaria tra due eventi, uno di cui è la causa effettiva dell'altro, ma solamente sufficiente.

La mera sufficienza della relazione di dipendenza causale è immediatamente evidente se consideriamo i cosiddetti casi di *early preemption*, ovvero casi in cui una causa è anticipata da un'altra. Vediamo un esempio:

ESEMPIO 2.1.1

Macchia e Romeo, due agili gatti, stanno dando la caccia ad un topo. Sono entrambi appostati a lato della tana del topo, pronti a catturarlo con un balzo, quando questo esce dal pertugio nel muro. Macchia, il più giovane dei due, salta una frazione di secondo prima di Romeo, il più anziano, e cattura il povero topo. Il balzo di Romeo non va a segno a causa del ritardo, nonostante fosse stato estremamente preciso e avrebbe determinato la medesima cattura, con l'unica differenza di un leggero ritardo. ■

Risulta evidente che in questa situazione non vi è relazione di dipendenza controfattuale tra l'attacco di Macchia M e la cattura del topo C . Questo perché, anche nel caso Macchia non avesse saltato per prendere il topo, il topo sarebbe stato preso da Romeo. Non sussistendo relazione di dipendenza controfattuale, in quanto $\neg C \Box \rightarrow \neg E$ risulta falso, non sussiste nemmeno relazione di dipendenza causale e, di conseguenza, il balzo di Macchia non è una causa effettiva della cattura del topo. Si tratta, quindi, di cercare strategie per affermare che Macchia è la causa della cattura del topo.

Non è possibile, naturalmente, discriminare i due eventi in base a chi ha effettuato la cattura, in quanto presupporrebbe informazioni causali che stiamo, per l'appunto, cercando di determinare; se provassimo ad utilizzare questo tipo di informazioni il nostro account sarebbe palesemente circolare e, di conseguenza, destinato al fallimento. Un'altra strategia, proposta anche da Lewis e poi abbandonata, è stata quella di promuovere un account estremamente *fragile* degli eventi.

Un account fragile è un account che sostiene che, se due eventi sono identici sotto tutti gli aspetti, ma tra essi vi è uno scarto, anche estremamente minuto, rispetto ad una proprietà o una caratteristica, essi sono due eventi distinti: bere un espresso da una tazzina inclinandola di un angolo di 35° è un evento differente da bere un espresso da una tazzina inclinandola di un angolo di $35,00 \dots 1^\circ$. Il problema di un account che sostiene la fragilità degli eventi è che porta ad una vasta gamma di conclusioni implausibili e controintuitive: se un sassolino scheggia a livello microstrutturale il mio parabrezza, quel sassolino è una causa del mio guidare l'auto questa mattina, in quanto se il sassolino non avesse scheggiato il parabrezza starei guidando un'auto differente e il mio *guidare l'auto* sarebbe un diverso *guidare l'auto*.

Di conseguenza, al fine di catturare queste situazioni causali, Lewis ha proposto di definire la causalità nei termini della nozione di *quasi-dipendenza*³.

DEFINIZIONE 2.1.1 (QUASI-DIPENDENZA)

Un evento E quasi-dipende rispetto ad un evento C se e solo se essi sono parti distinte di un processo che condivide le caratteristiche intrinseche con un processo nomologicamente possibile nel quale la controparte di E dipende dalla controparte di C .

Ma, negli anni recenti, una lunga serie di *puzzle* causali sono stati proposti e Lewis ha aggiornato nuovamente la propria definizione, questa volta basandosi sulla nozione di *influenza*, sempre legata alla nozione di fragilità:

³ Lewis, «Postscripts to 'Causation'» cit.

Dove C ed E sono eventi effettivi⁴ distinti, diciamo che C *influenza* E sse vi è un range sostanziale C_1, C_2, \dots di diverse non-troppo-differenti alterazioni di C (incluso l'alterazione effettiva di C) e vi è un range E_1, E_2, \dots di alterazioni di E , delle quali almeno qualcuna differente, tale che se C_1 fosse accaduto, E_1 sarebbe accaduto, e se C_2 fosse accaduto, E_2 sarebbe accaduto, e così via. Quindi, abbiamo un pattern di dipendenza controfattuale di se, quando e come rispetto a se, quando e come. (Come prima, la causazione è l'ancestrale: C causa E sse vi è una catena di influenza passo a passo da C ad E .)⁵

Così facendo, Lewis è riuscito a catturare un gran numero di intuizioni causali che normalmente abbiamo. Il problema è che questa definizione cattura fin troppo: qualsiasi evento che ha una anche minima influenza su di un altro viene considerato una causa di quest'ultimo, in linea con l'idea che le *alterazioni* siano estremamente fragili. Inoltre, le definizioni lewisiane di causa soffrono di molti altri problemi⁶.

La fragilità e le varie definizioni di Lewis, infine, non permettono di proteggersi dai casi di *sovradeterminazione*, come il seguente:

ESEMPIO 2.1.2 (Incendio Simultaneo)

In una foresta, dopo un lungo periodo di siccità, cade un fulmine. Simultaneamente, un piromane lancia un fiammifero acceso all'interno della foresta. Un incendio divampa nella foresta. ■

In questo esempio non c'è alcuna dipendenza controfattuale o *quasi-dipendenza* o *influenza* tra l'incendio nella foresta e la caduta del fulmine, in quanto avremmo avuto il medesimo effetto (al netto dell'assunzione che gli incendi partano nello stesso punto, si sviluppino nello stesso modo, ecc.) anche senza il fulmine (o il piromane). Lewis ritiene che in questo tipo di situazioni non si possa ricorrere ad una fragilità modale estrema a causa delle problematiche di causazione spuria che questa genera⁷. Questo caso, quindi, decide anche, in maniera più o meno definitiva, che la relazione di dipendenza causale sommata all'attualità degli eventi non è sufficiente per la relazione di causazione effettiva⁸.

La ricerca di una relazione necessaria per la causazione effettiva ha occupato, negli anni recenti, una vasta fetta del dibattito sulla causalità. In particolare, una forte linea di argomentazione a riguardo è stata prodotta dai teorici dei modelli causali, che hanno aggiunto un *twist* interventista alla relazione di Lewis, codificando le relazioni attraverso, appunto, i modelli causali. I principali teorici che hanno impostato la questione in tali termini, e che hanno avviato una linea di ricerca nella direzione della causalità effettiva formalizzata attraverso i **SEMs**, sono **Halpern e Pearl**.

⁴ Finora, *actual* è stato tradotto con effettivo. In questa sede viene mantenuta la convenzione, anche se forse meno applicabile al concetto di evento rispetto a quello di causa.

⁵ D. Lewis, «Causation as Influence», in **Collins et al.** (a cura di), *Causation and Counterfactuals* cit., pp. 75-106: p.91.

⁶ Si veda, ad esempio: G. Björnsson, «How Effects Depend on their Causes, Why Causal Transitivity Fails, and Why we Care About Causation», *Philosophical Studies*, vol. 133, n. 3 (2007), pp. 349-390, DOI: [10.1007/s11098-005-4539-8](https://doi.org/10.1007/s11098-005-4539-8); N. Hall, «Two Concepts of Causation», in **Collins et al.** (a cura di), *Causation and Counterfactuals* cit., pp. 225-276; J. Y. Halpern e C. Hitchcock, «Actual Causation and the Art of Modeling», in *Heuristics, Probability and Causality. A Tribute to Judea Pearl*, a cura di R. Dechter et al., College Publications, 2010; **Hiddleston**, «Causal Powers» cit.; C. Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason», *The Philosophical Review*, vol. 116, n. 4 (2007), pp. 495-532.

⁷ **Lewis**, «Causation» cit.: p. 198-199.

⁸ Lewis si è sempre dichiarato agnostico riguardo ai casi di sovradeterminazione, di conseguenza il fatto che la sua teoria non dia una risposta precisa è una caratteristica positiva più che un difetto a suo parere: **Lewis**, «Causation» cit.; **Lewis**, «Causation as Influence» cit.

Data questa premessa, seguiremmo l'evoluzione della loro proposta nel tempo al fine giungere al termine della sezione con una nozione di causa effettiva ben definita.

2.2 LA CAUSA EFFETTIVA SECONDO HALPERN E PEARL

Il primo lavoro di Halpern e Pearl a riguardo della causa effettiva è «Causes and Explanations: A structural-model approach. Part I: Causes»⁹, il cui obiettivo fondamentale è rettificare alcune problematiche riscontrate nella prima edizione di *Causality: Models, Reasoning, and Inference*¹⁰, risalente al 2000¹¹.

Oltre che ad offrire delle condizioni per la causa effettiva in grado di catturare situazioni ordinarie, la definizione si vuole attivamente confrontarsi con alcuni complessi controesempi alle ordinarie teorie della causalità, proposti da Bennet¹² e da Hall¹³. Il risultato è piuttosto interessante e, curiosamente, è in grado di affrontare anche dei complessi scenari di voto, considerati spesso un incubo per le teorie controfattuali¹⁴, che ricordano da vicino situazioni di sovradeterminazione causale.

Presenterò lo scenario e, dopo questo, la definizione di causa effettiva proposta da Halpern e Pearl.

ESEMPIO 2.2.1 (Voting Machine)

Due soggetti (V_1 e V_2) devono votare una mozione, che passa se uno dei due la accetta. Essi votano attraverso una macchina conta voti (M). Nel presente caso entrambi votano a favore e la mozione passa ($P = 1$). ■

Di seguito le equazioni del modello e, in Figura 2.1, la resa grafica:

$$M = V_1 + V_2$$

$$P = M \geq 1$$

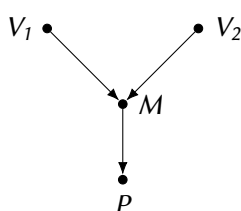


Figura 2.1: Modello causale dello scenario di voto

Questa situazione crea problemi alla definizione di causa effettiva proposta da Pearl nella prima edizione di *Causality: Models, Reasoning, and Inference*, in quanto egli sostiene che una causa effettiva è tale se e solo se, mantenendo le variabili esterne alla supposta causa stabili e intervenendo su

⁹ J. Y. Halpern e J. Pearl, «Causes and Explanations: A structural-model approach. Part I: Causes», in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001)*, 2001, pp. 194-201.

¹⁰ Pearl, *Causality: Models, Reasoning, and Inference* cit.

¹¹ Si veda in particolare *ivi*: Capitolo 10.

¹² J. Bennet, «Event Causation: The Counterfactual Analysis», in *Causation*, a cura di E. Sosa e M. Tooley, Oxford Readings in Philosophy, Oxford University Press, 1993.

¹³ Hall, «Two Concepts of Causation» cit.: non ancora pubblicato all'epoca.

¹⁴ J. Livengood, «Actual Causation and Simple Voting Scenarios», *Noûs*, vol. 47, n. 2 (2013), pp. 316-345, DOI: [10.1111/j.1468-0068.2011.00834.x](https://doi.org/10.1111/j.1468-0068.2011.00834.x).

quest'ultima modificandone il valore, si ottiene una variazione nell'effetto. Risulta evidente che in questo esempio la definizione di **Pearl** produce un risultato curioso: nessuno dei due voti è la causa di $P = 1$.

Da un lato, questo tipo di risultato cattura parzialmente una nostra intuizione, ovvero quella riguardo all'insicurezza di quale esattamente sia la causa, dato che entrambi i voti erano sufficienti. Dall'altro, è contrario alla posizione forse più normale da assumere, ovvero quella che entrambi i voti sono causa dell'approvazione della mozione. **Halpern e Pearl** ritengono che questa seconda posizione sia preferibile e, infatti, la loro definizione, presentata qui di seguito, porta esattamente a questo risultato. Verrà usata la notazione $[X \leftarrow x]$ come equivalente di $do(X = x)$, di conseguenza $(M, \vec{u}) \models [X \leftarrow x]$ è equivalente a $M_{X \leftarrow x}$. Definiamo, inoltre, un *evento primitivo* come una formula del seguente tipo: $X = x$, con $X \in \mathcal{V}$ e $x \in \mathcal{R}(X)$ si definisce, inoltre una *formula causale* come una formula dotata della seguente struttura: $[Y_1 \leftarrow y_1, \dots, Y_n \leftarrow y_n]\varphi$, dove φ è una combinazione booleana di eventi primitivi; Y_1, \dots, Y_n sono variabili distinte e $y_i \in \mathcal{R}(Y_i)$. Userò $(M, \vec{u}) \models \varphi$ se la formula causale φ è vera nel modello M con setting \vec{u} .

Questa è la definizione di causa effettiva *originale* della famiglia **HP**.

DEFINIZIONE 2.2.1 (CAUSA EFFETTIVA^{org})

$\vec{X} = \vec{x}$ è una causa effettiva di φ in (M, \vec{u}) se le seguenti condizioni sono verificate:

AC1. $(M, \vec{u}) \models (\vec{X} = \vec{x}) \wedge \varphi$

AC2. Esiste una partizione di \mathcal{V} in due sottoinsiemi \vec{Z} e \vec{W} (tali che $\vec{Z} \cap \vec{W} = \emptyset$) con $\vec{X} \subseteq \vec{Z}$ e un setting \vec{x}' e \vec{w}' delle variabili in \vec{X} e \vec{W} tale che, se $(M, \vec{u}) \models Z = z^*$ per $Z \in \vec{Z}$, allora:

$(a^{org}) (M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}'] \neg \varphi$

$(b^{org}) (M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}', \vec{Z}' \leftarrow \vec{z}^*] \varphi$ per tutti i sottoinsiemi \vec{Z}' di $\vec{Z} \setminus \vec{X}$

AC3. \vec{X} è minimale; ovvero nessun sottoinsieme di \vec{X} soddisfa le *Cond. AC1 e Cond. AC2*¹⁵

Come affermato in precedenza, si noti che l'utilizzo della notazione vettoriale serve a denotare insiemi insiemi; di conseguenza, quella fornita da **Halpern e Pearl** è una definizione di causa effettiva che accetta insiemi di variabili come cause¹⁶. Durante la discussione questa particolarità verrà omessa, quando non specificatamente rimarcato. Procediamo con la presentazione della definizione.

La prima condizione richiede che causa ed effetto si verifichino nel modello date le condizioni attuali (\vec{u} , ovvero il mondo attuale secondo la **Definizione 1.1.4**). Questa è una condizione condivisa da tutte le definizioni di causa effettiva della famiglia **HP** e, di conseguenza, quando verranno presentate altre definizioni di questa famiglia, la condizione verrà omessa. Vale lo stesso per la **Cond. AC3**, il cui scopo è quello di estromettere fattori non rilevanti dalla causa: senza questa condizione, se colpire una bottiglia ne causa la rottura, allora il colpirla mentre si fa un salto mortale sarebbe anch'essa una causa. La condizione si configura come una sorta di determinazione *Insufficient, but Necessary part of an Unnecessary but Sufficient condition (INUS)*¹⁷: gli eventi che fanno parte della causa sono necessari all'effetto.

¹⁵ Halpern e Pearl, «Causes and Explanations: A structural-model approach. Part I: Causes» cit.

¹⁶ Eiter e Lukasiewicz (2002) dimostra, in realtà, che le cause individuate dalla **Definizione Causa effettiva^{org}** sono sempre singoli congiunti (T. Eiter e T. Lukasiewicz, «Complexity results for structure-based causality», *Artificial Intelligence*, vol. 142, n. 1 (2002), pp. 53-89, DOI: [10.1016/S0004-3702\(02\)00271-0](https://doi.org/10.1016/S0004-3702(02)00271-0): p.70).

¹⁷ J. Mackie, «Causes and Conditions», *American Philosophical Quarterly*, n. 2 (1965), pp. 245-264.

Il fulcro della definizione, nonché suo motore, è la **Cond. AC2**. Questa condizione innanzitutto ci intima di dividere l'insieme delle variabili in due insiemi disgiunti, il primo (\vec{Z}) contenente le variabili che appartengono a quella che potremmo chiamare la *strada causale* o *percorso causale*, tra la variabile causa e la variabile effetto; il secondo rappresenta le variabili esterne a questo percorso. Questa divisione viene effettuata nella maniera esposta in quanto le variabili appartenenti a \vec{Z} devono essere libere di variare a seconda delle variazioni imposte dalla **Cond. AC2(a^{org})**, altrimenti l'influenza causale sarebbe impossibilitata a propagarsi e giungere a φ .

Ora è necessario trovare un *setting*, ovvero degli specifici valori delle variabili in \vec{X} e \vec{W} , anche differenti da quelli che si verificano nel modello (M, \vec{w}) lasciato intatto¹⁸. Se, dato questo *setting* le due sottocondizioni sono rispettate, allora \vec{X} è una causa effettiva di φ . La prima sottocondizione (**Cond. AC2(a^{org})**), richiede che l'intervento che realizza il *setting* non permetta a φ di verificarsi. Questa condizione è equivalente al test controfattuale di Lewis, solo valutato in una specifica contingenza $\vec{W} = \vec{w}'$.

La seconda sottocondizione (**Cond. AC2(b^{org})**), invece, richiede che il valore selezionato \vec{w}' non sia, di per sé, sufficiente a modificare φ indipendentemente da \vec{X} e tutti i sottoinsiemi \vec{Z}' settati a \vec{z}^* . Si configura, quindi, come una condizione che richiede la sufficienza di \vec{X} per il realizzarsi di φ . Questa seconda condizione non equivale ad affermare che il *setting* \vec{w}' non ha alcun effetto causale su φ bensì solo che questo non è in grado di modificarne il valore indipendentemente da \vec{X} e \vec{Z} . Di conseguenza, il cambiamento osservato nel valore di φ quando interveniamo $\vec{X} = \vec{x}'$ è attribuibile soltanto a \vec{X} . Si noti che il *setting* \vec{w}' può avere influenza sugli elementi di \vec{Z} in quel momento non settati a \vec{z}^* (ovvero non appartenenti all'insieme \vec{Z}' in quell'istanza sotto analisi).

Riprendendo l'**Esempio Voting Machine**, vediamo che, scegliendo $\vec{Z} = \{V_1, M, P\}$ e come $W = \{V_2\}$ con $w' = \{V_2 = 0\}$, la **Cond. AC2** è soddisfatta, rendendo $V_1 = 1$ una causa dell'esito del voto. La medesima argomentazione vale per V_1 .

Presentiamo anche la definizione ausiliaria del concetto di testimone, che verrà utilizzata successivamente:

DEFINIZIONE 2.2.2 (TESTIMONE)

Si dice testimone una tripla $(\vec{W}, \vec{w}, \vec{x}')$ tale che $\vec{X} = \vec{x}$ è una causa effettiva di φ , secondo una delle definizioni **HP**.

L'obiettivo principale di questa definizione di causa effettiva era quello di affrontare casi di sovradeterminazione e di *early preemption*. Abbiamo analizzato il caso della sovradeterminazione o di relazioni simili attraverso l'**Esempio Voting Machine**. Procediamo ora con la trattazione dei casi di *early preemption*.

ESEMPIO 2.2.2 (Backup)

Un Assassino Apprendista (A) è incaricato di uccidere un soggetto SM . Il suo Mentore M lo segue, per assicurarsi la riuscita della missione: nel caso A non spari, si occuperà lui di eliminare il soggetto. Entrambi gli assassini sono infallibili dalla distanza da cui spareranno. L'assassino in addestramento spara. Il soggetto muore ($SM = 1$)¹⁹. ■

¹⁸ Per la variabile \vec{X} questa differenza è necessaria (equivale, grossomodo, al test controfattuale proposto da Lewis (**Lewis**, «**Causation**» cit.)), per \vec{W} invece no.

¹⁹ C. Hitchcock, «The Intransitivity of Causation Revealed in Equations and Graphs», *The Journal of Philosophy*, vol. 98, n. 6 (2001), pp. 273-299, DOI: [10.2307/2678432](https://doi.org/10.2307/2678432); p.276.

Come per i casi precedenti, di seguito equazioni strutturali e grafo causale (Figura 2.2).

$$M = \neg A$$

$$SM = A \vee M$$

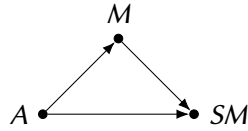


Figura 2.2

La domanda è, ovviamente, se $A = 1$ sia una causa di $SM = 1$. L'approccio di Lewis fallisce, in quanto non abbiamo relazione di dipendenza controfattuale tra le due variabili: nella fattispecie $\neg A \Box \rightarrow \neg SM$ è chiaramente falso. Ma se mantenessimo la contingenza in cui il mentore non spara, avremmo indietro la nostra dipendenza controfattuale e, quindi, saremmo in grado di sostenere che il fatto che l'assassino spari è la causa della morte del soggetto. Questa è l'idea dietro al setting \vec{w}' : l'imporre alcune contingenze, verificatesi o meno (e vedremo che questo sarà un fatto chiave, nel prosieguo), ci permette di reinstaurare una relazione di dipendenza controfattuale, permettendoci di individuare la causa effettiva.

L'idea fondamentale che questa definizione di causa effettiva racchiude è la seguente: la causalità effettiva non richiede una relazione di dipendenza controfattuale *tout court*, bensì la relazione di dipendenza controfattuale sotto alcune specifiche circostanze. Questa intuizione è dovuta, principalmente, a Pearl²⁰, che in *Causality: Models, Reasoning, and Inference* propone, grossomodo, il seguente criterio: $X = x$ è causa effettiva di $Y = y$ se tra queste vi è dipendenza controfattuale nel sottomodello in cui tutte le variabili, escluse X e Y , sono settate al valore che avrebbero avuto nel contesto attuale \vec{u} ²¹. L'idea di sottomodelli realizzato attraverso interventi sui quali testare dipendenze controfattuali è, potremmo dire, la chiave di volta per la questione causale, secondo questi autori.

Una proprietà interessante della **Definizione Causa effettiva**^{org}, dimostrata da Hopkins e Pearl²², è che qualsiasi condizione **INUS** è classificata come causa. Questa proprietà ci suggerisce che la definizione potrebbe essere, e vedremo, a breve, che questo è il caso, troppo permissiva rispetto ad un account corretto della causa effettiva. Si consideri, infatti, il seguente caso:

ESEMPIO 2.2.3 (Plotone di Esecuzione)

Un plotone di esecuzione è formato da B e C . A è l'incaricato della manutenzione e del caricamento dell'arma di B , mentre C è autonomo. Il giorno dell'esecuzione del prigioniero A carica il fucile di B . C spara al prigioniero D , che muore. B decide di non fare fuoco. ■

L'equazione strutturale che governa il modello è unica, ed è la seguente: $D = (A \wedge B) \vee C$. Il grafo causale, elementare, è riportato in Figura 2.3.

²⁰ Pearl, *Causality: Models, Reasoning, and Inference* cit.

²¹ La definizione di Pearl ha dei vincoli più elaborati ma immateriali ai fini della presentazione in questa sede.

²² M. Hopkins e J. Pearl, «Clarifying the usage of structural models for commonsense causal reasoning», in *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 2003, pp. 83-89.

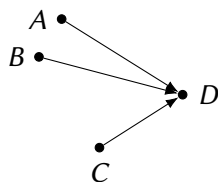


Figura 2.3: Grafo dell'Esempio Plotone di Esecuzione

Il problema della **Definizione Causa effettiva^{org}** è che considera $A = 1$ come una causa di $D = 1$, con $\vec{W} = C, B$ e $\vec{w} = (0, 1)$: nel caso C non avesse sparato e, invece, B avesse sparato, il prigioniero sarebbe morto. Selezionando questo setting di W abbiamo, quindi, dipendenza controfattuale tra $A = 1$ e $D = 1$ e, di conseguenza, un'attribuzione di causalità con la quale non siamo minimamente d'accordo e che ci pare palesemente scorretta. Si noti, inoltre, che anche $B = 1$ (interpretando questo come “ B preme il grilletto”) verrebbe considerata una causa di $D = 1$ nel contesto in cui $A = 0$ e $C = 1$, in quanto con il setting $A = 1$ e $C = 0$, allora D dipende da B .

Una delle soluzioni praticabili, e che **Halpern** sostiene sia una modalità generale di affrontare problematiche di questo tipo per la **Definizione Causa effettiva^{org}**, sia quella di aggiungere una variabile $B_1 = A \wedge B$, a rappresentare il far fuoco di B con un fucile carico, e $D = B_1 \vee C$ (producendo il grafo in **Figura 2.4**)²³: ora, con $B_1 \in \vec{Z}$ non è più possibile stabilire una relazione di dipendenza controfattuale tra $A = 1$ e $D = 1$, in quanto, nel mondo attuale $B_1 = 0$ e, quindi, anche se $A = 1$, $D = 0$.

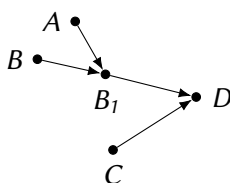


Figura 2.4: Grafo con aggiunta della variabile B_1

Questa soluzione non è sembrata soddisfacente²⁴ e, quindi, **Halpern e Pearl** hanno preferito proporre una nuova versione della loro definizione, modificata in modo da catturare anche questo tipo di casi e fronteggiare la problematica posta in luce da **Hopkins e Pearl**²⁵. Vediamo questa nuova definizione, che chiameremo *aggiornata* (AC^a). Come precedentemente affermato, le **Cond. AC1** e **Cond. AC3** verranno omesse in quanto identiche. In questa definizione anche la **Cond. AC2(a^{org})** verrà omessa per il medesimo motivo.

²³ Halpern, *Actual Causality* cit.: p.59-60.

²⁴ Vedremo che in seguito, invece, questo tipo di mosse verranno incoraggiate da Blanchard e Schaffer.

²⁵ Hopkins e Pearl, «Clarifying the usage of structural models for commonsense causal reasoning» cit.

DEFINIZIONE 2.2.3 (CAUSA EFFETTIVA^a)

$\vec{X} = \vec{x}$ è una causa effettiva di φ in (M, \vec{u}) se le seguenti condizioni sono verificate:

$AC_2(b^a)$. Se \vec{z}^* è tale che $(M, \vec{u}) \models \vec{Z} = \vec{z}^*$, allora per tutti i sottoinsiemi \vec{W}' di \vec{W} e tutti i sottoinsiemi \vec{Z}' di $\vec{Z} \setminus \vec{X}$, abbiamo: $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*] \varphi^{26}$

A parole, questa versione aggiornata della definizione richiede che non solo la dipendenza controfattuale tra causa ed effetto si verifichi in un certo setting, per ogni sottoinsieme del percorso causale tra causa ed effetto, ma che questa dipendenza sussista anche nel caso in cui gli interventi sono limitati a soli dei sottoinsiemi di W . Questa definizione di causalità corregge, quindi, parte di quella permissività che era il motivo per cui qualsiasi parte di condizione **INUS** veniva considerata una causa, limitandosi, ora, a considerare cause solo condizioni sufficienti a produrre l'effetto in una varietà di casi.

Con questa definizione, nell'**Esempio Plotone di Esecuzione**, $A = 1$ non è una causa di $D = 1$. Immaginiamo, infatti, di replicare il ragionamento fatto in precedenza. Abbiamo $(M, \vec{u}) \models A = 1 \wedge D = 1$; con $\vec{W} = \{B, C\}$ e $\vec{w} = (1, 0)$, $(M, \vec{u}) \models [A \leftarrow 0, B \leftarrow 1, C \leftarrow 0] D = 0$, ma se ora, seguendo la **Cond. AC₂(b^a)**, dobbiamo testare il setting su tutti i sottoinsiemi di \vec{W} , compreso $\vec{W}' = \emptyset$ o $\vec{W}' = \{C\}$, allora in questi casi $(M, \vec{u}) \models [A \leftarrow 0, C \leftarrow 0] D = 0$, quindi non vi è alcuna relazione di dipendenza controfattuale tra le due variabili: B , assumendo il valore che ha nel mondo attuale ($B = 0$), interromperebbe l'influenza causale di A su D . Lo stesso discorso può essere fatto per $B = 1$ nel caso in cui $A = 0$.

Questa definizione sembra, quindi, catturare le nostre intuizioni anche in casi problematici, e, infatti, a titolo di esempio, si consideri il seguente caso di *double prevention* proposto da Hall²⁷:

ESEMPIO 2.2.4 (Doppia prevenzione)

Suzy e Billy sono ormai cresciuti quando scoppia la Terza Guerra Mondiale. Suzy, diventata un pilota di bombardieri, partecipa ad una missione con l'obiettivo di attaccare un nodo di comunicazione centrale per i nemici. Billy, pilota di caccia, fa da sua scorta in questa missione. Un aereo nemico si avvicina minaccioso, pilotato dal Nemico. Billy, rilevata la segnatura radar di Nemico, effettua una manovra di ingaggio e, premuto il grilletto, abbatte l'aereo di Nemico. Suzy procede indisturbata e il bombardamento va a buon fine, come previsto. ■

È Billy parte del successo della missione? Se consideriamo il fatto che se non avesse abbattuto Nemico, quest'ultimo avrebbe abbattuto Suzy, intuitivamente la nostra risposta è affermativa: Billy è, quantomeno, parte di una causa del successo della missione (assieme, naturalmente, al bombardamento di Suzy).

Il modello è semplicemente formato da variabili binarie che rappresentano gli eventi, con grafo in **Figura 2.5**.

- BS , per “Billy spara”
- NS , per “Nemico schiva”
- NSS , per “Nemico spara a Suzy”

²⁶ Halpern e Pearl, «Causes and Explanations: A structural-model approach. Part I: Causes» cit.

²⁷ Hall, «Two Concepts of Causation» cit.

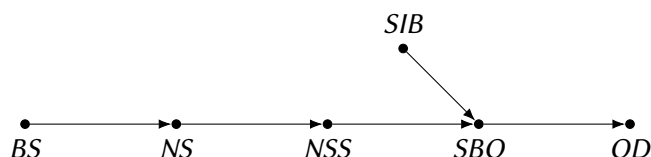
- *SBO*, per “Suzy bombarda obiettivo”
- *OD*, per “Obiettivo distrutto”



Figura 2.5

In questo modello *BS* è una causa di *OD*. Risulta, però, strano, sostenere che Billy sia la causa del fatto Suzy programmava di fare sin dall’inizio. Personalmente, e so di essere in minoranza, non mi sembra una conclusione così controintuitiva. Dopotutto vi era il rischio concreto che Suzy non riuscisse nella sua missione e, di conseguenza, l’essere causa di questo fatto per Billy è perfettamente normale. Ritengo che la problematica sia legata, principalmente, a quale tipo di controfattuali il soggetto guarda quando fa questo tipo di ragionamenti.

In ogni caso, per rendere il modello più flessibile e in grado di gestire questo tipo di differenze nei giudizi, è sufficiente aggiungere una variabile. Osserviamo il network aumentato in [Figura 2.6](#), dove *SIB* rappresenta l’intenzione di Suzy a bombardare l’obiettivo.

Figura 2.6: Network aumentato rispetto alla [Figura 2.5](#)

Ora *SIB* e *BS* sono entrambi cause di *OD*, e il problema si risolve.

Questa definizione aggiornata, però, ha ancora delle problematiche, peraltro sempre legate al teorema provato da [Hopkins e Pearl](#)²⁸. Questo controesempio è dovuto a [Weslake](#)²⁹.

ESEMPIO 2.2.5 (Lampadina)

Vi sono tre interruttori (*A*, *B*, *C*) con tre possibili posizioni $\{-1, 0, 1\}$. Una lampadina (*L*) si accende solo quando due di questi interruttori sono nella stessa posizione. Nella situazione attuale $A = 1, B = -1, C = -1$, e la lampadina è accesa. ■

L’equazione che descrive il comportamento della lampadina è la seguente:

$$L = (A = B) \vee (B = C) \vee (A = C)$$

Questa situazione rappresenta un problema per la definizione alla [Cond. AC2\(b^a\)](#). Si supponga di voler accertare se $A = 1$ sia una causa dell’accensione della lampadina. L’idea, conoscendo la storia, sembra assolutamente ridicola: l’interruttore *A* nella posizione 1 non svolge alcun ruolo causale nella storia: infatti, la lampadina è accesa proprio perché gli altri due interruttori sono nella posizione -1 . Purtroppo, non è questo che la definizione [HP^a](#) ci dice: se, infatti, prendiamo $\vec{W} = \{B, C\}$ e $\vec{w} = (1, -1)$, risulta, come rappresentato nella [Tabella 2.1](#), che $A = 1$ soddisfa tutte le condizioni per essere considerata una causa.

²⁸ Hopkins e Pearl, «Clarifying the usage of structural models for commonsense causal reasoning» cit.

²⁹ B. Weslake, «A Partial Theory of Actual Causation», *British Journal for the Philosophy of Science* (2015), in uscita.

\vec{W}'	Variabili			
	A	B	C	L
$B=1$	1	1	-1	1
$C=-1$	1	-1	-1	1
$C=-1, B=1$	1	1	-1	1
\emptyset	1	-1	-1	1

Tabella 2.1: Tabella con i valori dei sottoinsiemi \vec{W}' riguardanti l'Esempio Lampadina

Questo esempio nasconde assai più di quanto possa apparire e, infatti, sarà oggetto di discussione nel **Capitolo 3**, dove verrà riconsiderato alla luce dei risultati ottenuti. Al momento, in ogni caso, verrà trattato come un controesempio, per affrontare il quale è necessario proporre una nuova definizione di causa effettiva *modificata*.

DEFINIZIONE 2.2.4 (CAUSA EFFETTIVA^m)

$\vec{X} = \vec{x}$ è una causa effettiva di φ in (M, \vec{u}) se le seguenti condizioni sono verificate:

$AC2(a^m)$. C' è un insieme \vec{W} di variabili in \mathcal{V} e un setting \vec{x}' delle variabili in \vec{X} tale che se $(M, \vec{u}) \models \vec{W} = \vec{w}^*$, allora: $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}^*] \neg \varphi^{30}$

La definizione modificata incorpora l'idea che lo scenario che ci interessa quando vogliamo indicare che un evento (o equivalente relato causale) è causa di un altro è quello attuale. L'intuizione è che ciò che accade è il nostro *metro di misura* per la causazione. Di conseguenza, ora il setting \vec{W} permette solo valori attuali delle variabili, e solo questi possiamo, quindi, considerare quando cerchiamo di stabilire relazioni di causazione diretta.

Infatti, analizzando l'Esempio Lampadina attraverso questa definizione, è evidente che il contro-intuitivo risultato che $A = 1$ sia una causa di $L = 1$ non può essere ottenuto: considerando solamente un $\vec{W} = \vec{w}^*$ tale che $(M, \vec{u}) \models \vec{W} = \vec{w}^*$, modificare il valore di A non comporta alcuna variazione in L , cosa che, invece, B e C fanno.

Particolarità della **Definizione Causa effettiva^m** è che, insieme alla **Cond. AC₁**, implica le condizioni della **Definizione Causa effettiva^a** e anche la condizione della **Definizione Causa effettiva^{org}**, rendendole, di conseguenza, superflue.

Questa ultima definizione, pur essendo la più efficace nel difendersi da controesempi, implica delle conseguenze poco gradite. Si consideri il seguente esempio:

ESEMPIO 2.2.6 (Incendio^{mod})

All'interno di una foresta vi sono due piromani (P_1 e P_2). Il secondo dà fuoco alla foresta solo nel caso in cui anche il primo lo faccia. Nel frattempo, l'aria è carica di elettricità. Il clima è ideale per una tempesta di fulmini. Vi è la possibilità che uno di questi fulmini cada proprio sulla foresta, dandole fuoco e bruciandola. La foresta brucerà asimmetricamente ($FF = 2$) se solo uno dei piromani le dà fuoco. Invece, nel caso entrambi i piromani diano fuoco alla foresta, l'incendio sarà simmetrico. Anche nel caso in cui cadesse un fulmine (F), l'incendio sarebbe simmetrico. ■

Le equazioni che rappresentano la storia sono le seguenti, con network in **Figura 2.7**.

³⁰ Halpern, «A Modification of the Halpern-Pearl Definition of Causality» cit.; Halpern, *Actual Causality* cit.

$$P_2 = P_1$$

$$FF = \begin{cases} 2 & \text{se } P_1 \neq P_2 \\ 1 & \text{se } (F = 1 \wedge P_1 = P_2) \vee (P_1 = P_2 = 1) \\ 0 & \text{se } F = P_1 = P_2 = 0 \end{cases}$$

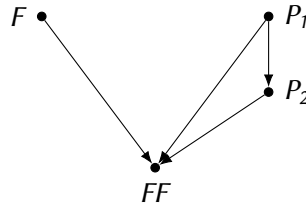


Figura 2.7: Grafo causale dell'Esempio Incendio^{mod}

Innanzitutto, quali sono le nostre intuizioni riguardo il contesto $\vec{u} = \{F = 1, P_1 = 1\}$, ovvero il caso in cui un fulmine cade e il primo piromane dà fuoco alla foresta (emulato, secondo equazioni, dal secondo piromane P_2)? Idealmente $F = 1$ dovrebbe essere una causa, perlomeno in questo contesto³¹. Il problema è che, nonostante quello che sostiene Halpern³², non risulta tale per nessuna delle definizioni proposte, se non per la prima.

Si immagini di selezionare come testimone $(\{P_1\}, 0, 0)$, ovvero lo scenario dove il primo piromane non dà fuoco alla foresta. Halpern sostiene che la Definizione Causa effettiva^a ci porta a concludere che $F = 1$ sia una causa in quanto la Cond. AC2(a^{org}) è rispettata: $(M, \vec{u}) \models [F \leftarrow 0, P_1 \leftarrow 0](FF = 0)$; inoltre, anche la Cond. AC2(b^a) è rispettata: $(M, \vec{u}) \models [F \leftarrow 1, P_1 \leftarrow 0](FF = 1)$. Sulla prima non vi è nulla da dire, ma la seconda affermazione (che la Cond. AC2(b^a) sia rispettata) mi sembra un grossolano errore: infatti, non tiene conto del fatto che $\vec{Z} = \{L, P_2, FF\}$ per definizione (è, infatti, il complemento di W rispetto a \mathcal{V}), e che, quindi, la Cond. AC2(b^a) richiede anche il rispetto della seguente condizione:

$$(M, \vec{u}) \models [L \leftarrow 0, P_1 \leftarrow 0, P_2 \leftarrow 1](FF = 1)$$

in quanto settare ogni sottoinsieme di \vec{Z} dovrebbe comunque risultare in φ , ovvero $FF = 1$. Ma, chiaramente, nella nostra situazione $(M, \vec{u}) \models [L \leftarrow 0, P_1 \leftarrow 0, P_2 \leftarrow 1](FF = 2)$, quindi la Cond. AC2(b^a) non è rispettata e, di conseguenza, $F = 1$ non è una causa nel contesto \vec{u} .

Si potrebbe pensare che la soluzione sia utilizzare l'insieme $\vec{W} = \{P_1, P_2\}$, ma anche questa strada è destinata a fallire per lo stesso motivo: secondo la Cond. AC2(b^a), dobbiamo testare la condizione per ogni sottoinsieme \vec{W}' di \vec{W} , ma questo significa che sarà necessario testare la condizione in una situazione in cui $P_1 \neq P_2$, portando, quindi, a $FF = 2$.

³¹ Questo secondo l'intuizione di Halpern, che qua seguo ai fini della discussione. Personalmente, questa intuizione non è così forte e sarei più propenso a considerare il contesto $\{F = 1, P_1 = 1\}$ come una situazione in cui vi è sovradeterminazione causale. Questa è una situazione in cui le intuizioni sono diverse e alcuni, ad esempio Lewis, suggeriscono di non dargli troppo peso (Lewis, «Postscripts to 'Causation'» cit.). Inoltre, nel Capitolo 5 argomenterò che questo, in realtà, è un caso di trumping (J. Schaffer, «Trumping Preemption», in Collins et al. (a cura di), *Causation and Counterfactuals* cit., pp. 59-74) e non una semplice sovradeterminazione e che, di conseguenza, l'intuizione che L sia da considerarsi una causa è scorretta.

³² Halpern, *Actual Causality* cit.: p.63.

Anche la **Cond. AC2(a^m)** purtroppo fallisce nel considerare $F = 1$ una causa. Questo perché P_1 è l'unica causa nel presente setting \vec{u} . Questa problematica, non risolvibile, a quanto mi risulta, attraverso una raffinazione della definizione, ci conduce a considerare soluzioni di diversa natura, che verranno analizzate nei capitoli successivi e, in particolare nel **Capitolo 3**.

2.3 CONCLUSIONE

Abbiamo analizzato varie definizioni di causalità della famiglia **HP**, anche attraverso dei test empirici. Il risultato è una tripla di definizioni di causa effettiva, ognuna con i suoi vantaggi e i suoi svantaggi. La definizione originale, in specie, soffre di un'esagerata permissività che non le consente di avere una buona performance, se non attraverso l'aggiunta di variabili nel modello causale al fine di specificarlo maggiormente.

Qualcosa di simile può essere sostenuto per la **Definizione Causa effettiva^a**: nonostante una correzione che la rende meno vulnerabile, come abbiamo visto anche questa definizione finisce per fallire e richiedere ulteriori specificazioni del modello al fine di produrre i risultati intuitivamente corretti.

Come abbiamo visto, la **Definizione Causa effettiva^m** è, certamente, in grado di catturare in maniera più precisa le nostre intuizioni, specie attraverso la specificazione che limita i *setting* valutabili a quelli attuali (ovvero con valori definiti esclusivamente dal contesto \vec{u}); ma presenta anch'essa delle problematiche, come quelle che l'**Esempio Incendio^{mod}** fanno emergere.

Sembra che ancora non si sia trovata una definizione di causa effettiva sufficientemente solida. La vena *definizionale* sembra essersi esaurita a questo punto. Si tratta, ora, di aprirsi a diverse prospettive al fine di affinare la discreta base che fin'ora è stata prodotta. Alcuni di questi tentativi verranno portati avanti nella **Discussione**.

DISCUSSIONE

3

IL PROBLEMA DELL'ISOMORFISMO E L'ARTE DI MODELLARE

L'utilizzo dei modelli causali ha permesso al discorso e al dibattito riguardante la causalità di raggiungere livelli di raffinatezza mai pensati; ha, inoltre, prodotto un notevole incremento nella pubblicazione di letteratura vicina a queste tematiche. I motivi sono stati esplorati nei **Preliminari** e sono, a mio parere, la precisione, la potenza, l'elegante formalismo e la capacità di affrontare i cosiddetti casi di ridondanza (*early e late preemption*, *sovradeterminazione*, ecc.). Ma, recentemente, sono emerse delle nuove problematiche e, in particolare, degli ostacoli tali da far preoccupare alcuni autori riguardo le potenzialità della ricerca filosofica inerente questo programma. Il problema forse più interessante è il cosiddetto **PI**¹, che è anche quello che ha suscitato più pessimismo.

Il **PI** consiste nella discrepanza tra giudizi causali di fronte a strutture e modelli causali identici, o isomorfi. Il fenomeno sembra condannare un'analisi della causalità in termini di soli modelli causali, in quanto suggerisce che i nostri giudizi causali siano informati non esclusivamente da strutture controfattuali, catturate dai modelli causali; e che, quindi, «[...] ci deve essere, per quanto riguarda la causalità, qualcosa di più delle sole equazioni strutturali»². Di conseguenza, gli sforzi degli ultimi anni si sono rivolti all'arricchimento dei modelli causali tramite distinzione tra stati (o comportamenti) di default e devianti³, misure di plausibilità^{4,5} o altre soluzioni⁶.

¹ P. Menzies, «Difference-making in Context», in **Collins et al.** (a cura di), *Causation and Counterfactuals* cit.; P. Menzies, «Causation in Context», in *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, a cura di H. Price e R. Corry, Oxford University Press, 2007; P. Menzies, «The Problem of Counterfactual Isomorphs», in **Beebe et al.**, *Making a Difference* cit., DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001); **Hitchcock**, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.; N. Hall, «Structural Equation and Causation», *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 132, n. 1 (2007), pp. 109-136; **Halpern**, «Defaults and Normality in Causal Structures» cit.; **Halpern**, *Actual Causality* cit.; **Halpern e Hitchcock**, «Graded causation and defaults» cit.; F. Huber, «Structural Equations and Beyond», *The Review of Symbolic Logic*, vol. 6, n. 4 (2013), pp. 709-732, DOI: [10.1017/s175502031300018x](https://doi.org/10.1017/s175502031300018x); G. Statham, «Causes as Deviations from the Normal: Recent Advances in the Philosophy of Causation», in *Perspective on Causation. Selected Papers from the Jerusalem 2017 Workshop*, a cura di E. A. B.-A. Siegal e N. Boneh, Jerusalem Studies in Philosophy and History of Science, Springer International Publishing, 2020, pp. 445-462, DOI: [10.1007/978-3-030-34308-8_14](https://doi.org/10.1007/978-3-030-34308-8_14).

² **Halpern**, «Defaults and Normality in Causal Structures» cit.: 204.

³ **Hall**, «Structural Equation and Causation» cit.; **Menzies**, «Difference-making in Context» cit.; **Menzies**, «Causation in Context» cit.; **Menzies**, «The Problem of Counterfactual Isomorphs» cit.

⁴ **Friedman e Halpern**, «Plausibility Measures: A User's Guide» cit.

⁵ Tra coloro che hanno impiegato misure di plausibilità, seppur con diversi apparati matematici, troviamo: **Halpern**, «Defaults and Normality in Causal Structures» cit.; **Halpern**, *Actual Causality* cit.; **Halpern e Hitchcock**, «Actual Causation and the Art of Modeling» cit.; **Halpern e Hitchcock**, «Graded causation and defaults» cit.; **Huber**, «Structural Equations and Beyond» cit.

⁶ **Blanchard e Schaffer**, «Cause without Default» cit.

3.1 IL PROBLEMA

Proseguiamo ora con l'analisi del **PI**. Questo sorge quando due modelli causali sono isomorfi, data la seguente definizione di grafi isomorfi:

DEFINIZIONE 3.1.1 (ISOMORFISMO TRA GRAFI DIREZIONATI)

Siano $G_1(V_1, E_1, \alpha_1, \omega_1)$ e $G_2(V_2, E_2, \alpha_2, \omega_2)$ due grafi direzionati con α_n e ω_n denotanti i vertici di partenza e di arrivo degli archi E_n . G_1 e G_2 si dicono isomorfi se esistono due funzioni biettive e una funzione di rinominazione dei vertici $\rho: V_1 \mapsto V_2$ e una funzione di trasformazione degli archi $\epsilon: E_1 \mapsto E_2$ tale che:

$$(i) \quad \forall e \in E_1: \rho(\alpha_1(e)) = \alpha_2(\epsilon(e))$$

$$(ii) \quad \forall e \in E_1: \rho(\omega_1(e)) = \omega_2(\epsilon(e))$$

DEFINIZIONE 3.1.2 (ISOMORFISMO TRA MODELLI CAUSALI (MENZIES))

Due modelli causali sono isomorfi sse i loro «pattern di indipendenza e dipendenza controfattuale sono identici, modulo l'uso di differenti variabili»⁷.

Vi è un problema nel definire esattamente da dove sorga il **PI**: Menzies⁸ parla di isomorfismo in situazioni tra strutture controfattuali e non a livello di struttura causale, che, anzi, diverge. Halpern e Hitchcock⁹, invece, sembrano sostenere che le strutture causali siano isomorfe e che i giudizi causali divergano. La mia intuizione è che entrambe queste definizioni siano parzialmente fuorvianti, in quanto suggeriscono che l'isomorfismo sia una relazione tra sistemi e non tra modelli di sistemi. Io tratterò l'isomorfismo, a mio parere correttamente, come una relazione tra modelli di sistemi causali. Di conseguenza, il **PI**, nella mia trattazione, in linea con quanto sostenuto da Blanchard e Schaffer¹⁰, risulterà un problema che sorge nel momento in cui vi sono giudizi causali differenti rispetto ad una medesima struttura causale.

Di seguito un esempio isomorfo all'Esempio Backup

ESEMPIO 3.1.1 (Avvelenamento Prudente)

Assistente Guardia del Corpo mette dell'innocuo antidoto nel caffè della vittima. A quel punto, Collega avvelena il caffè della vittima con un veleno normalmente letale ma inattivato dall'antidoto, al fine di far fare bella figura a Assistente e farlo promuovere. Collega non avrebbe avvelenato il caffè se Assistente non avesse versato l'antidoto prima. Vittima beve il caffè e sopravvive.¹¹ ■

Il modello causale per l'Esempio Avvelenamento Prudente è il seguente:

$$A = 1$$

$$V = \neg A$$

$$VS = A \vee V$$

dove con $A = 1$ indicheremo che l'antidoto è stato versato; di conseguenza, con $V = 1$, indicheremo il fatto che il veleno non è stato versato nel caffè¹² e con $VS = 1$ la sopravvivenza di Vittima.

⁷ Menzies, «The Problem of Counterfactual Isomorphs» cit.: 159.

⁸ Ivi.

⁹ Halpern e Hitchcock, «Graded causation and defaults» cit.

¹⁰ Blanchard e Schaffer, «Cause without Default» cit.

¹¹ Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.: 519.

¹² Questo accorgimento è comune in letteratura ed è fatto al fine di rendere l'isomorfismo ancora più evidente.

L'isomorfismo strutturale con l'**Esempio Backup** è rivelato attraverso la sostituzione di A con la variabile che rappresenta se Apprendista spara o meno (in questo caso A), di V con la variabile M , che rappresenta l'eventuale sparo di Mentore e di VS con SM . I modelli e le equazioni sono chiaramente identici, e lo sono anche i rapporti controfattuali tra variabili (al netto del significato della variabile V che ci appare anomalo a prima vista).

L'applicazione delle varie definizioni di causalità date finora ha il medesimo effetto per ognuna di queste: dobbiamo dire che l'avvelenamento è una causa effettiva della sopravvivenza di Vittima. Questo avviene nonostante il fatto che, abbiamo un'indipendenza tra A e VS . La dipendenza controfattuale è, però, instaurata mantenendo fisso il fatto che avvenga l'avvelenamento. Si noti che questi pattern di dipendenza e indipendenza controfattuale, come previsto dalla **Definizione Isomorfismo tra modelli causali (Menzies)**, sono i medesimi nell'**Esempio Backup**: la morte del soggetto ($SM = 1$) non dipende dall'eventualità che Apprendista faccia fuoco o meno. Se, invece, manteniamo fisso il fatto che Mentore non spara ($M = 0$), equivalente a mantenere fisso, nell'**Esempio Avvelenamento Prudente**, il fatto che avvenga l'avvelenamento, ecco apparire la dipendenza controfattuale tra il premere il grilletto di Apprendista e la morte di Vittima.

Nonostante tutte queste similitudini tra questi due casi, i nostri giudizi causali sono estremamente differenti: nell'**Esempio Backup** classificheremmo Apprendista $A = 1$ come una causa della morte della vittima. Nel caso dell'**Esempio Avvelenamento Prudente**, invece, la nostra intuizione è esattamente opposta: l'aggiunta dell'antidoto al caffè della vittima non è una causa della sua sopravvivenza. I motivi per cui siamo spinti ad un giudizio simile sono principalmente due:

- (i) L'esito, ovvero la sopravvivenza di Vittima, è quello che ci si aspettava in ogni caso, quindi non ha senso parlare di causazione in una situazione che non presenta alcunché di anomalo.
- (ii) La somministrazione dell'antidoto inizia un processo causale che esso stesso disinnesci: è solo perché la somministrazione avviene che avviene l'avvelenamento, che è evitato dalla somministrazione stessa. Questo è caso noto in letteratura come *short-circuit*¹³.

Il problema sembra essere che, per noi, le azioni di Apprendista e quelle di Assistente sono estremamente differenti; non così per lo scarno modello causale. Questo è il succo del **PI**.

Vediamo un altro caso classico, trattato da **Halpern e Hitchcock** e dovuto ad **Hitchcock**, ispirandosi ad un esempio proposto da **Hiddleston**¹⁴.

ESEMPIO 3.1.2 (Bogus Prevention)

Assassino ha un veleno da somministrare a Vittima, ma all'ultimo minuto cambia idea e non versa il veleno. Bodyguard mette l'antidoto nel caffè, che avrebbe neutralizzato il veleno, ce ne fosse stato. Vittima beve il caffè e sopravvive. ■

Il caso è rappresentato dal seguente modello causale:

$$A = \{0, 1\}$$

$$B = \{0, 1\}$$

$$VS = B \vee A^{15}$$

¹³ Hall, «Structural Equation and Causation» cit.: 120.

¹⁴ Halpern e Hitchcock, «Graded causation and defaults» cit.; Hiddleston, «Causal Powers» cit.; Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.

Nella situazione attuale $\vec{u} = \{A = 1, B = 1\}$, quindi Vittima sopravvive ($VS = 1$).

Le Definizioni 2.2.1 e 2.2.3 danno il controintuitivo esito che il Bodyguard sia una causa della sopravvivenza della Vittima. Non così la **Definizione Causa effettiva^m**, che non considera Bodyguard una causa, bensì parte di una causa (assieme al fatto che Assassino non metta il veleno). Nonostante si possa sostenere che questa non sia un verdetto così scorretto¹⁶, non è questo l'esito che ci piacerebbe avere. Sarebbe assolutamente preferibile che la nostra definizione fosse in grado di portarci un verdetto in linea con la nostra intuizione, secondo cui il fatto che Bodyguard metta l'antidoto nel caffè non ha alcun ruolo causale nella sopravvivenza della Vittima.

Sembra che si tratti di trovare una definizione in grado di catturare questo caso *extra*, come fatto per tanti altri esempi. Si consideri, ora, l'**Esempio Incendio Simultaneo**. In questo caso, eccezionalmente rispetto a tutte le altre definizioni di causalità incontrate, la **Definizione Causa effettiva^m** è in grado di catturare più o meno correttamente la situazione, affermando che il fulmine è parte di una causa, così come il lancio del fiammifero da parte del piromane, e che la congiunzione di questi due eventi è la vera causa effettiva dell'incendio. Questo, pur non essendo un esito che ci lascia particolarmente soddisfatti, è perfettamente logico e, sulla base delle informazioni fornite (ovvero nessuna distinzione di tempo, caratteristiche, aspetti, ecc. dell'incendio), assolutamente corretto.

Il problema sorge quando ci accorgiamo che questo caso è isomorfo all'**Esempio Bogus Prevention**. Si noti, infatti, il medesimo pattern di dipendenza, o meglio dire, indipendenza controfattuale. L'esito non dipende da nessuno dei casi presi indipendentemente l'uno dall'altro: non dipende da Assassino, in quanto Bodyguard potrebbe evitare l'avvelenamento; non dipende da Bodyguard, in quanto Assassino potrebbe non aver avvelenato il caffè. Il medesimo ragionamento si può fare nell'**Esempio Incendio Simultaneo**: l'incendio non dipende dal fulmine, come non dipende dal fiammifero, in quanto l'*evento avversario* può sempre essere attivo. Nel momento in cui, invece, fissiamo l'azione di Assassino o l'inazione di Bodyguard e rispettivamente fissiamo la non caduta del fulmine o il non lancio del fiammifero, la dipendenza controfattuale con l'*evento avversario* si manifesta.

Eccoci quindi giunti alla grossa problematica: tutti i nostri sforzi per produrre una definizione adeguata di causalità, che sia in grado di catturare i classici controesempi alle teorie controfattuali della causalità andati in fumo a causa di una strana *simmetria asimmetrica*: infatti, o rinunciamo ad alcuni risultati finora ottenuti, e iniziamo a considerare delle non-cause quelle che prima consideravamo cause, oppure troviamo delle modalità che siano in grado di discriminare queste situazioni, per così dire, differenti ma identiche, riuscendo ad individuare ciò che le rende diverse, pur mantenendo intatti e non modificando gli aspetti per cui sono simili. Si profila una situazione per cui siamo costretti ad affermare che i modelli causali mancano di qualcosa, qualcosa di fondamentale e che possa spiegare il perché di queste discrepanze di giudizi. Siamo, quindi, posti di fronte ad un bivio.

La prima strada è, chiaramente, quella dell'ammissione della sconfitta e, per questo, non perseguibile. Significherebbe non poter fornire una definizione adeguata di causalità basata sullo strumento che sta permettendo di fare passi da gigante in campi come la statistica, l'epidemiologia, l'inferenza e la scoperta causale e l'Intelligenza Artificiale, e negarne la possibilità di applicazione in campi legali, morali e filosofici.

¹⁵ Ho, nuovamente, invertito l'intuitiva assegnazione dei valori alla variabile *A*, ovvero indicando con 1 il non avvelenamento, al fine di far emergere in maniera più chiara l'isomorfismo con un altro esempio.

¹⁶ Personalmente, e in disaccordo con Halpern (Halpern, *Actual Causality* cit.), ritengo che questo esito sia corretto, seppur con una riserva, che verrà analizzata nel Capitolo 4, nel quale si sostiene il fatto che parte del motivo per cui non percepiamo l'antidoto come causa è che non è un meccanismo in azione.

La seconda strada, quella più promettente, è quella di arricchire i nostri modelli in qualche modo, al fine di fornire alle nostre teorie li strumenti che permettano loro di discriminare questi casi problematici, nel frattempo mantenendo intatta la loro capacità di individuare la genuina causa effettiva come finora dimostrato. Sarà necessario, però, non stravolgere la cornice metodologica, al fine di non creare discrepanze interne al campo della trattazione a modelli causali. Una discrepanza di questo genere sarebbe, infatti, come correttamente notato da **Blanchard e Schaffer**¹⁷ fortemente penalizzante e quasi deleteria, in quanto si troverebbe a dover rifiutare tutto un apparato che si è dimostrato, negli ultimi anni, senza alcun rivale in quanto a precisione e potenzialità di sviluppo. È proprio nella direzione di un arricchimento dell'apparato metodologico, che ne lasci intatto il nocciolo, ma lo fornisca di nuovi bracci meccanici in grado di effettuare nuove operazioni, che ci muoviamo ora, andando ad analizzare alcune delle soluzioni più interessanti al **PI** proposte dai teorici della causalità e dei modelli causali.

3.2 LE SOLUZIONI

Come già anticipato, tre sono le possibili vie per risolvere il **PI** che possono essere percorse: gli approcci basati sulla distinzione *default/deviante*, gli approcci basati sulle misure di plausibilità, oppure la terza via. Di seguito, esploreremo le varie proposte.

3.2.1 *Default e deviazioni*

Una delle più comuni modalità attraverso le quali affrontare il **PI** è la distinzione tra stati di default e stati devianti. Il più influente tra gli approcci di questo tipo è stato proposto da **Menzies** in una serie di lavori incentrati sul tema e non¹⁸. **Menzies**, in tutti questi scritti, parla della causa attuale come *difference-maker*, 'fare la differenza'. Questa nozione viene definita in termini prettamente controfattuali, seguendo il paradigma lewisiano, ma con una differenza ragguardevole. In particolare, **Menzies** critica la proposta di **Lewis** sulla base del fatto che definisce la causa effettiva come una relazione assoluta e indipendente dal contesto, una strada metafisica che, a sua opinione, è necessario abbandonare al fine di analizzare correttamente la causa come *difference-maker*.

L'approccio di **Menzies** parte dal fatto che è la stessa valutazione dei controfattuali a risultare errata nella proposta di **Lewis**, di conseguenza, propone una differente semantica dei controfattuali, in cui la distinzione default/deviante svolge il compito di centrare il mondo corretto rispetto al quale valutare i controfattuali.

Nella fattispecie, l'approccio di **Lewis** è quello di centrare la sfera di mondi simili sul mondo attuale, in modo che il mondo rispetto al quale avvengono tutti i confronti necessari alla valutazione dei controfattuali sia il mondo attuale. Differente è l'approccio di **Menzies**, che ritiene che il mondo rispetto al quale valutare i controfattuali non sia il mondo attuale, bensì un altro mondo. L'idea di fondo è che vi sia un parametro contestuale che determina il modo in cui valutiamo i controfattuali¹⁹. Questo parametro viene impostato dal contesto e può variare. Il concetto di causa diventa, così, dipendente dal contesto. Per essere precisi: la valutazione dei controfattuali, su cui il concetto di

¹⁷ Blanchard e Schaffer, «Cause without Default» cit.: 178.

¹⁸ Menzies, «Difference-making in Context» cit.; Menzies, «Causation in Context» cit.; Menzies, «The Problem of Counterfactual Isomorphs» cit.

¹⁹ Menzies, «Difference-making in Context» cit.

causa effettiva (o *difference-maker*) si basa, diventa contestuale. Si badi bene, il *concetto* di causa effettiva è contestuale. **Menzies** non fa proposte antirealiste o sostiene che la causazione sia una relazione soggettiva o mentale, bensì afferma che la mera struttura controfattuale di una situazione non è sufficiente a determinare i valori di verità dei giudizi causali inerenti la situazione.

L'intuizione chiave, che guida la nostra valutazione causale è, per **Menzies**, la seguente:

Nell'applicare il concetto causale in una particolare situazione, astraiamo e generalizziamo interpretando la situazione in termini del modo in cui un particolare sistema si comporta. Il secondo elemento è che tipicamente supponiamo *che un sistema di un certo tipo, se lasciato a se stesso, segue un corso di evoluzione che è 'normale' o 'naturale'* per sistemi di quel tipo quando non sono soggetti ad interferenza esterna. Il terzo ed ultimo elemento è che, quando il sistema esemplificato nella particolare situazione ha *deviato dal suo normale corso di evoluzione, cerchiamo qualcosa che ha fatto la differenza*[...]²⁰

Questa idea suggerisce il passo successivo: in una cornice a modelli causali, quello che usualmente si fa, è mantenere lo stato effettivo delle variabili come la *baseline* sulla quale basare i propri giudizi e controfattuali. **Menzies** rovescia completamente questa idea, e propone, invece, di valutare i nostri controfattuali e emettere i nostri giudizi causali su di uno stato che rappresenti la vera *baseline* del sistema sotto indagine: lo stato di default.

DEFINIZIONE 3.2.1 (MODELLO CAUSALE DI DEFAULT)

*Un modello causale di default è un modello causale nel quale le variabili esogene sono impostate ai loro valori di default*²¹.

A questo punto, si tratta di introdurre la nozione di *dipendenza controfattuale a condizioni ideali*, al fine di catturare il fatto che «i nostri giudizi causali sono sensibili a considerazioni che vanno oltre le equazioni strutturali-ovvero, considerazioni riguardo le condizioni ideali»²². Questa nozione è semplicemente una dipendenza controfattuale che sussiste in condizioni ideali. Ecco svelato quello che sta sotto al **PI**: le nostre valutazioni sono differenti perché differenti sono le condizioni ideali nelle quali dobbiamo valutare la dipendenza controfattuale. Si tratta, ora, di definire quali sono queste condizioni ideali. Il primo passo è definire una condizione *abilitante*, che è, grossomodo, una condizione che permette l'instaurarsi di una connessione causale tra una potenziale causa effettiva e un effetto effettivo. Questa condizione deve essere, naturalmente, presente nello stato attuale del sistema e deve fare la differenza per la presenza dell'effetto. *En passant* si fa notare che questa definizione può facilmente essere oggetto di critiche in quanto il limite che separa una condizione *abilitante* da una causa vera e propria è estremamente sottile e, a volte, non discriminabile: questo avviene nei casi in cui una congiunzione di fattori sono considerabili una causa.

DEFINIZIONE 3.2.2 (CONDIZIONE ABILITANTE)

Una condizione $V_j = v_j$ è una condizione abilitante per una possibile connessione causale tra la causa attuale $X = x$ e $Y = y$ nel modello M se (i) V_j sfocia in un percorso che da X porta a Y e $(M, \vec{u}) \models V_j = v_j$; e (ii) l'imposizione del proprio valore effettivo alle variabili che sfociano nel percorso $V^ = v^*$, rende i seguenti controfattuali veri:*

²⁰ **Menzies**, «Causation in Context» cit.: enfasi mia.

²¹ *Ivi*.

²² **Menzies**, «The Problem of Counterfactual Isomorphs» cit.: 155.

$$(a) (X = x \wedge V_j = v_j \wedge V^* = v^*) \Box \rightarrow Y = y$$

$$(b) (X = x \wedge V_j \neq v_j \wedge V^* = v^*) \Box \rightarrow Y \neq y$$

Arrivati a questo punto, siamo in grado di definire la nostra nozione di *causa come difference-maker*.

DEFINIZIONE 3.2.3 (MENZIES DIFFERENCE-MAKER)

C fa una differenza rispetto a E in una situazione attuale rispetto al modello M sse $C \Box \rightarrow_M E$ e $\neg C \Box \rightarrow_M \neg E$ ²³.

dove, con il simbolo $\Box \rightarrow_M$ si denota il controfattuale valutato in una situazione ideale, ovvero in cui le condizioni abilitanti, se attive, sono mantenute attive, altrimenti vengono impostate in modo da rappresentare l'assenza dell'evento (ovvero il suo *default*)²⁴.

Questa definizione è una raffinazione di quelle proposte precedentemente in quanto si limita ad effettuare modifiche sul sistema solo in quelle variabili che potrebbero essere rilevanti per la determinazione della relazione di causa attuale. La semantica dei controfattuali proposta da Menzies²⁵ può essere lasciata intatta senza intaccare questa seconda proposta sulla causa attuale.

Valutiamo ora i risultati di questo approccio nei confronti del **PI**. Si noti che esso ci fornisce una elegante soluzione al problema dell'**Esempio Bogus Prevention**: lo stato intrinseco di questo modello, ovvero il più normale, è quello in cui Assassino non avvelena il caffè. Di conseguenza, l'azione di Bodyguard non fa alcuna differenza rispetto alla sopravvivenza di Vittima e, di conseguenza, non va considerata come una causa. Si prenda ora l'isomorfo, ma intuitivamente causalmente differente **Incendio Simultaneo**. Qui lo stato intrinseco del sistema è quello in cui non vi sono né piromani, né fulmini. In questa situazione, rispetto a questo *modello di default* entrambi gli eventi sono *difference-makers* e, quindi, vanno considerati come cause.

L'intuizione di fondo di Menzies è la seguente: vi è un operatore contestuale che determina lo stato di default rispetto al quale valutare i controfattuali e, quindi, la relazione di causa attuale è, perlomeno parzialmente, *sensibile al contesto*. Non si tratta, come sostenuto da Hitchcock²⁶, di un'ambiguità del concetto di causalità, ma di una caratteristica che emerge dal di fuori rispetto alla relazione intrinseca rappresentata dalla struttura causale nella sua crudezza, e che fa uso di un riferimento contestuale per *sbrigliare la matassa*.

Compiute queste osservazioni, passiamo alla trattazione delle misure di plausibilità.

3.2.2 Le misure di plausibilità

Chi propone un approccio basato sulle misure di plausibilità non ha intenzione di modificare le condizioni di verità dei controfattuali, bensì vuole emendare i modelli causali attraverso una loro estensione conservativa. In questa maniera, si tenta di incorporare una sorta di distinzione tra stati di default e stati devianti attraverso un ordine (parziale o totale, a seconda degli autori) sui mondi, definiti secondo la **Definizione 1.1.4**.

²³ Menzies, «Difference-making in Context» cit.:166.

²⁴ Menzies non è estremamente preciso in questo contesto in quanto, in alcuni casi, il default di un evento è la sua presenza (ad esempio la presenza di ossigeno nell'aria è un default).

²⁵ Ivi.

²⁶ C. Hitchcock, «Of Humean Bondage», *The British Journal for the Philosophy of Science*, vol. 54, n. 1 (2003), pp. 1-25, DOI: 10.1093/bjps/54.1.1.

Questi account non vanno a modificare la semantica dei controfattuali ma cercano di dare un *twist* psicologico al **PI**, trattandolo come una discrepanza di giudizi riguardo alla medesima struttura causale, originata da un differente modo di ordinare la classe di contrasto sulla quale i giudizi si basano²⁷. Halpern sostiene che «la soluzione involve l'assumere che un agente ha, oltre ad una teoria della causalità (come modellata dalle equazioni strutturali), una teoria della “normalità” o “tipicalità”»²⁸.

La trattazione introduttiva della proposta si baserà su *Actual Causality*, dopodiché verranno presentate rapidamente alcuni differenti approcci. Si noti che non sempre gli autori parlano di misure di plausibilità, e a volte, come nel caso di Halpern, fanno riferimento a ordini parziali, altre volte a funzioni di ranking²⁹. Tratterò gli approcci in maniera equivalente e userò funzioni di ranking e misure di plausibilità interscambiabilmente: pur essendo la prima parte dell'insieme delle seconde, la distinzione è, in questo caso, immateriale.

3.2.2.1 Modelli causali estesi secondo Halpern

Halpern³⁰ ritiene che vi siano tre concetti strettamente imparentati: *default*, *tipicalità* e *normalità*. Il primo descrive «assunzioni riguardo a cosa succede, [...], quando non viene data nessuna informazione aggiuntiva». Se ci viene detto che Macchia è un gatto, una delle inferenze naturali che abbiamo è che abbia il pelo. Naturalmente, questa inferenza può venir annullata nel momento in cui ci vengono date delle informazioni aggiuntive: se ci viene detto che Macchia è un Canadian Sphynx, allora sappiamo che la nostra inferenza è scorretta.

Il secondo concetto, la *tipicalità*, ha a che vedere con quelle che sono le caratteristiche di un “tipo”. «Nonostante non tutti gli uccelli volino, volare è qualcosa che caratteristicamente associamo agli uccelli»³¹. Il tipo uccello ha la caratteristica di essere in grado di volare: quando pensiamo ad un uccello, pensiamo a qualcosa che vola.

Il terzo concetto, la *normalità*, è un concetto complesso. Ha a che fare, sommariamente, con norme societarie, di funzionamento, morali e statistiche³². Vengono inclusi, quindi, aspetti sia descrittivi, che prescrittivi³³. Halpern propone di utilizzare i primi due riguardo a variabili, ed di trattare il terzo come una relazione tra mondi.

Viene definito più normale rispetto ad un altro un mondo in cui le variabili assumono valori in linea con le equazioni strutturali del modello in numero maggiore rispetto ad un altro. Non esiste un unico ordine di normalità, differenti soggetti possono avere diverse misure di plausibilità riguardo alla normalità. Non ci soffermeremo se questo punto nel corso della presente discussione.

²⁷ Questo non vale, in realtà, per la proposta di Huber, che sì, si basa sulle funzioni di *ranking*, ma propone un'estensione controfattuale della funzione di ranking, al fine di fornire una semantica dei controfattuali. L'approccio permette di risolvere alcune problematiche degli altri account, ad esempio Halpern e Hitchcock, «Graded causation and defaults» cit.: nota 18. Si vedano Huber, «Structural Equations and Beyond» cit. e J. Y. Halpern e C. Hitchcock, «Compact Representations of Extended Causal Models», *Cognitive Science*, vol. 37, n. 6 (2013), pp. 986-1010, DOI: 10.1111/cogs.12059: sec. 5 per dettagli.

²⁸ Halpern, «Defaults and Normality in Causal Structures» cit.

²⁹ *Ivi*.

³⁰ Halpern, «Defaults and Normality in Causal Structures» cit.; Halpern e Hitchcock, «Graded causation and defaults» cit.; Halpern, *Actual Causality* cit.

³¹ Halpern, *Actual Causality* cit.:78.

³² Si vedano: Knobe e Fraser, «Causal Judgement and Moral Judgement: Two Experiments» cit.; Hitchcock e Knobe, «Cause and Norm» cit.; F. Cushman et al., «Moral appraisals affect doing/allowing judgments», *Cognition*, vol. 108, n. 1 (2008), pp. 281-289, DOI: 10.1016/j.cognition.2008.02.005; Halpern e Hitchcock, «Graded causation and defaults» cit.; T. F. Icard et al., «Normality and actual causal strength», *Cognition*, vol. 161 (2017), pp. 80-93, DOI: 10.1016/j.cognition.2017.01.010.

³³ Statham, «Causes as Deviations from the Normal: Recent Advances in the Philosophy of Causation» cit.:448.

Halpern definisce un modello causale esteso come una tupla $M = (S, \mathcal{F}, \geq)$, dove \geq rappresenta la relazione di normalità (se $s \geq s'$ s è almeno normale quanto s') e viene assunta essere un preordine parziale. Questo implica che:

- La relazione è riflessiva: $s \geq s$
- La relazione è transitiva: se $s \geq s'$ e $s' \geq s''$, allora $s \geq s''$
- La relazione non è necessariamente antisimmetrica, ovvero può essere il caso che $s \geq s'$, $s' \geq s$ e $s \neq s'$

Halpern assume, inoltre, che la relazione non sia totale.

La definizione di causa attuale viene, quindi, modificata nel seguente modo:

DEFINIZIONE 3.2.4 (CAUSA ATTUALE⁺)

$\vec{X} = \vec{x}$ è una causa effettiva di φ in (M, \vec{u}) se le seguenti condizioni sono verificate:

$AC_2^+(a)$ Vi è una partizione di \mathcal{V} in due sottoinsiemi distinti \vec{Z} e \vec{W} con $\vec{X} \subseteq \vec{Z}$ e un setting \vec{x}' e \vec{w} , rispettivamente, delle variabili in \vec{X} e \vec{W} , tale che: $s_{\vec{X}=\vec{x}', \vec{W}=\vec{w}} \geq s_{\vec{u}}$ e $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}] \neg \varphi$

Purtroppo, anche incorporando questa funzione di ranking, il **PI** permane. Si consideri, infatti, l'isomorfismo tra l'**Esempio Backup** e l'**Esempio Avvelenamento Prudente**. Riportati in **Tabella 3.1** abbiamo l'ordinamento tra mondi nei due esempi in questione³⁴.

A	M	VS
0	0	0
1	0	1
0	1	1
1	1	1

(a) Ordine di normalità dei mondi nell'**Esempio Backup**

A	V	VS
0	1	1
1	1	1
0	0	0
1	0	1

(b) Ordine di normalità dei mondi nell'**Esempio Avvelenamento Prudente**

Tabella 3.1: Ordini di normalità nel **PI**

Come risulta evidentemente dalla **Tabella 3.1**, ora, grazie ai modelli estesi, corredati con la funzione di ranking, non siamo più di fronte ad una situazione di stretto isomorfismo, in quanto l'ordine tra mondi è differente nei due casi. Si tratta, quindi, ora, di mostrare che $A = 1$ non è una causa effettiva di $VS = 1$. Vediamo come procedere.

Innanzitutto, la scelta più naturale per il mondo di default è quello in cui né Assistente, né Collega modificano la composizione chimica del caffè: $\vec{w} = \{A = 0, V = 0, VS = 1\}$. Appliciamo, ora, la nuova **Definizione Causa attuale⁺**. Selezioniamo come testimone per la potenziale causa effettiva $A = 1$ la seguente tripla: $(\{V\}, V = 0, A = 0)$. L'esito che desideriamo è che questo testimone, ed il mondo da esso generato, risulti meno normale del mondo $w = \{A = 1, V = 0, VS = 1\}$, in modo da non permetterci di definire $A = 1$ come causa, in modo da allinearsi con le nostre intuizioni. Il problema è che questo questo testimone non è meno normale, come si evince dalla **Tabella 3.1b**. Questo cattura l'intuitivo ragionamento che, se il valore normale per A è 0, e il valore normale per V è 1; allora un mondo in cui solo una di queste variabili assume un valore anomalo è più normale

³⁴ Sono stati omessi alcuni mondi in quanto non rilevanti nella trattazione che qui si svolge.

di quello in cui entrambe assumono valori anomali. Se questo è il ragionamento a cui ci appoggiamo, allora la definizione non è sufficiente.

Ma qui, Halpern e Hitchcock fanno notare che questo è un «sottile errore nella maniera di pensare all'ordine di normalità»³⁵. Infatti, si noti che la variabile V non è fissata da \vec{u} , bensì da A , una variabile interna al modello. Questo significa che quando valutiamo la normalità, è necessario valutare la normalità della relazione tra A e V . E, in questo modello, non assistiamo alla più normale relazione tra queste due variabili, rappresentate nella Tabella 3.2.

N°	$P = ?$
1	1
2	A
3	0

Tabella 3.2: Ordine di normalità delle relazioni tra A e V

Questo approccio equivale a misurare la normalità delle intenzioni di Collega, considerando massimamente normale il non avere l'intenzione di avvelenare il caffè, seguito dal non complotare con Assistente, e, infine, trattare l'intenzione omicida come la meno normale³⁶. Queste considerazioni ci permettono di rivalutare in maniera corretta il modello esteso e vedere che, come i nostri giudizi prevedono, $A = 1$ non è una causa effettiva di $VS = 1$. Infatti, il mondo attuale $w_a = \{A = 1, P = 0, VS = 1\}$ è caratterizzato da un valore deviante (quello di A) ed un comportamento di una variabile non massimamente tipico ($P = A$). Il mondo prodotto dal testimone, ovvero $w_w = \{A = 0, V = 0, VS = 0\}$, è caratterizzato da un valore di default (si tratta nuovamente di A), ed un comportamento di una variabile massimamente atipico ($P = 1$ indipendentemente da tutto). Le opzioni sono due: o (i) i due mondi sono incomparabili; o (ii) il mondo del testimone è meno normale del mondo attuale³⁷. In entrambi i casi, $A = 1$ non è una causa di $VS = 1$. Di conseguenza, il **PI** viene risolto attraverso qualche attenzione dedicata e maggiore rigore.

Un altro risultato che viene ottenuto senza alcun prezzo dall'integrazione delle misure di plausibilità in un account **HP**, e la possibilità di differenziare cause e condizioni di sfondo molto elegantemente³⁸. Il problema è che «quali eventi contano come cause di un altro evento, [...], è una materia drammaticamente sensibile al contesto»^{39,40}.

Si consideri il caso di un incendio avvenuto a causa di un cortocircuito. Naturalmente, in assenza di ossigeno, l'incendio non sarebbe avvenuto; eppure nessuno considererebbe la presenza di ossigeno come una causa. Questa è la discrepanza di cui gli account basati sulle misure di plausibilità ci permettono di rendere conto.

³⁵ Halpern e Hitchcock, «Graded causation and defaults» cit.:451.

³⁶ *Ibidem*.

³⁷ Si noti che, ad essere rigorosi, questa è un'interpretazione scorretta, e i due mondi andrebbero considerati incomparabili. Viene qui presentata questa alternativa per esautività.

³⁸ Si veda: H. L. A. Hart e A. M. Honoré, *Causation in the law*, Oxford University Press, 1959:30-41.

³⁹ E. Swanson, «Lessons from the Context Sensitivity of Causal Talk», *The Journal of Philosophy*, vol. 107, n. 5 (2010), pp. 221-242, <http://www.jstor.org/stable/25764443>:222.

⁴⁰ Sull'argomento si veda anche: B. C. van Fraassen, *The Scientific Image*, Oxford, England: Oxford University Press, 1980:112-130.

Si prenda il seguente modello:

$$\begin{aligned}C &= \{1, 0\} \\ O &= \{1, 0\} \\ F &= M \wedge O\end{aligned}$$

dove, come al solito, denotiamo con 1 la presenza del evento/fattore che la variabile rappresenta e con 0 la sua assenza.

In questo esempio, $F = 1$ dipende controfattualmente da $C = 1$ e da $O = 1$, di conseguenza sono entrambe cause secondo le definizioni HP. Ma ora, analizzando l'ordine di normalità, notiamo la seguente caratteristica:

$$(C = 0, O = 1, F = 0) \geq (C = 1, O = 1, F = 1) \geq (C = 1, O = 0, F = 0) \geq$$

Quello che emerge è che il mondo in cui non avviene il cortocircuito, che è il testimone che permette di determinare che $C = 1$ è una causa di $F = 1$, è almeno normale quanto il mondo effettivo. Il motivo di questa equi-normalità (o maggiore normalità) è il fatto che la presenza di ossigeno ($O = 1$) è la situazione di default. Per quanto riguarda il testimone che ci permette di determinare l'ossigeno come causa del fuoco, il discorso è radicalmente differente. Infatti, il mondo in cui si verifica il cortocircuito e in cui è assente l'ossigeno è meno normale del mondo effettivo. Questo spiega il motivo per cui trattiamo le cause di sfondo come tali: essendo valori di default, una loro modifica rende il testimone meno normale del mondo attuale.

Si noti che questo trattamento ha il grande vantaggio di essere in grado di rendere conto di una moltitudine di giudizi contestuali che discriminano tra cause e condizioni di sfondo. Questo è dovuto al fatto che l'ordine di normalità è dipendente dal contesto. Questo permette di considerare valori di variabili che sono default in alcuni casi estremamente anormali in altri: se ci trovassimo in una stazione spaziale e volessimo determinare la causa di un incendio, il valore di default di O sarebbe differente: non è, infatti, da dare per scontata la presenza di ossigeno in una tale situazione e, anzi, potrebbe trattarsi di un'anomalia (ad esempio, una perdita dal serbatoio di ossigeno).

Durante questa sezione, abbiamo potuto apprezzare la grande potenza di queste proposte basate sulle misure di plausibilità. Quella di Halpern, come già detto, non è l'unica proposta in questo senso. Halpern e Pearl, ad esempio estendono i modelli con una sorta di funzione di ranking che *permette o vieta* particolari valori per le variabili endogene, sulla base del fatto che consideriamo alcuni scenari estremamente improbabili⁴¹. Successivamente, Halpern ha proposto di indicizzare la relazione di normalità su contesti invece che su mondi⁴².

Non tutti, però, sono soddisfatti dall'inserire quelle che considerano considerazioni psicologiche ed euristiche (gli ordini di normalità e i default) all'interno di un account che si fregia di un rigore matematico e scientifico quasi senza pari nel campo della riflessione sulla causalità. Temendo che l'estensione, conservativa o meno, dei modelli con delle misure di plausibilità possa in un certo qual modo inquinare la purezza del macchinario metodologico dei modelli causali, Blanchard e Schaffer

⁴¹ Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.:869-870.

⁴² Halpern, *Actual Causality* cit.:sec. 3.5.

hanno cercato di argomentare, credo anche parzialmente ispirati dall'idea di Halpern e Pearl⁴³ citata poco sopra, per una soluzione alternativa, basata sulla raffinazione dei nostri modelli. Stiamo per entrare in quel dominio che viene chiamato "arte di modellare".

3.3 MODELLI ADATTI E L'ARTE DI MODELLARLI

Le soluzioni al **PI** affrontate sin ora si basano tutte su una qualche distinzione basata sul contesto e su distinzioni tra stati normali o di default. Per quanto questi approcci siano estremamente efficaci, vi è qualcosa in queste estensioni che lascia alcuni non soddisfatti. La distinzione è, infatti, certamente utile quando si tratta di *giudizi inerenti la causalità*, ma incorporare questi strumenti contestuale e pericolosamente carichi di assunzioni pragmatiche, in quella che è una soluzione elegantemente matematica e che si fregia di grande rigore, fa storcere ad alcuni il naso. Come argomentano Blanchard e Schaffer⁴⁴, sembra che facendo questo passo si rischi di iniettare concetti che pertengono maggiormente all'aspetto *psicologico-cognitivo* della causalità, rispetto all'analisi concettuale e metafisica.

In particolare, una soluzione migliore ad alcuni di quei casi che sono stati presentati in letteratura come afferenti al **PI**, è affermare che in realtà non si tratta di un problema genuino⁴⁵. La vera problematica è, infatti, di tipo rappresentativo: i nostri modelli causali sono, per l'appunto, modelli, rappresentazioni matematiche di sistemi, delle loro relazioni e rapporti, delle loro strutture. Quand'è che queste strutture vengono catturate adeguatamente dai nostri modelli? Questa è una domanda a cui rispondere non è facile. Benché si tratti di una domanda inerente alla questione della rappresentazione in generale, i nostri scopi e obiettivi sono, in questa sede, decisamente più ristretti. Riformuleremo, quindi, la domanda in questi termini: *che cosa rende un modello causale adeguato alla situazione che vuole rappresentare?*

In letteratura, le risposte non sono moltissime. Quella dell'adeguatezza dei modelli è, infatti, una questione generalmente lasciata in disparte, anzi, Halpern e Pearl ci dicono, addirittura, che «dibattere su quale di due (o più) modelli del mondo sia una migliore rappresentazione [...] è il tipo di dibattito che muove dibattiti informali (e legali) tutte le volte». Abbiamo qualche linea guida su cui basarci, ma, in generale «cosa costituisca un modello appropriato è una affare intricato, più una questione di arte che di scienza»⁴⁶. Vediamo cosa ha da offrirci la letteratura:

1. Le *equazioni strutturali* non devono implicare controfattuali falsi.
2. Le *equazioni strutturali* non devono rappresentare relazioni di dipendenza controfattuale tra eventi che non sono distinti.
3. \mathcal{V} non deve contenere variabili i cui valori corrispondono a possibilità che consideriamo essere troppo remote^{47,48}.

⁴³ Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.

⁴⁴ Blanchard e Schaffer, «Cause without Default» cit.

⁴⁵ *Ivi*.

⁴⁶ Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.: 503.

⁴⁷ Hitchcock, «The Intransitivity of Causation Revealed in Equations and Graphs» cit.: 287.

⁴⁸ Si noti che questa condizione è simile a quella stabilita da Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.

4. Valori di differenti variabili non dovrebbero corrispondere ad eventi/proprietà che hanno una relazione logica⁴⁹.
5. Differenti valori della medesima variabile devono rappresentare eventi/proprietà mutualmente esclusivi⁵⁰.
6. I valori delle variabili devono rappresentare caratterizzazioni intrinseche⁵¹.
7. Il modello deve includere variabili sufficienti a catturare la struttura essenziale della situazione modellata⁵².
8. Aggiungere variabili non dovrebbe invertire i giudizi causali⁵³.

Tutte queste regole sembrano essere sufficientemente chiare, esclusa la **Cond. 7**. Personalmente, nell'interpretazione di questa direttiva mi appoggio all'analisi proposta da **Paul**⁵⁴, ovvero identificando come componenti di assoluta rilevanza a livello strutturale gli aspetti della situazione, ma senza trattare di per sé questi aspetti come relata causali. **Paul** parla degli aspetti in termini di tropi o di istanze di proprietà⁵⁵. Tralasciando l'impegno metafisico, che in questa sede non ci concerne, ritengo che la discussione svolta da **Paul** catturi perfettamente quelli che sono le componenti basilari della struttura da rappresentare attraverso un modello causale. Di che cosa si intenda parlando di struttura *intrinseca* di un sistema (o, *insieme di oggetti*) si è occupato **Menzies** nel suo «Difference-making in Context», a cui rimando⁵⁶ per una discussione approfondita.

Tornando a **Blanchard e Schaffer**: l'argomento degli autori parte con la constatazione che, al fine di individuare correttamente le cause attuali, è necessario che il modello sul quale ci basiamo sia appropriato. Questo requisito, per noi che, al momento, stiamo seguendo le definizioni della famiglia **HP**, risulta lapalissiano: queste definizioni, infatti, come già affermato, sono definizioni relative ad un modello⁵⁷; di conseguenza è evidente che, se il sottostante modello è scorretto, è facile che anche le nostre conclusioni ereditino questa sfortunata caratteristica.

Il passo successivo, come già anticipato, è quello di affermare che alcuni di questi presunti casi afferenti al **PI** vengano caratterizzati attraverso modelli scorretti. Vediamo l'**Esempio Avvelenamento Prudente**. Il modello con cui abbiamo sin ora rappresentato la situazione è tradotto in grafo in **Figura 3.1**.

Questo modello tralascia di rappresentare un fattore fondamentale ai fini della resa corretta della situazione, violando la **Cond. 7**. Si tratta, quindi, di correggere questo difetto. Di seguito propongo un esempio raffinato, che mette in luce l'aspetto fondamentale che viene tralasciato.

49 Hitchcock aggiunge «logica o concettuale»: Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.: 502.

50 Halpern e Hitchcock, «Actual Causation and the Art of Modeling» cit.: 397.

51 Blanchard e Schaffer, «Cause without Default» cit.: 183.

52 Hitchcock, «Prevention, Preemption, and the Principle of Sufficient Reason» cit.: 503.

53 Si noti che questa condizione è, in principio, problematica se non corredata da un modello esteso con funzione di ranking, come dimostra J. Y. Halpern, «Appropriate Causal Models and the Stability of Causation», *The Review of Symbolic Logic*, vol. 9, n. 1 (2016), pp. 76-102, DOI: [10.1017/s1755020315000246](https://doi.org/10.1017/s1755020315000246); Teorema 6.1. Non risulta, comunque, problematica nella pratica: Halpern, *Actual Causality* cit.: 96.

54 Paul, «Aspect Causation» cit.

55 Riguardo ai tropi, alle istanze di proprietà e alle differenze tra questi, si veda: M. Carrara et al., *Introduzione alla metafisica contemporanea*, il Mulino, 2021: cap.VII.

56 Menzies, «Difference-making in Context» cit.:155-158.

57 Si noti infatti che vi è sempre il termine $(M, \vec{u}) \models$ all'inizio di ogni clausola.

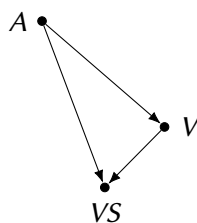


Figura 3.1: Grafo associato all'**Esempio Avvelenamento Prudente**

ESEMPIO 3.3.1 (Avvelenamento Prudente 2)

Tutto si svolge come da copione dell'**Esempio Avvelenamento Prudente**, fino a quando il veleno non è versato nel caffè. A quel punto, avviene una reazione chimica nel caffè R : l'antidoto, grazie alla propria struttura tridimensionale, cattura le molecole di veleno al suo interno, rendendole inassorbibili. Vittima beve il caffè e sopravvive. ■

Analizzando il modello causale associato a questa storia, ci si presenta un modello assai differente:

$$A = \{0, 1\}$$

$$V = \{0, 1\}$$

$$R = A \wedge \neg V$$

$$VS = R \vee \neg V$$

Il grafo che rappresenta la situazione è il seguente:

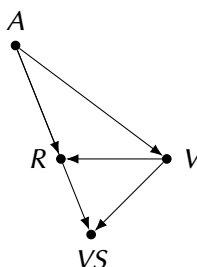


Figura 3.2: Grafo associato all'**Esempio Avvelenamento Prudente 2**

In questo caso ci troviamo di fronte ad una situazione estremamente differente. Innanzitutto, il modello non è più isomorfo, e così, di conseguenza, i pattern di dipendenza controfattuale. Ciò che ci motiva ad arricchire il modello in questa maniera è l'idea che, nel modello originale, non è catturato uno degli aspetti fondamentali dell'azione compiuta da Assistente: essa avvia una serie di azioni che, potenzialmente, mettono a rischio la sopravvivenza di vittima. Vorremmo che il nostro modello fosse in grado di catturare questo aspetto e che, attraverso i giusti interventi, emerga il fatto che, imponendo $A = 1$ al posto di $A = 0$, Vittima muore. Questo è il ruolo che svolge la nuova variabile R introdotta arricchendo il modello: nella situazione rappresentata dalla **Figura 3.2**, se imponiamo $R = 0$, l'azione di Assistente causa la morte di Vittima⁵⁸. Si noti che, quindi, in questo caso, stiamo rispettando la **Cond. 8**. Non è così nell'altra istanza del **PI** presentata.

⁵⁸ Questa trattazione risulta nel verdetto che Assistente causi la sopravvivenza di vittima. Per un commento si veda: Blanchard e Schaffer, «Cause without Default» cit.:p.205, nota 24.

La strategia consiste, nuovamente, nell'interpolare una variabile tra altre due in un sistema che risulta *impoverito* e che, quindi, non rispetta, nuovamente, la **Cond. 7**. Includiamo, come nel caso precedente, una variabile che rappresenti la disattivazione del veleno, con equazione $R = A \wedge B$ e otteniamo il modello rappresentato in **Figura 3.3**.

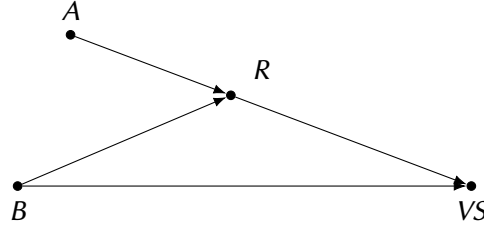


Figura 3.3: Grafo corretto per l'Esempio Bogus Prevention

Proviamo ad applicare le nostre definizioni di causa effettiva a questo caso. Le Definizioni **Causa effettiva^{org}** e **Causa effettiva^a** non considerano $B = 1$ come una causa di $VS = 1$. Per vederlo, si noti che è necessario utilizzare, al fine di far emergere la dipendenza controfattuale, $A = 0$ (Assassino avvelena) come parte del testimone. Ora la questione diventa stabilire in che insieme mettere R . Abbiamo le due seguenti opzioni:

1. Inserirlo nell'insieme \vec{Z} . Questo implica che, per valutare la **Cond. AC2(b^{org})** dovremo testare

$$(M, \vec{u}) \models [B \leftarrow 1, A \leftarrow 0, R \leftarrow 0](VS = 1)$$

ma questa condizione chiaramente fallisce, in quanto, nel mantenere R al suo valore attuale, non permettiamo alla reazione di neutralizzazione di avvenire.

2. Inserirlo nell'insieme \vec{W} . In questo caso, allora, si dovrà inserire nel testimone $R = 0$, altrimenti non emerge alcuna relazione controfattuale tra VS e B , in quanto il veleno sarebbe disattivato di default. Ma a questo punto ci ritroviamo nella situazione precedente.

Per la **Definizione Causa effettiva^a** vale un ragionamento analogo. Con la **Definizione Causa effettiva^m** non vi erano problemi nemmeno in principio, in quanto B non era considerato una causa in quanto, fissate le condizioni attuali del sistema, non vi era dipendenza controfattuale tra l'azione di Bodyguard e la sopravvivenza di Vittima.

Si noti, tuttavia, che tutte e tre le definizioni identificano $A = 1$ come una causa di $VS = 1$. Questo esito è forse leggermente controintuitivo, ma in un certo senso condivisibile. Dopotutto, se Assassino avesse messo il veleno, Vittima non sarebbe sopravvissuta, tenendo fermo il fatto che non vi è stata alcuna reazione chimica che ha permesso di annullare l'effetto del veleno.

Questa idea dell'interpolazione di una variabile era già stata discussa precedentemente rispetto al lavoro di **Blanchard e Schaffer**⁵⁹. Quello che i due autori aggiungono, fondamentale, è che questo fatto avrebbe dovuto creare dello scompiglio e mettere qualche dubbio sull'adeguatezza del modello rispetto alla situazione rappresentata. Dopotutto, due condizioni suggerite dalla letteratura come necessarie per un modello causale adeguato (le **Cond. 7** e **Cond. 8**) vengono violate da questo esempio. **Blanchard e Schaffer** pongono l'accento su questa problematica: se ci trovassimo in un

59 Ad esempio: **Halpern e Hitchcock**, «Graded causation and defaults» cit.: sec. 7.4 e **Halpern**, *Actual Causality* cit.: sec. 3.4.2.

contesto legale e si cercasse di far passare il primo modello come corretto, gli avvocati avversari potrebbero fare la seguente «devastante obiezione»⁶⁰:

Lei non si è reso conto di una parte essenziale della storia, ovvero che l'antidoto del suo cliente non ha mai effettivamente neutralizzato alcun veleno; si consideri questo e otteniamo il risultato opposto.⁶¹

La critica di **Blanchard e Schaffer** non è finita qui. Nella prossima sezione presenteremo in maniera organica il loro argomento, approfittandone per collegarci ad un'altra problematica, questa volta passata relativamente inosservata, perlomeno nella declinazione legata ai modelli causali⁶²: il *Problema del Disaccordo* (**PD**).

3.4 SUL PROBLEMA DELL'ISOMORFISMO E SUL PROBLEMA DEL DISACCORDO

La distinzione default/devianti, sia nel suo *twist* controfattuale (la proposta di **Menzies**, per intenderci), che nella sua declinazione funzionale (l'approccio di **Halpern e Hitchcock**) non risultano, personalmente, soddisfacenti. Nonostante condivida gran parte delle considerazioni di **Blanchard e Schaffer**, e parte di questo capitolo dovesse essere dedicata ad argomentazioni simili, non tutto il lavoro da loro svolto è completo. Soprattutto, alcuni dubbi rimangono. Questi pertengono a quello che **Halpern e Hitchcock** hanno definito il **PD**⁶³.

Parte di questa sezione sarà dedicata al sostenere perché ritengo che la distinzione tra stati di default e stati devianti sia superflua. Una parte sarà dedicata ad illustrare quanto sia diffusa la tendenza ad utilizzare modelli non adeguati. Ritengo che la diffusione di questi casi dimostri un eccessivo attaccamento da parte dei filosofi ai *neuron diagrams* di matrice lewisiana, contro il suggerimento di **Hitchcock**⁶⁴, e che questa poca sensibilità rallenti grandemente il dibattito presentando casi problematici che, in realtà, non sono tali una volta rappresentati correttamente. Ritengo che questa situazione vada a rafforzare l'argomento di **Blanchard e Schaffer** contro la distinzione default deviante.

3.4.1 *Isomorfismo e modelli non adatti*

Il fulcro argomentativo di **Blanchard e Schaffer** è esposto nel seguente passaggio:

*Quando ci si trova di fronte a case strutturalmente isomorfe ma causalmente distinti, si sospetta che almeno uno dei modelli sia impoverito o altrimenti non adeguato.*⁶⁵

Le regole proposte nella sezione precedente servono, per l'appunto, ad avere una qualche guida quando ci si trova in queste situazioni.

In letteratura non mancano i casi in cui vengono presentati modelli causali che palesemente infrangono queste regole. Stupisce che, a volte, nemmeno nozioni basilari come quelle di utilizzare

⁶⁰ **Blanchard e Schaffer**, «Cause without Default» cit.: 202.

⁶¹ *Ibidem*.

⁶² Di questo tipo di problematiche si è occupato, ad esempio **van Fraassen**, *The Scientific Image* cit.

⁶³ **Halpern e Hitchcock**, «Graded causation and defaults» cit.: 415.

⁶⁴ **Hitchcock**, «What's Wrong with Neuron Diagrams?» cit.

⁶⁵ **Blanchard e Schaffer**, «Cause without Default» cit.: 205.

variabili non logicamente legate vengano rispettate. Prendiamo il un classico esempio di *switching*, discusso da Collins *et al.*

ESEMPIO 3.4.1 (Switching)

Un treno si dirige verso un bivio. Poco prima del bivio vi è uno scambiatore di binario. Quest'ultimo, se attivo, porterà il treno sul binario di sinistra, se disattivo sul binario di destra. Poco dopo, i due binari si riuniscono. Una damigella in pericolo è legata ai binari poco dopo l'unione dei due binari

Il grafico del modello standard è riportato di seguito, in Figura 3.4.

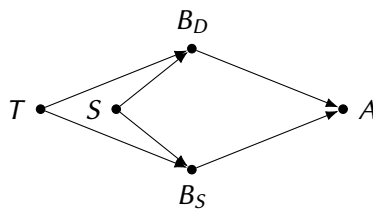


Figura 3.4: Grafo del modello che rappresenta l'Esempio Switching

Risulta evidente, dopo la discussione fatta precedentemente, che questo modello presenta due variabili che hanno una relazione logica, contro i dettami della Cond. 4. Queste sono B_D e B_S . Nella fattispecie, queste due variabili si prestano ad essere fuse in una, in quanto, come ci dice la Cond. 5, i valori della medesima variabile devono rappresentare situazioni incompatibili, o non compostibili.

Questo errore di modellazione genera una dipendenza *spuria*, in quanto afferma che vi è una dipendenza tra il fatto di *non essere su di un binario* e arrivare. L'aspetto corretto che, invece, ci interessa catturare, è che *l'essere su almeno uno dei due binari* è ciò che conta al fine di arrivare alla damigella, e non l'essere su di un binario od un altro.

Si potrebbe ribattere affermando che questo modello diviso permette di effettuare interventi separati sui binari. Non potrei essere più d'accordo. Il fatto è che anche il modello alternativo, proposto in Figura 3.5, è in grado di rendere conto di questi potenziali interventi (od ostacoli, se si vuole modellare l'impossibilità di muoversi liberamente su di un particolare binario).

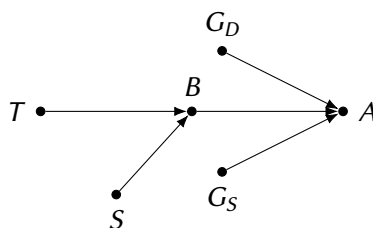


Figura 3.5

Un altro caso che in cui abbiamo una situazione simile, chiaramente ereditata dalla discussione basata sui network neurali, è il classico esempio del lancio del sasso da parte di Billy e Suzy: si tratta

66 J. Collins *et al.*, «Counterfactuals and Causation: History, Problems and Prospects», in Collins *et al.* (a cura di), *Causation and Counterfactuals* cit., pp. 1-58: 40.

67 Per alcune discussioni, vedere: Paul, «Aspect Causation» cit.; N. Hall, «Causation and the Price of Transitivity», in Collins *et al.* (a cura di), *Causation and Counterfactuals* cit., pp. 181-203; Swanson, «Lessons from the Context Sensitivity of Causal Talk» cit.; J. Schaffer, «Contrastive Causation», *The Philosophical Review*, vol. 114, n. 3 (2005), pp. 327-358, DOI: 10.1215/00318108-114-3-327.

del nostro **Esempio Rock Throwing**, rappresentato in **Figura 1.4**. Questo modello, che avevamo originariamente strutturato in maniera differente, rappresentata in **Figura 1.3**, risente della classica rappresentazione fornita dai teorici della causalità influenzati dall'approccio di stampo lewisiano. Un esempio tra gli altri sono **Collins et al.**⁶⁸, di cui riportiamo la rappresentazione in **Figura 3.6**.

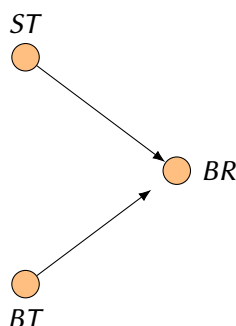


Figura 3.6: Diagramma neurale dell'**Esempio Rock Throwing**

Questa rappresentazione è sommaria e scorretta, in quanto non rappresenta una delle caratteristiche fondamentali dell'esempio in questione e rappresenta scorrettamente altri aspetti. Nella fattispecie, rappresenta scorrettamente la situazione rispetto al lancio di Billy. Nell'esempio, infatti, non si dà il caso che il sasso non raggiunga la regione spaziale della bottiglia, ma è piuttosto che, quando vi arriva, essa non c'è più. La rappresentazione è fuorviante in quanto elide questo fatto.

Inoltre, questione, a mio parere, più grave, è che il rapporto tra l'esito del lancio del sasso di Suzy, nel caso vada a segno, ed il lancio di Billy non è correttamente caratterizzato: il fatto che Suzy rompa la bottiglia, infatti, *impedisce* al sasso di Billy di raggiungere (e rompere) la bottiglia. Nel grafico non è reperibile nessuna relazione di questo tipo. Di conseguenza, mi sento di dire che l'intrinseca struttura causale della situazione è catturata malamente dalla **Figura 3.6**. Pertanto, una rappresentazione più adeguata potrebbe essere quella offerta nella **Figura 3.7**, una versione ridotta di quella proposta da **Lewis**⁶⁹.

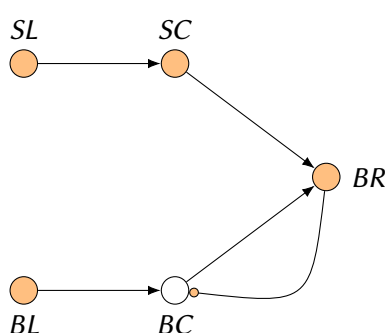


Figura 3.7: Rappresentazione alternativa dell'**Esempio Rock Throwing**

Anche questo grafo, però, è piuttosto problematico: questa modalità di rappresentazione, oltre che a prestare il fianco a potenziali confusioni in materia di quali neuroni fanno fuoco e quali no, quali neuroni stimolano altri neuroni e quali no, ecc., soffre di un ben più grave difetto, ovvero il *loop* tra i neuroni *BR* e *BC*. Nei casi in cui si generano cicli causali, abbiamo, generalmente, una severa problematica: non siamo in grado di capire cosa sia la causa effettiva di cosa. Ma, in questo

⁶⁸ Collins et al., «Counterfactuals and Causation: History, Problems and Prospects» cit.: fig. 1.3, pag.23.

⁶⁹ Lewis, «Postscripts to 'Causation'» cit.: 204.

caso, abbiamo una chiara intuizione di quale sia la causa effettiva (i) della rottura della bottiglia e (ii) del fatto che Billy non colpisca il sasso: il fatto che Suzy colpisca la bottiglia. Inoltre, situazioni con *feedback* causali di questo tipo non sono modellabili attraverso un modello causale⁷⁰.

Una corretta rappresentazione della situazione può essere ottenuta attraverso i *dynamic models* proposti da Pearl⁷¹, che, infatti, li utilizza per rappresentare correttamente la situazione⁷². Questo tipo di rappresentazione evita anche le problematiche discusse da Hall a proposito del modello rappresentato in Figura 1.4. Si veda il grafo in Figura 3.8.

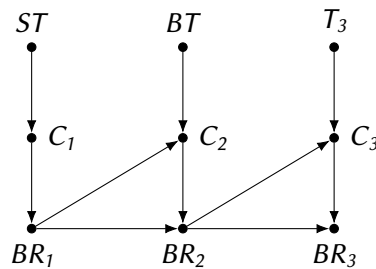


Figura 3.8: Grafo indicizzato sul tempo dell'Esempio Rock Throwing

Il modello che genera questo grafo rappresenta correttamente tutte le nostre intuizioni, utilizzando il semplicissimo espediente di associare alle variabili un'indicizzazione temporale. Naturalmente, questo significa accettare una concezione fragile degli eventi, che abbiamo già visto essere problematica. In questo caso, però, non credo che lo sia: la distinzione temporale ci è necessaria in questi casi in quanto è all'interno della struttura temporale che si manifestano le caratteristiche principali dei casi di *late preemption*.

Rifiutare di catturare questi aspetti fondamentali della situazione per questioni di principio a livello di individuazione di eventi mi sembra assolutamente non corretto: considerare la rottura della bottiglia da parte di Billy e Suzy come il medesimo evento è un'istanza di quell'impovertimento di modelli di cui ci parlano Blanchard e Schaffer in quanto stiamo privando il modello del numero di variabili necessarie a catturare correttamente la situazione.

Presenterò di seguito un ultimo esempio tratto da Halpern e Hitchcock⁷³, che lo discutono al fine di mostrare quanto la distinzione default/deviante sia utile in contesti legali. Anche in questo caso abbiamo a che fare, a mio parere, con un modello impoverito. Il modello vuole rappresentare una situazione da analizzare in astratto in un contesto legale. In questi contesti, si ritiene che un agente non sia responsabile per degli esiti a cui le proprie azioni hanno portato se (i) un altro agente ha agito successivamente, in maniera deliberata, al fine di realizzare l'esito; o (ii) l'esito è prodotto da un evento estremamente improbabile⁷⁴.

Immaginiamo la seguente situazione:

ESEMPIO 3.4.2 (Dolo o colpa? Colpa)

Anna e Bob sono in campeggio. Anna vuole fare un bel barbecue e ha portato della benzina per aiutarsi a fare un bel fuoco (controllato, naturalmente). Mentre si sta versando la benzina, una splendida farfalla le passa accanto. Anna, che è affascinata dalle farfalle, si distrae e rovescia la benzina.

⁷⁰ Pearl, *Causality: Models, Reasoning, and Inference* cit.: sec. 10.3-5.

⁷¹ *Ivi*.

⁷² Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.: 864.

⁷³ Halpern e Hitchcock, «Graded causation and defaults» cit.: sec. 7.6.

⁷⁴ Hart e Honoré, *Causation in the law* cit.

Bob, che proprio in quel momento ha finito di fumare la sua sigaretta, la lancia senza guardare, e appicca l'incendio. ■

In questo caso, Anna è da considerarsi colpevole⁷⁵. Ma immaginiamo uno scenario leggermente differente:

ESEMPIO 3.4.3 (Dolo o colpa? Dolo)

Tutto si svolge esattamente come prima, fino ad un momento prima che Bob lanci la sigaretta. Il nostro Bob è, in questa versione della storia, un piromane e, avendo visto Anna rovesciare la benzina, coglie l'occasione al volo e fa finta di lanciare la sigaretta senza guardare. L'incendio divampa di lì a poco, nella stessa maniera. ■

In questo caso è evidente che il colpevole è Bob. Immaginiamoci in un contesto legale e di trovarci nella necessità di svolgere questo tipo di ragionamenti al fine di determinare un colpevole. Halpern e Hitchcock sostengono che, attraverso un modello causale esteso, siamo in grado di catturare questo tipo di differenze di giudizio. Di seguito, il modello proposto, con grafo corredato in Figura 3.9:

$AN=0$ se Anna è negligente

$AS = AN$

$BM=1$ se Bob è malvagio

$BC=1$ se Bob è noncurante

$BT = \max(BC, BM)$

$F = \min(AS, BT)$

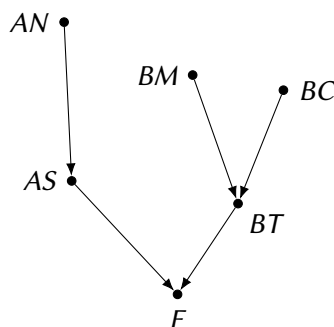


Figura 3.9: Modello proposto

L'argomentazione di Halpern e Hitchcock è basata sulla proposta di utilizzare l'ordine di normalità al fine di rappresentare quella che sono i diversi gradi di *mens rea*. In questo caso, il fatto che Bob sia distratto viene trattata come lo stato di *default*, ovvero una totale assenza di *mens rea*, il fatto che Anna sia stata negligente come un primo grado di *mens rea* e, quindi, nel modello causale esteso, come una leggera deviazione dal default. Infine, il fatto che Bob sia malvagio viene modellato come una deviazione sostanziale, in quanto rappresenta l'apice della colpevolezza, in quanto ha agito con intenzione criminale.

In questa sede, però, non ci interessano i verdetti del modello. Quello che ho intenzione di far notare è quanto sia comune effettuare errori di modellazione, anche quando questi modelli sono

⁷⁵ Halpern e Hitchcock, «Graded causation and defaults» cit.: 447.

costruiti da “*artisti*” della modellazione. Ritengo che mostrare questa diffusione rafforzi, in qualche maniera, l'argomento di **Blanchard e Schaffer**. Torniamo al modello.

Nuovamente, anche questo modello rappresenta un'istanza di violazione della **Cond. 7**. Non viene, infatti, caratterizzata correttamente una parte della situazione, a causa della mancanza di una variabile e delle relazioni che la sua inclusione genererebbe, e che, invece, sono presenti nella situazione da modellare.

La variabile che va aggiunta a questo modello è quella che rappresenta il fatto che, nel caso Bob sia malvagio, coglie l'occasione, volontariamente, per scatenare l'incendio. Vi sono due modi di modellare questa situazione: o si fa in modo che la variabile vada a sfociare in un percorso già attivo, generando un modello simile a quello in **Figura 3.10a**, o si seleziona una dipendenza esclusiva che va dalla variabile all'incendio, una scelta volta a determinare una separazione di meccanismi causali⁷⁶, generando un modello come in **Figura 3.10b**.

$$\begin{aligned}
 AN &= 1 \text{ se Anna è negligente} \\
 AS &= AN \\
 BM &= 1 \text{ se Bob è malvagio} \\
 OCC &= AS \times BM \\
 BC &= 1 \text{ se Bob è noncurante} \\
 BT &= \begin{cases} \max(OCC, BC) & \text{in Figura 3.10a} \\ BC & \text{in Figura 3.10b} \end{cases} \\
 F &= \begin{cases} \min(AS, BT) & \text{in Figura 3.10a} \\ \min(AS, OCC, BT) & \text{in Figura 3.10b} \end{cases}
 \end{aligned}$$

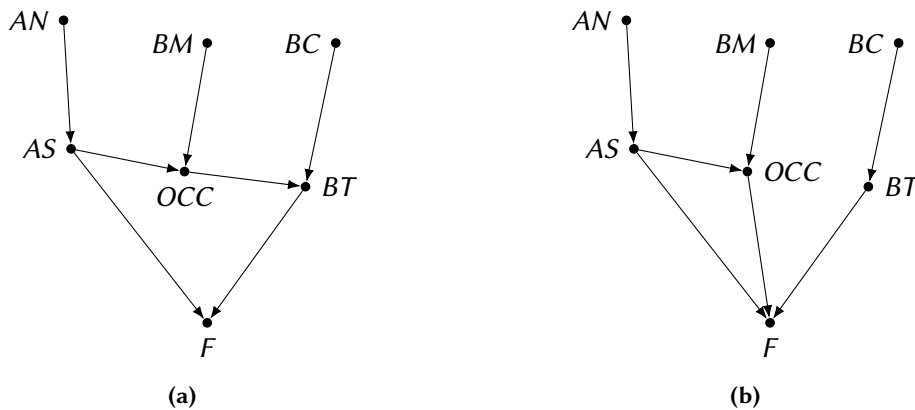


Figura 3.10: Correzione del modello proposto da **Halpern e Hitchcock**

Un ultimo esempio che tratterò è stato proposto da **Weslake** e poi discusso da **Halpern**⁷⁷. L'esempio è Lampadina (**Esempio 2.2.5**)

Qual'è la causa effettiva dell'accensione della lampadina? Le definizioni **HP**, esclusa la **Definizione Causa effettiva^m**, ci informano che anche $A = 1$ è una causa di $L = 1$. Ma non è questa la nostra intuizione. Naturalmente, solo $(B = -1, C = -1)$ sono cause dell'accensione della lampadina. Ma si consideri il caso in cui L si accende solo se (i) nessuna tra $A, B, e C$ è in posizione -1; (ii) tra

⁷⁶ Una scelta che risulta tutto sommato criticabile: dopotutto, l'incendio dipende ancora dal lancio della sigaretta.

⁷⁷ **Halpern, Actual Causality** cit.: 110–111.

A, B , e C è in posizione 0; (iii) tra A, B , e C è in posizione 1. Ora che $A = 1$ sia una causa non è così controintuitivo.

Questo è chiaramente un caso di isomorfismo. Questo esempio è nato come una potenziale critica alle definizioni della famiglia **HP**. Il fulcro dell'argomento è il fallimento di queste definizioni in casi di questo tipo, che definirei di *switching composto*. Il problema è che l'esempio, utilizzato come critica, fa una leva tacita sul fatto che, invece di trattarsi di un unico modello che possiede un'unica struttura causale, è piuttosto quello che Van Fraassen chiama una *famiglia di modelli*⁷⁸. Esso nasconde, infatti, al suo interno, un po' come un cavallo di Troia, un contenuto che ne determina il successo per vie trasverse: un altro modello. Quello che la storia fa è renderci cognitivamente disponibile il primo modello, e poi valutare l'efficacia della definizione sulla coppia di modelli, su di una struttura causale parzialmente determinata e che quindi produce un giudizio scorretto rispetto alla situazione. La forza dell'argomento contro le definizioni **HP** deriva proprio dal fatto che l'esempio *sfrutta* tacitamente questo isomorfismo con un'altra situazione causale, al fine di mostrare come la definizione porti ad un esito errato.

Come si risolve questo isomorfismo? *Rappresentando correttamente l'intrinseca struttura causale della situazione*, ovvero specificando quali sono i meccanismi⁷⁹, ovvero aggiungendo le variabili e le relazioni necessarie a questo fine. Se non si utilizza un numero di variabili sufficienti, la struttura risulta *sottodeterminata*, e, quando essa è sottodeterminata, le definizioni di causalità devono arrangiarsi a lavorare con quello che viene loro fornito, e i giudizi espressi non possono essere considerati né corretti, né definitivi.

Vi sono molti altri casi in cui i modelli non sono sufficientemente caratterizzati⁸⁰. Spesso, come in questo caso, sono differenze che, nel contesto in cui sono presentate, non sono essenziali ma comunque, a mio parere, rilevanti. Ritengo che una stretta fedeltà del modello rispetto alla situazione che si vuole modellare sia fondamentale, e che l'assenza di variabili che svolgono ruoli anche secondari rappresenti, a volte, possibili problematiche nascoste, che sfuggono concentrandosi sui pattern di dipendenza dei quali l'autore sta trattando e per cui ha prodotto il modello.

Inoltre, lo sforzo massimo per aderire alle condizioni presentate all'inizio di questa sezione è una preconditione fondamentale al fine di far procedere in maniera costruttiva il dibattito sulla causa attuale. Risulta, purtroppo, difficile caratterizzare adeguatamente le strutture causali. Questo perché, oltre a esserci disaccordo tra gli autori su cosa possa essere considerato una causa⁸¹ e, quindi, cosa faccia parte della struttura intrinseca di una situazione o di un sistema e quali siano i rapporti all'interno di questa struttura; ci sono forti *bias* cognitivi in questo senso, che ci rendono immediatamente disponibili modelli che non sono quelli corretti, o che sono *sottodeterminati*. La prima delle due questioni, ovvero quella del disaccordo, costituisce, per l'appunto, il cosiddetto **PD**⁸².

⁷⁸ Van Fraassen, *The Scientific Image* cit.

⁷⁹ In questo caso il termine meccanismo appare estremamente adatto, si veda: Woodward, «What Is a Mechanism? A Counterfactual Account» cit.; P. Menzies, «The causal structure of mechanisms», *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n. 4 (2012), pp. 796-805, DOI: 10.1016/j.shpsc.2012.05.008.

⁸⁰ Solo per citarne alcuni, si veda Halpern, *Actual Causality* cit.: cap.4. In molti di questi casi, Halpern fa notare che le strutture proposte sono compatibili con differenti storie causali.

⁸¹ Si vedano, ad esempio, H. Beebe, «Causing and Nothingness», in Collins et al. (a cura di), *Causation and Counterfactuals* cit., pp. 291-308 e M. S. Moore, *Causation and Responsibility. An Essay in Law, Morals, and Metaphysics*, Oxford University Press, Oxford 2009: cap. 3,5,18, dove si sostiene che le omissioni non possono essere cause, contro il normale uso che facciamo delle omissioni come cause e l'opinione di gran parte degli altri autori.

⁸² Halpern e Hitchcock, «Graded causation and defaults» cit.: 415.

3.4.2 *Il Problema del Disaccordo*

Gran parte del contenuto di questa sezione è ispirato dal lavoro di Fischer proprio sul PD⁸³, con il quale condivido la maggior parte delle riflessioni svolte. È necessario ammettere che il lavoro è estremamente dettagliato, ma cercherò di aggiungere le mie considerazioni personali. Infatti, pur seguendo il medesimo percorso, gli esiti della riflessione saranno parzialmente differenti. Procediamo ora con l'analisi del PD.

Blanchard e Schaffer rifiutano la distinzione default/deviante attraverso tre differenti linee argomentative. La prima si basa sulla vaghezza e sulle situazioni poco chiare che la distinzione genera. Come abbiamo visto, Halpern e Hitchcock distinguevano tra tre concetti di normalità. Il problema è che tutti e tre questi concetti sono poco chiari e estremamente flessibili. Inoltre, non vi è modo per definire correttamente un equilibrio tra i tre. Cosa dovremmo considerare normale quando abbiamo due norme contraddittorie? Per esempio, la legge impone limiti di velocità in auto, ma la maggior parte degli automobilisti non li rispetta⁸⁴. Dovremmo considerare normale la norma statistica degli automobilisti, o quella prescrittiva della legge? Inoltre, quando si valuta una norma statistica, bisogna selezionare una classe di riferimento. Quali criteri per selezionarla? Sulla base di che caratteristiche salienti della situazione? Non credo che vi possano essere dei criteri obiettivi sulla base dei quali selezionare le risposte corrette a queste domande.

La seconda linea argomentativa si basa sul fatto che la distinzione default deviante non sia genuinamente causale, ma sia parassitaria rispetto ad un'euristica cognitiva che seleziona gli scenari controfattuali sui quali basare i nostri ragionamenti. L'influenza di fattori di tipo normale sarebbe, quindi, l'influenza di un'euristica carica di questo tipo di considerazioni che viene proiettata su di un concetto che normativo non è. Per sostenere questo argomento paragonano questa euristica alla famosa fallacia probabilistica: quando sappiamo che Alice è laureata in Scienze Filosofiche ed è iscritta ad un partito di sinistra, giudichiamo l'enunciato

1. Alice è impiegata in banca e femminista.

più probabile di

2. Alice è impiegata in banca.⁸⁵

Questa fallacia è dovuta all'utilizzo di un concetto di probabilità che è, naturalmente, privo di condizionamenti di tipo normativo, ma aiutato da un'euristica che, invece, è, da questo punto di vista, estremamente carica. Come questo argomento non supporta una visione *norm-laden* della probabilità, l'argomento basato sul fatto che i nostri giudizi causali sono *norm-laden* non va a sostenere un concetto di causalità anch'esso normativo.

Il terzo argomento, che abbiamo già presentato, è quello basato sull'adeguatezza dei modelli. Pur trovandomi perfettamente d'accordo con questo argomento, esso non si applica nel momento in cui si tratta di risolvere il cosiddetto PD⁸⁶.

Nella fattispecie, mi trovo d'accordo con Fischer nel sostenere che il successo di questa linea argomentativa di Blanchard e Schaffer comporti non la soluzione della questione dei default, bensì un suo mero spostamento in un dominio al di là di quello del PI. Entriamo nello specifico.

83 Fischer, «Causation and the Problem of Disagreement» cit.

84 Blanchard e Schaffer, «Cause without Default» cit.

85 Esempio ripreso da A. Tversky e D. Kahneman, «Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment», in *Reasoning*, Cambridge University Press, 2008, pp. 114-135, DOI: [10.1017/cb09780511814273.008](https://doi.org/10.1017/cb09780511814273.008).

86 Fischer, «Causation and the Problem of Disagreement» cit.

Ricordiamo che la distinzione e gli ordinamenti tra stati normali e stati anormali è necessaria per catturare delle disuniformità di giudizio tra due modelli causali che appaiono isomorfi. La soluzione era rendere uno dei due modelli adeguati, al fine di catturare correttamente la situazione causale e produrre il verdetto corretto. Blanchard e Schaffer ci presentano un altro caso di modellazione potenzialmente scorretta: quei casi in cui un'infinità di omissioni permettono ad un evento di accadere⁸⁷. L'esempio più classico⁸⁸ è quello del giardiniere:

ESEMPIO 3.4.4 (Il Re e il giardiniere)

Giardiniere, il cui compito è quello di curare delle begonie, non le bagna. Gli splendidi fiori muoiono poco dopo. ■

In questa situazione, chiunque considera il mancato innaffiamento delle piante *da parte Giardiniere* come una causa della morte dei fiori. Nessuno, invece, considera il fatto che il Re d'Inghilterra non si sia scomodato per innaffiare le piante come una causa di questo evento; e come vale per il Re d'Inghilterra, così vale per tutti gli altri (ecluso, naturalmente, Giardiniere). Questa asimmetria causale è problematica, nonché estremamente diffusa⁸⁹. Il problema è che gli autori che negano che le omissioni possano svolgere ruolo causale negano che l'omissione di Giardiniere sia una causa. Qui sorge il PD. Halpern e Hitchcock ci informano che:

Prima facie, non sembra che una teoria della causa effettiva possa rispettare tutti questi giudizi senza cadere in inconsistenze.⁹⁰

In questo caso, la ricetta proposta da Blanchard e Schaffer è guidata dalla Cond. 3: escludere le variabili che non consideriamo rilevanti, o che rappresentano possibilità troppo remote. Questa soluzione, però, porta con sé dei problemi.

In particolare, si immagini di essere in un contesto in cui si discute, ad esempio, un'attribuzione di colpa. I due partiti rivali presentano due modelli causali, estremamente simili, ma uno dei due risulta privo di una variabile che l'altro possiede. Di fronte a che tipo di situazione ci troviamo? Siamo di fronte ad un caso in stile Halpern e Pearl? «[U]no scenario estremamente irragionevole»⁹¹? Oppure uno dei due modelli non riconosce alcuna rilevanza causale alla variabile omessa, nel senso che la espelle dalla struttura causale intrinseca della situazione, in quanto priva di influenza? Oppure, semplicemente, è una contingenza non considerata?

87 Lo status metafisico e validità come relata causali delle omissioni è piuttosto dubbio: ci sono autori che non le riconoscono come cause: P. Dowe, *Physical Causation*, Cambridge University Press, 2000; Beebe, «Causing and Nothingness» cit.; Moore, *Causation and Responsibility* cit. Altri che invece le difendono: D. Lewis, «Void and Object», in Collins *et al.* (a cura di), *Causation and Counterfactuals* cit., pp. 277-290; J. Schaffer, «Causation by Disconnection», *Philosophy of Science*, vol. 67, n. 2 (2000), pp. 285-300, DOI: 10.1086/392776; Schaffer, «Trumping Preemption» cit.

88 Menzies, «Difference-making in Context» cit.: 145, S. McGrath, «Causation By Omission: A Dilemma», *Philosophical Studies*, vol. 123, n. 1-2 (2005), pp. 125-148, DOI: 10.1007/s11098-004-5216-z, C. Sartorio, «The Prince of Wales Problem for Counterfactual Theories of Causation», in *New Waves in Metaphysics*, Palgrave Macmillan UK, 2010, pp. 259-276, DOI: 10.1057/9780230297425_14.

89 Vi sono, infatti, un'infinità di mancanze che fungono da cause di ogni evento: il mio scrivere questa tesi è permesso dal fatto che non manca l'elettricità, o che la stanza non stia andando a fuoco. Il problema risale addirittura a J. S. Mill, *Sistema di logica deduttiva e induttiva*, a cura di M. Trinchero, Classici della Filosofia, UTET, 1988. Del problema si sono occupati diversi autori: Lewis, «Void and Object» cit.; Mackie, «Causes and Conditions» cit.; Menzies, «Difference-making in Context» cit.; P. Menzies, «Platitudes and counterexamples», in Beebe *et al.* (a cura di), *The Oxford Handbook of Causation* cit., pp. 341-367, DOI: 10.1093/oxfordhb/9780199279739.001.0001, solo per citarne alcuni.

90 Halpern e Hitchcock, «Graded causation and defaults» cit.: 415.

91 Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.: 869.

La domanda fondamentale è, insomma, *come costruiamo e giustifichiamo un modello?* Perché «la scelta rispetto a quali variabili sono rilevanti e quali vanno ignorate e la decisione rispetto a quali contingenze considerare seriamente è soggettiva e *dipende [...] da che uso facciamo del modello*»⁹²?

Un modello è sempre informato da informazioni di background riguardo al dominio in due sensi:

1. Rispetto al dominio inteso come dominio causale che stiamo modellando, ovvero la conoscenza del modellista della situazione e della sua struttura e relazioni intrinseche.
2. Il dominio rispetto al quale il modello va messo a frutto, il contesto nel quale mettiamo all'opera il modello e nel quale lo utilizziamo.

La scelta delle variabili è informata da questo contesto, e questo contesto è normativamente carico: definisce quali sono i controfattuali da valutare come plausibili, quali sono invece troppo remoti e non vanno inclusi. Questo implica una selezione delle variabili che faccia giustizia a queste considerazioni contestuali. Ora osserviamo che la nostra *segnatura* *S* non è né *causalmente neutra*, né *normativamente neutra*.

A questo deve aggiungersi il fatto che un determinato dominio definisce, oltre che le variabili, la classe di contrasto che ci interessa valutare: anche questa, infatti, descrive

1. Quali altre “opzioni” ha una variabile, come può muoversi all'interno del proprio dominio di pertinenza. Ad esempio, è possibile caratterizzare un malfunzionamento in diversi modi: trattarlo come una variabile binaria (malfunzionamento sì, malfunzionamento no), oppure specificare di che tipo di malfunzionamento si tratta (un problema idraulico? Meccanico? A livello di *software*?).
2. Quali aspetti vanno considerati come alternative plausibili e quali no.

Io credo che la risposta alle nostre domande sia da ricercarsi nella distinzione tra stati di default e devianti e la concezione che abbiamo della norma nel dominio all'interno del quale modelliamo. L'argomento di **Blanchard e Schaffer** ha, quindi, successo per quanto riguarda il **PI**, ma per un altro verso, non è definitivo, e non fa altro che spostare una problematica prima indirizzata dal concetto di default ad un livello ulteriore, il livello della selezione e giustificazione del modello.

3.4.3 *Quale il posto per la distinzione default/deviante?*

Per quanto concerne la distinzione default/deviante, non credo che essa debba essere incorporata nella nostra definizione di causa effettiva, anche se vi sono stati tentativi di successo in questo senso nella letteratura recente⁹³. Il rendere soggettivi dei giudizi che sono principalmente dettati da rapporti strutturali intrinseci ad un sistema risulta, a mio parere problematico⁹⁴. Questa mia visione non è condivisa da **Halpern e Hitchcock**, che ritengono che i giudizi di causa effettiva siano soggettivi. Io penso che la dicitura *soggettivi* sia problematica. La dicitura più corretta è contestuali. La contestualità rispetto ad un modello è accettabile, in quanto rappresenta una contestualità rispetto

⁹² *Ivi*: 878. Enfasi mia.

⁹³ F. Zhu, «A New Halpern-Pearl Definition of Actual Causality by Appealing to the Default World», *Axiomathes* (2022), DOI: [10.1007/s10516-021-09613-z](https://doi.org/10.1007/s10516-021-09613-z).

⁹⁴ Questa è una posizione condivisa da molti. Ad esempio, **McGrath**, «Causation By Omission: A Dilemma» cit.: 125: «[...] la causalità è comunemente ritenuta essere un esempio paradigmatico di relazione naturale, e quindi interamente non normativa.».

alla nostra migliore conoscenza del fenomeno e alla situazione nel quale il modello deve essere utilizzato, e non è una contestualità dettata da giudizi psicologici o *bias* cognitivi. Una volta che il modello è determinato, la causa effettiva è determinata.

Anche l'evidenza psicologica a favore degli account estensivi dei modelli causali⁹⁵, incontra qualche resistenza⁹⁶. Nella fattispecie, sembra giustificato il sospetto che quando vengono sottoposti quesiti come quelli preparati da Hitchcock e Knobe e Icard *et al.* e viene chiesto quale, o chi, sia la causa di un qualche evento, questa domanda non venga percepita come una domanda sulla causalità come relazione *pura*, bensì come una domanda riguardo chi è colpevole o cosa avremmo dovuto fare per evitare l'esito. Le risposte che forniamo non sono risposte veramente causali: è vero che individuiamo la causa, ma la risposta effettiva che forniamo cattura un aspetto che per noi risulta sì fondamentale, ma non per questo esso è l'esclusiva causa e neghiamo che gli altri aspetti siano causali o abbiano rapporti di tipo causale rispetto all'effetto.

Inoltre, Danks *et al.* hanno sperimentalmente testato in maniera ulteriore le tesi di Hitchcock e Knobe, non rilevando alcuna differenza nei giudizi causali tra due gruppi di partecipanti, ai quali sono state presentate due storie differenti, una moralmente carica e l'altra no⁹⁷.

Quello che questi esperimenti sembrano suggerire, a mio parere, è che quando siamo posti di fronte a quesiti causali produciamo un'interpretazione contestuale riguardo allo scenario presentatoci. Per esempio, è a mio parere rilevante che le vignette utilizzate da Hitchcock e Knobe differiscano parzialmente da quelle utilizzate da Danks *et al.*: le prime sembrano, infatti, suggerire che quello che ci viene chiesto sia quale intervento effettuare al fine di una correzione della situazione, cosa che le seconde non fanno. La sensazione che, a mio parere, si ha leggendo le due storie è piuttosto differente e le prime elicitano più facilmente giudizi correttivi⁹⁸.

Hitchcock e Knobe confermano questa visione in qualche maniera, e infatti sostengono che:

questi [criteri] sono progettati in maniera tale da dirigersi verso le migliori strategie di intervento.⁹⁹

Il modo corretto di rendere conto di questo fenomeno, alla luce degli esperimenti di Danks *et al.*, non è sostenere che i nostri giudizi causali siano influenzati da norme, bensì che, quando sottoposti a quesiti causali, questi vengano interpretati nel senso di *domande perché*¹⁰⁰. Non è il giudizio di

95 Knobe e Fraser, «Causal Judgement and Moral Judgement: Two Experiments» cit.; Hitchcock e Knobe, «Cause and Norm» cit.; Icard *et al.*, «Normality and actual causal strength» cit.; D. Pettit e J. Knobe, «The Pervasive Impact of Moral Judgment», *Mind & Language*, vol. 24, n. 5 (2009), pp. 586-604, DOI: 10.1111/j.1468-0017.2009.01375.x; J. Sytsma *et al.*, «Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions», *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n. 4 (2012), pp. 814-820, DOI: 10.1016/j.shpsc.2012.05.009.

96 D. Danks, «The psychology of causal perception and reasoning», in Beebe *et al.* (a cura di), *The Oxford Handbook of Causation* cit., pp. 447-470, DOI: 10.1093/oxfordhb/9780199279739.001.0001; D. Danks *et al.*, «Demoralizing causation», *Philos Stud*, vol. 171, n. 2 (2013), pp. 251-277, DOI: 10.1007/s11098-013-0266-8; D. Danks, «Functions and Cognitive Bases for the Concept of Actual Causation», *Erkenntnis*, vol. 78, n. S1 (2013), pp. 111-128, DOI: 10.1007/s10670-013-9439-2; C. Glymour *et al.*, «Actual causation: a stone soup essay», *Synthese*, vol. 175, n. 2 (2009), pp. 169-192, DOI: 10.1007/s11229-009-9497-9.

97 Danks *et al.*, «Demoralizing causation» cit.

98 Con questo non si vuole affermare che l'esperimento sia mal progettato. Hitchcock e Knobe si soffermano lungamente sul come alcune storie siano costruite in modo da evitare di suscitare intuizioni morali e come sia stata attentamente testata l'ipotesi che ci siano nozioni di colpa coinvolte (ad esempio utilizzando scenari impersonali). Resta il fatto che i test utilizzati non mi convincono in quanto, in qualche modo, collegati ad un'idea di *agency*. Ho il sospetto che utilizzando storie in cui è meno coinvolta l'idea di manipolazione, il risultato potrebbe essere differente. Se questo fosse vero, implicherebbe che anche l'inferenza di Hitchcock e Knobe sul trovare i migliori obiettivi di intervento non sarebbe giustificata.

99 Hitchcock e Knobe, «Cause and Norm» cit.: 612.

100 Van Fraassen, *The Scientific Image* cit.

causalità effettiva che è carico di norme, ma è l'interpretazione del quesito che, avvenendo all'interno di un contesto, si carica delle norme proiettate da quest'ultimo. Quello che succede è che i soggetti a cui vengono sottoposte le storie non rispondono a quesiti sulla causalità in senso metafisico, bensì rispondono a richieste di spiegazione rispetto al fenomeno avvenuto. Fondamentalmente, rispondono ad una domanda del tipo: *se dovessi spiegare cosa è andato storto*¹⁰¹, *quali di questi due fattori citeresti?* Infatti: «È da tempo riconosciuto il fatto che gli ordinari giudizi di causalità fanno uso di informazioni che vanno al di là rispetto a qualsiasi cosa si possa includere in una struttura causale. Le persone sembrano affidarsi a informazioni extra-strutturali per selezionare come candidati alcune cause rispetto ad altre»¹⁰². Questo tipo di fattori extra-strutturali sono, appunto, fattori pragmatici che determinano la spiegazione migliore, non la selezione della causa. Il contesto, fondamentalmente, determina cosa è appropriato *trattare come causa*, ma non *quale sia* la causa effettiva.

Questa tendenza è perfettamente coerente con un utilizzo *funzionale*¹⁰³ del concetto di causalità: diciamo che *E* è causato da *C* se riteniamo che, in questo contesto, la correzione dell'aspetto *C* sia la migliore correzione da effettuare per ottenere *E'*. Questa nozione non ha a che fare minimamente con questioni normative. Segue semplicemente il principio che il miglior intervento correttivo è quello che richiede la minima deviazione dallo stato del sistema.

Il ruolo del concetto di default non è, quindi, definitorio¹⁰⁴; esso non deve svolgere alcun ruolo nella definizione di causa effettiva, bensì deve caratterizzarsi come un mezzo per esplicitare e giustificare scelte modellistiche. La funzione è, esattamente quella proposta da Halpern e Hitchcock, che, dopo averla formulata così chiaramente, paiono essersela dimenticata nei successivi lavori:

Un avvocato che, in tribunale, cerchi di argomentare che il malfunzionamento dei freni sia stato la causa dell'incidente dovrà essere in grado di giustificare il suo modello; similmente, il suo avversario dovrà comprendere in che cosa consista un legittimo attacco al modello.¹⁰⁵

ovvero, è nella fase di giustificazione e “contrattazione” rispetto a quale sia il modello adeguato che la distinzione default/deviante entra in campo.

Propongo, a questo fine, la seguente procedura, che incorpora la nozione di default come guida alla modellazione al fine di una massimizzazione dell'accordo. Questa procedura fa uso di un approccio basato sulle funzioni di ordinamento simile a quello proposto da Huber¹⁰⁶, ma con un intento radicalmente differente. Di seguito la procedura.

PROC1 *Si includa nel modello causale ogni variabile afferente al dominio sotto analisi in modo che le dipendenze controfattuali siano massimamente rappresentate e la struttura non sia sottodeterminata sotto nessun aspetto e in nessun senso.*

PROC2 *Si utilizzi un ordinamento di normalità (o si esplicitino dei default) al fine di caratterizzare e giustificare la scelta delle variabili salienti e non salienti. Nella fattispecie:*

¹⁰¹ In tutte le storie proposte da Hitchcock e Knobe vi è sempre *qualcosa che va storto*, per così dire: non siamo mai di fronte a casi di causazione *pulita*.

¹⁰² Hitchcock e Knobe, «Cause and Norm» cit.: 592.

¹⁰³ J. Woodward, «A Functional Account of Causation: or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)», *Philos. of Sci.*, vol. 81, n. 5 (2014), pp. 691-713, DOI: 10.1086/678313.

¹⁰⁴ Contra Fischer, «Causation and the Problem of Disagreement» cit.

¹⁰⁵ Halpern e Hitchcock, «Actual Causation and the Art of Modeling» cit.: 385.

¹⁰⁶ Huber, «Structural Equations and Beyond» cit.

- PROC2(a) *Si raggruppino assieme tutte le variabili caratterizzate da una normalità massimale. Con normalità massimale si intende che una violazione, anche se minima, di queste condizioni caratterizza uno scenario altamente anormale.*
- PROC2(b) *Si includano tra le variabili massimamente normali tutte le omissioni e condizioni di sfondo che normalmente non vengono modellate.*
- PROC2(c) *Si associ alle variabili che vogliamo caratterizzare come salienti e rilevanti nella situazione che vogliamo modellare un ordinamento di normalità minore, in base alla rilevanza che ad esse associamo. In questo passo, si seguano semplicemente i procedimenti standard adottati da Halpern e Hitchcock¹⁰⁷.*
- PROC3 *Si associ alle variabili con ordinamento massimamente normale una variabile esogena e la si caratterizzi, nuovamente, come massimamente normale. Questa variabile, \vec{U}_{MAX} rappresenta la baseline della situazione. Nuovamente, una modifica di questa variabile andrà caratterizzata come massimamente anormale.*
- PROC4 *Si includano nel modello le altre variabili come esogene o endogene e si modellino le equazioni strutturali in base alla conoscenza riguardo al dominio sotto esame.*
- PROC5 *Si rispetti la Cond. 7 nella modellazione e non si escludano dipendenze per alcuna ragione.*
- PROC6 *Si associ alle variabili \vec{V}_{SEC} , che esplicano una funzione di dipendenza considerata secondaria rispetto al dominio, un ordinamento massimamente non normale per tutto il range $\mathcal{R}(\vec{V}_{SEC})$. In questo modo una loro violazione verrà considerata altamente normale e non inciderà sull'ordine di normalità rispetto alle variabili salienti.*
- PROC7 *In caso di disaccordo riguardo \vec{U}_{MAX} , si proceda con un'ulteriore distinzione delle variabili $\vec{U} \in \vec{U}_{MAX}$ sino all'accordo. Si ricaratterizzi l'ordine di normalità (o i default) rispetto alla nuova segnatura \mathcal{S} ottenuta dalla ulteriore caratterizzazione.*
- PROC8 *In caso di disaccordo riguardo \vec{V}_{SEC} si proceda in modo analogo alla Cond. PROC7.*

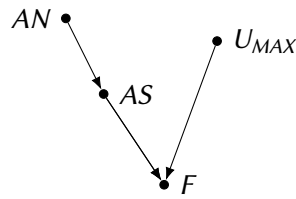
Il processo consiste, semplicemente, nell'associare alle condizioni di sfondo una normalità massima, in modo che una loro violazione venga considerata molto rara e anormale. All'estremo opposto, le variabili che riteniamo non avere una grande influenza causale rispetto alla situazione vengono caratterizzate da un ordinamento estremamente anormale su tutto il loro range, in modo che siano considerate fondamentalmente irrilevanti rispetto al contesto. Se non si è d'accordo su di una qualche variabile, semplicemente si contratta fino all'ottenimento di un ruolo causale ritenuto da entrambi i partiti soddisfacente.

Stabilito il modello causale adeguato, la determinazione della causa effettiva è solo materia di applicare la definizione. In questo modo si arriva ad ottenere, all'interno del modello, quella che van Fraassen definisce una *razionalità Prussiana*¹⁰⁸: vietiamo tutto ciò che non è esplicitamente permesso. Questo ci mette nella condizione di rendere i modelli causali uno strumento rigoroso, come dovrebbero essere. Il pragmatico e contestuale viene, infatti, determinato prima che il modello entri in scena. Una volta che siamo al cospetto del modello, solo ciò che è intrinseco e strutturale conta.

¹⁰⁷ Halpern e Hitchcock, «Graded causation and defaults» cit.

¹⁰⁸ B. C. van Fraassen, *Laws and Symmetry*, Oxford University Press, 1989: 171–172.

Si prenda l'**Esempio Dolo o colpa? Colpa**. In questo caso, facendo agire l'ordine di normalità descritto da **Halpern** al di fuori della definizione di causalità, ed esclusivamente rispetto al modello, otteniamo il seguente grafo causale relativamente al quale AN è la causa attuale.



Questa procedura oltre a fornire una sorta di algoritmo per ricercare l'accordo su di un modello causale, permette di fare una esplicita distinzione tra l'esclusione di variabili: se considerate non causalmente rilevanti, si caratterizzeranno includendole nella segnatura \mathcal{S} ma escludendole dalle equazioni strutturali; se considerate, invece, rappresentare eventualità troppo remote e irragionevoli, questa decisione verrà rappresentata dalla forte penalità di normalità associata con una loro differenza rispetto allo stato di default.

UNA NATURA, DIVERSE FUNZIONI: SUL PLURALISMO CAUSALE

UN PICCOLO AMMONIMENTO Un'estensiva studio di coorte svolto dai (più o meno) rigorosissimi scienziati dello stato di Correlandia ha messo in luce che il 100% della popolazione che dichiarava di non essere stata visitata da un dottore negli ultimi 10 anni, dichiarava anche di non avere un tumore¹. Lo studio ha inoltre evidenziato che i pazienti che, invece, effettuavano periodici check-up medici erano più propensi ad ammalarsi di tumore: fra questi, infatti, coloro che riportavano un tumore erano ben l'1,7%.

Lo studio crea un enorme scalpore e, di lì a poco, i dottori dello stato di Correlandia non riescono a trovare un paziente neanche offrendo gratuitamente i loro servizi, dato che il governo spinge i cittadini a non farsi visitare, vista la forte correlazione evidenziata dal famigerato studio. Esattamente come ci si aspettava, negli anni successivi, il numero di pazienti con tumori si riduce enormemente. Medaglie di merito piovono sui petti degli scienziati che progettaron lo studio, e tutto sembra procede per il meglio con il nuovo piano sanitario nazionale.

Correlandia, si ricordi, è uno stato che tiene molto alla salute dei suoi cittadini: quando si inizia a notare un aumento della mortalità nella popolazione, promulga una legge che obbliga ad effettuare l'autopsia di ogni soggetto deceduto, in modo da comprendere come indirizzare la problematica (dopotutto bisognerà pur far lavorare questi medici nullafacenti!). Contro ogni previsione, gli esiti autoptici che riportano tumori sono in numero considerevolmente maggiore rispetto a quello che ci si aspetterebbe. Gli scienziati di Correlandia sono perplessi: la correlazione tra il non andare dal medico e l'assenza di tumori era marcata. Come mai, incentivando le persone a non andare dal dottore, i casi di tumore rilevati in pazienti deceduti sono enormemente aumentati?

Perché non andare dal dottore per evitare un tumore è diverso rispetto allo smettere di fumare e svolgere regolare attività fisica rispetto allo stesso obiettivo? Perché tutte le teorie sulla causalità distinguono attentamente tra causalità e mera correlazione? Perché questo contrasto non è solo una questione statistica e matematica, bensì riguarda la nostra vita di tutti i giorni, ed in questa si esplica costantemente? Il motivo è che, costantemente, nel nostro agire nel mondo, cerchiamo di produrre effetti, realizzare obiettivi, tutelarci contro potenziali esiti di situazioni. Per fare questo, ci serve uno strumento che ci dica come e, soprattutto *dove* agire. Questo strumento è la causalità.

La differenza fondamentale tra le correlazioni e le relazioni causali è che le seconde possono essere usate ai fini di *produrre* o *determinare*² stati, esiti o eventi. Le relazioni causali ci offrono una base per *manipolare* il mondo, in una maniera che le correlazioni non possono fare. Il criterio che utilizziamo per discriminarle è esattamente questo: le mere correlazioni non sempre supportano

¹ Esempio adattato da Pearl, *Causality: Models, Reasoning, and Inference* cit.

² Pur non essendo equivalenti, uso questo termine per indicare quello che in inglese viene reso con *bring about*.

modifiche nel comportamento di un correlato obiettivo, dati interventi che modificano l'altro correlato; le relazioni causali, invece, per la loro intrinseca natura, supportano modifiche dell'effetto ogni qualvolta si manipolino le sue cause³.

4.1 L'APPROCCIO FUNZIONALE

Da queste caratteristiche emerge un rapporto di tipo funzionale⁴ che gli umani hanno con la nozione di causa. Il fatto che noi esprimiamo questa modalità di interazione con la causalità dovrebbe guidare il modo in cui noi *indaghiamo* la causalità. Seguendo Woodward⁵, la nostra cognizione causale è votata ad informare un comportamento guidato da obiettivi, essa è una *tecnologia epistemica* che svolge un compito, e come tale va analizzata.

L'approccio funzionale si pone due obiettivi: uno *descrittivo* e l'altro *normativo*. Questi due obiettivi non vanno interpretati come separati, bensì come strettamente interconnessi: apprendere il modo in cui produciamo giudizi causali in determinate situazioni può aiutarci a stabilire dei criteri normativi. Si prenda lo studio di Lombrozo, «Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions». Lo studio si focalizza, in parte, sui giudizi causali di alcuni soggetti esaminati riguardo casi di doppia prevenzione⁶. La caratteristica principale di questi casi è che non coinvolgono trasferimenti di quantità. Nella terminologia di Hall, non si tratta di casi di *produzione*, ma di casi di *dipendenza*⁷. Ebbene, Lombrozo ha concluso che alcuni casi di doppia prevenzione vengono ritenuti più paradigmatici di altri, e che questi casi sono quelli che riguardano fenomeni naturali, percepiti quindi come più stabili.

L'approccio funzionale trae da questo aspetto descrittivo un aspetto normativo:

non è misterioso il perché persone ordinarie e scienziati dovrebbero dare maggior valore a relazioni che sono più stabili[...] sono più generalizzabili o proiettabili a nuove situazioni. I giudizi delle persone riguardo ai casi di doppia prevenzione hanno senso, o sono razionali, dato questa preoccupazione per la stabilità. In questo senso, è normativamente o metodologicamente appropriato per entrambi, scienziati e persone comuni, avere i giudizi che hanno.⁸

Una conseguenza a mio parere obbligata dell'aderire all'approccio funzionale è lo sposare una visione pluralista della causalità⁹. Questa affermazione può essere intesa in diversi sensi. Potrebbe sembrare, in un senso concettuale, un truismo. Dopotutto vi sono diversi tipi di concetti di causalità: quello impiegato nella fisica è radicalmente differente da quello utilizzato, per esempio, da uno

³ Naturalmente, *modulo* ragionevoli assunzioni.

⁴ Danks, «Functions and Cognitive Bases for the Concept of Actual Causation» cit.; Woodward, «A Functional Account of Causation: or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)» cit.

⁵ Woodward, «A Functional Account of Causation: or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)» cit.

⁶ Un esempio di doppia prevenzione è quello presentato nell'Esempio Doppia prevenzione.

⁷ Hall, «Two Concepts of Causation» cit.

⁸ Woodward, «A Functional Account of Causation: or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)» cit.: 705.

⁹ C. Hitchcock, «On the Importance of Causal Taxonomy», in *Causal Learning. Psychology, Philosophy and Computation*, a cura di A. Gopnik e L. Schulz, Oxford University Press, 2007; C. Hitchcock, «How to be a Causal Pluralist», in *Thinking About Causes. From Greek Philosophy to Modern Physics*, a cura di P. Machamer e G. Walters, University of Pittsburgh Press, Pittsburgh 2007.

psicologo sociale; quest'ultimo è diverso da quello utilizzato da un neuroscienziato, che è differente da quello utilizzato da un economista. Un altro senso in cui si può intendere questa affermazione è un senso *ontologico*, secondo cui vi sono diversi tipi di causalità: una statistica, una meccanica, una controfattuale, ecc.

Un altro senso è quello proposto da **Hitchcock**: il «pluralismo intramurale»¹⁰. Questo consiste nel sostenere la presenza di una varietà di concetti causali, definiti tutti all'interno della medesima cornice metodologica. Nel campo dei modelli causali, questa non è una posizione infrequente, anzi, sembrerebbe maggioritaria: **Woodward** definisce i concetti di *causa totale, diretta, contributiva*; **Hitchcock** parla di *effetto netto e effetto componente*; **Pearl** presenta una varietà estrema di concetti causali. Vediamo un esempio paradigmatico in cui queste distinzioni vengono alla luce in maniera preponderante: il caso della pillola contraccettiva discusso da **Hesslow**¹¹:

ESEMPIO 4.1.1 (Pillola contraccettiva e trombosi: causa in che senso?)

La pillola contraccettiva è uno dei metodi più efficaci nella prevenzione della gravidanza. Le alte dosi di estrogeni (o equivalenti), comportano, però, un aumento del rischio di trombosi non indifferente. Di conseguenza, potremmo dire che prendere la pillola contraccettiva causi trombosi. Anche la gravidanza, però, comporta un aumentato rischio di trombosi. Dato che la pillola contraccettiva riduce grandemente la probabilità di gravidanza (nelle donne che hanno rapporti sessuali con uomini), essa abbassa la probabilità complessiva di andare incontro a trombosi in questa classe di popolazione. Sembra allora, che la pillola contraccettiva sia un fattore preventivo rispetto alla trombosi, e non una sua causa. ■

In **Figura 4.1** abbiamo una rappresentazione della situazione.

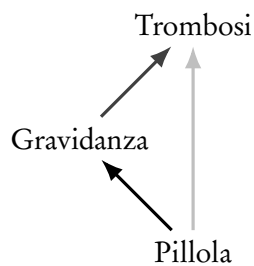


Figura 4.1: Grafico che rappresenta le relazioni descritte dall'**Esempio 4.1.1**, con l'intensità di colore a rappresentare la forza dell'effetto.

Quello che emerge da questa situazione è che la pillola contraccettiva ha diversi effetti: essi si esplicano lungo i differenti percorsi di influenza causale che vanno dall'assunzione della pillola allo sviluppo della trombosi. Per discriminare questi effetti sono necessari diversi concetti afferenti al concetto di causalità. Questa è la funzione che svolgono, ad esempio, i concetti di *causa contributiva* ed *effetto componente*: essi rappresentano e misurano gli effetti di una variabile su di un solo percorso causale. In questo caso l'assunzione della pillola ha un effetto (*diretto*, per usare un'ulteriore specificazione utilizzata da **Woodward**) che promuove leggermente la trombosi nel percorso non interpolato dalla gravidanza; sul percorso interpolato dalla gravidanza, la pillola ha un effetto fortemente preventivo sulla gravidanza e quindi sulla trombosi. L'effetto *netto* o *totale*, osservando che le

¹⁰ **Hitchcock**, «How to be a Causal Pluralist» cit.

¹¹ G. Hesslow, «Two Notes on the Probabilistic Approach to Causality», *Philos. of Sci.*, vol. 43, n. 2 (1976), pp. 290-292, DOI: [10.1086/288684](https://doi.org/10.1086/288684).

donne appartenenti alla classe di interesse che prendono la pillola sono meno soggette a trombosi rispetto a quelle che non la prendono, è preventivo.

Ma perché questi concetti? Perché non semplicemente chiamare tutto una causa, e poi specificarne le caratteristiche? Io ritengo che l'utilità di questi concetti sia spiegata dal paradigma funzionale: essi svolgono diverse funzioni e, seguendo Hitchcock e Knobe, ci aiutano ad individuare specifiche modalità di intervento^{12,13}. Il classico test controfattuale, ad esempio, non sempre ritorna il giusto tipo di intervento da compiere al fine di manipolare il sistema nel senso desiderato, ed, anzi, a volte non è in grado di restituire nemmeno un target¹⁴. Ma noi conosciamo sempre, o quanto meno molto spesso, l'aspetto adeguato su cui intervenire al fine di modificare la situazione verso un obiettivo specifico. Cognitivamente, il possedere una vasta gamma di concetti causali ci permette di discriminare situazioni in cui, al fine di ottenere l'esito desiderato, sono necessarie diverse manipolazioni del sistema sotto esame. In particolare, diversi concetti di causa svolgono la funzione di individuare un *target* di intervento specifico. Si analizzi il seguente esempio.

ESEMPIO 4.1.2 (Il Miglior Impiegato Del Mese)

Impiegato è un tipo fannullone: al lavoro sonnecchia ed è spesso in ritardo con le consegne dei suoi progetti. Per sua sfortuna, è anche una persona molto orgogliosa. Questo mese si è convinto di doversi adoperare per essere eletto impiegato del mese. Il modo più comodo per ottenere il suo scopo è corrompere chi assegna il titolo. ■

Come dovrebbe comportarsi Impiegato? Nella fattispecie, se dovessimo guidarne le azioni (trascurando, naturalmente, l'immoralità della situazione) che cosa suggeriremmo? Innanzitutto, per prima cosa, vorremmo sapere di più su come si assegna il titolo. Questa domanda ha un chiaro fine: individuare il target di intervento adeguato, al fine di far nominare Impiegato "impiegato del mese": infatti, la nostra risposta sarà differente nei seguenti tre casi:

IC1 Il premio è assegnato da una commissione.

IC2 Il premio è assegnato da una commissione, ma il caporeparto può ribaltare l'esito della decisione.

IC3 Il premio è assegnato dal caporeparto, ma poi sottoposto al caposezione, che può apporre un veto alla nomina.

Le risposte sono differenti, perché differenti sono le strutture causali coinvolte in questi esempi e differenti sono i concetti causali che utilizziamo per trattare la situazione ed individuare una manipolazione adeguata. Ora si tratta di analizzare la componente descrittiva, al fine di produrre quella normativa. Individuando il concetto di causa adeguato che viene applicato ad ogni caso, siamo in grado di produrre uno strumento concettuale e metodologico in grado di guidare i nostri interventi in maniera rigorosa; o, quantomeno, più rigorosa rispetto all'esserne privati.

È mia intenzione, nella prossima sezione, proporre alcuni concetti causali che catturino queste differenti selezioni cognitive di target di intervento. Due sono i concetti che ritengo più interessanti tra quelli che propongo: i blocchi causali e le cause per scambio.

¹² Hitchcock e Knobe, «Cause and Norm» cit.

¹³ Parzialmente a riguardo, si veda anche S. A. Sloman e D. A. Lagnado, «Do We “do”?», *Cognitive Science*, vol. 29, n. 1 (2005), pp. 5-39, DOI: [10.1207/s15516709cog2901_2](https://doi.org/10.1207/s15516709cog2901_2).

¹⁴ Si pensi ai casi di preemption.

4.2 ALCUNI CONCETTI CAUSALI

In questo capitolo introdurrò due concetti causali che risultano utili nel guidare i nostri interventi. Questi concetti saranno sempre definiti relativamente ad un modello causale. Il primo si basa sull'idea di un *blocco causale*, ovvero un cammino direzionato caratterizzato da una sufficienza forte¹⁵ ed alcune sue variazioni. L'idea è che l'agire su questo tipo di cammini *blocca* l'esito, nel senso che lo garantisce¹⁶. L'idea è parzialmente ispirata ad una proposta di Halpern¹⁷. Il secondo concetto, invece, cattura la modalità di intervento adatta in quei casi in cui vi sono potenzialmente molteplici processi in grado di produrre l'effetto in questione. Questo concetto ha, più che un'utilità di produzione, un'utilità preventiva: ci aiuta ad evitare esiti quando questi sono sovradeterminati. Non si tenterà un approccio alla Halpern e Hitchcock, ovvero non si stabilirà nessuna nozione di *causalità per grado*, anche se, in questa definizione, potrebbe essere utile in contesti legali.

4.2.1 La causalità per blocco

In questo capitolo verranno presentati quelli che io definisco *blocchi causali*. L'obiettivo di questo concetto è catturare quei casi in cui si è in grado di bloccare, nel senso di garantire, un esito attraverso un unico intervento.

La più interessante applicazione di questo concetto si ha in quei casi in cui compiamo azioni preventive, volte ad evitare danneggiamenti futuri ad esempio¹⁸. Utilizzerò “causa”, tra virgolette, in quanto la relazione che stiamo analizzando è sì legata alla causalità, ma per vie traverse e solo sotto determinate condizioni. Ritengo che concettualizzare correttamente questa tipologia di “causalità”, che presenta comportamenti, per così dire, *alla rovescia*, sia necessario in quanto risulta fondamentale in situazioni in cui vi è il bisogno di attribuire colpe e responsabilità. Come sostenuto lungo tutta la trattazione precedente, ritengo che la nozione di causa e, quindi, con essa, le nozioni causali, svolgano un ruolo di primaria importanza nella formulazione dei nostri giudizi morali e nell'attribuzione di responsabilità¹⁹ e, di conseguenza, risulta, a mio parere, necessaria la produzione e l'esplorazione di un apparato concettuale atto ad affrontare questo tipo di problematica.

È, quindi, nell'ottica di questo progetto che si inserisce la successiva trattazione. Di seguito una breve presentazione delle sezioni che succederanno.

Nella Sezione 4.2.1.1 mi occuperò di discutere alcuni casi problematici che, a mio parere, sono trattati con estrema facilità una volta che ci forniamo di una concettualità adatta, come, ad esempio, il concetto di blocco causale.

Nella Sezione 4.2.1.2 presenterò una soluzione alle problematiche trattate nella sezione precedente, anche appoggiandomi ad alcune idee accennate da Menzies²⁰, che però non riconosce la

15 Originariamente questo particolare concetto causale doveva chiamarsi *locking cause*, con l'idea di trasmettere un senso di *chiudere a chiave l'esito*. Per questioni di uniformità linguistica, si è optato per la dicitura «blocco causale», che pur perdendo la caratterizzazione originale, ne mantiene, di fondo, lo spirito.

16 Garantendo un esito, evita, naturalmente, l'altro. È importante tenere a mente questa simmetria, in quanto molte di quelle che io definisco istanze di blocchi causali non sopravviverebbero ad un'analisi basata sui default proprio per questo aspetto: potrebbero non modificare la situazione in termini di normalità.

17 Oltre a Halpern, «Defaults and Normality in Causal Structures» cit.: 8, anche Halpern, *Actual Causality* cit.: 54, anche se strettamente più debole della mia proposta, nonché entrambe differenti in spirito.

18 Quello che in inglese viene definito *safeguarding*.

19 Si veda, ad esempio: C. Sartorio, «A New Asymmetry Between Actions And Omissions», *Noûs*, vol. 39, n. 3 (2005), pp. 460-482, DOI: <https://doi.org/10.1111/j.0029-4624.2005.00509.x>.

20 Menzies, «The Problem of Counterfactual Isomorphs» cit.

problematica e, quindi, la aggira. In particolare mi rifarò all'idea di *enabling condition*, però rovesciandola. Sosterrò, quindi, che alcuni blocchi causali producono cause che si configurano come il rovescio di un'istanza di causalità per omissione dove è necessario che si presenti non una *enabling condition*, bensì che la causa in questione si comporti come una *disabling condition*, una condizione disabilitante. Argomenterò, infine, che quando una condizione disabilitante è attiva e contrastata, abbiamo causalità per blocco, e che l'attività della condizione disabilitante non vada considerata come una causa, ma vada considerata tale solo in certune contingenze. Infine, presenterò alcuni casi pratici che mostrano perché la nozione sia rilevante e ne discuterò l'utilità pratica.

4.2.1.1 *Casi problematici*

La causalità per blocco è, in una certa maniera, simile alla causalità per disconnessione²¹: invece di trattarsi di casi in cui un certo meccanismo si disconnette, producendo un effetto causale, si tratta di casi in cui un evento instaura un processo causale duraturo e 'dormiente', costantemente connesso ma inattivo, pronto a diventare una causa nel momento in cui un altro processo contrario si connette e cerca di produrre l'effetto contrario. Questo è un fenomeno particolarmente interessante perché, pur trattandosi di un concetto causale, non è un concetto di causa se non in determinate condizioni.

Nel mondo, ci troviamo circondati da cause di questo tipo: assicurazioni, cinture di sicurezza, parapetti, ecc. In generale, questi svolgono il loro ruolo causale, per così dire, nell'ombra: sono quasi sempre processi attivi, ma non *effettivi*, non hanno alcuna relazione diretta con lo scorrere degli eventi salienti nella situazione, non intrattengono relazioni di tipo controfattuale diretto e non svolgono, apparentemente, alcun ruolo. Se, riprendendo la **Definizione Causa effettiva^m** considerassimo le contingenze nel contesto attuale, questi processi *bloccanti* non svolgono alcun ruolo e passano inosservati.

La salienza di questo tipo di cause viene alla luce solo nel momento in cui sono contrastata da processi che, se lasciati indisturbati, modificherebbero l'esito della situazione: immaginiamo la seguente situazione:

ESEMPIO 4.2.1

Un soggetto malato di cuore sul quale è stato effettuato un impianto di pacemaker. Il pacemaker invia, a intervalli di tempo regolari, un impulso elettrico che fa battere il cuore regolarmente, anche se l'esito, non fosse presente l'impulso prodotto dal pacemaker, sarebbe il medesimo e, quindi, il ritmo cardiaco resterebbe invariato. Ad un certo punto il cuore del soggetto sta per andare in crisi e il ritmo cardiaco sta per farsi irregolare. Il pacemaker invia una scarica proprio in quel momento e stabilizza il battito cardiaco all'usuale ritmo. ■

Pur essendo un processo "causale" (si noti, infatti, il trasferimento di energia costante tra il pacemaker e il cuore del soggetto), la scarica elettrica prodotta dal pacemaker non è in alcuna relazione di dipendenza controfattuale con il ritmo cardiaco, fino a quando questo non devia dalla sua norma. Il processo, pur perdurando nel tempo, diventa attivo solo nel momento in cui è minacciato. Queste cause si comportano come delle *anti-condizioni* nel background causale di una situazione: non hanno alcuna salienza nella situazione e quello che fanno è, semplicemente, impedire che una certa situazione si verifichi. Ma nel momento in cui svolgono la loro funzione sembrano assurgere al rango di causa.

²¹ Schaffer, «Causation by Disconnection» cit.

Questo comportamento peculiare è riscontrabile in diversi esempi discussi in letteratura. Si veda, tra gli altri, l'**Esempio Bogus Prevention**: in questo caso l'antidoto garantisce la sopravvivenza di Vittima, ma abbiamo fatto di tutto pur di non farlo figurare come causa della sopravvivenza. Nel simile **Esempio Avvelenamento Prudente**, invece, vediamo che l'antidoto svolge il suo ruolo causale, in quanto blocca l'azione del veleno. La differenza tra le due situazioni non è rappresentata dalla non presenza dell'antidoto, bensì dalla presenza o assenza di un processo che vada a tentare di interferire con l'esito che l'antidoto vuole garantire.

Un altro esempio, discusso da **Hart e Honoré**²² è il seguente:

ESEMPIO 4.2.2 (Carestia)

Una terribile siccità si abbatte sull'India. Fa seguito una terribile carestia. Il governo, poco previdente, non ha immagazzinato sufficienti scorte per far fronte a questa evenienza. ■

Chiaramente, anche in questo caso, ci troviamo di fronte ad un esempio di causalità di blocco: se il governo avesse immagazzinato le scorte, non ci sarebbe stata la carestia. Si noti la curiosa asimmetria per cui, da un lato, compiere l'azione non sia causa di un esito, dall'altro ometterla possa renderla tale.

Un altro tipo di causalità di blocco si presenta quando, invece, la nostra presunta causa è attiva e svolge il suo ruolo causale solo in virtù di questo fatto. Naturalmente, anche qui, garantisce un esito.

ESEMPIO 4.2.3 (Virus)

Il *virus varicella-zoster* (**VVZ**), una volta infettato un soggetto, può potenzialmente causare la varicella. Anche dopo la soluzione dell'infezione attraverso il lavoro del sistema immunitario, il virus permane dormiente nei neuroni del soggetto per tutta la vita, dove può riattivarsi. ■

In questo caso, l'infezione da **VVZ** garantisce la presenza del virus all'interno dei neuroni in maniera permanente. Si noti l'assenza di condizioni abilitanti al fine della garanzia dell'esito, differenza rispetto al caso precedente.

Ora che il fenomeno è stato individuato, propongo una definizione nella prossima sezione.

4.2.1.2 Una proposta di causalità per blocco

Ciò che vogliamo catturare attraverso la definizione della nozione di causa di blocco è stato esplicitamente chiarito: l'idea di manipolare qualcosa al fine di garantire un esito. La formalizzazione di questo concetto ci è utile in quanto andrà a svolgere un ruolo normativo e metodologico all'interno della cornice funzionale: vuole dare uno strumento atto a guidarci in determinate situazioni in cui ci troviamo di fronte a cause di questo tipo.

Il primo passo per la formalizzazione di un concetto di causa di blocco è la definizione di un *blocco causale*.

DEFINIZIONE 4.2.1 (BLOCCO CAUSALE)

Un blocco causale per $Y = y$, denominato $\vec{R}_{Y=y}$, in un modello causale M è una tupla $\langle I, < \rangle$, con $I = \{V_0: X_0 = x_0, \dots, V_n: Y = y\}$, con $<$ rappresentante la relazione di dipendenza diretta forte in M e I tale che, per tutti i $V_i, V_{i-1} < V_i$. Definiamo l'elemento V_0 la testa del blocco e l'elemento V_n la coda del blocco. Indichiamo con $\text{Var}(\vec{R})$ le variabili che compongono \vec{R} .

²² Hart e Honoré, *Causation in the law* cit.: 33-34.

Dove la relazione di dipendenza diretta forte è definita nel seguente modo:

DEFINIZIONE 4.2.2 (DIPENDENZA DIRETTA FORTE)

$X = x$ è in una relazione di dipendenza diretta forte con $Y = y$ in M sse, sia $\vec{W} = \mathcal{V} \setminus \{X, Y\}$, per ogni $w \in \mathcal{R}(\vec{W})$ e per ogni $\vec{u} \in \mathcal{R}(\mathcal{U})$:

$$1. (M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}] Y = y$$

In parole, un blocco causale è un insieme di eventi²³ tale per cui ogni elemento del blocco causale ha un rapporto di dipendenza diretta forte con l'elemento immediatamente successivo all'interno del modello causale, dove con dipendenza diretta forte si indica una relazione tra due eventi all'interno di un modello causale tale per cui l'imposizione del primo evento implica necessariamente il secondo, indipendentemente dal contesto \vec{u} e dalle altre variabili in \mathcal{V} escluse le due variabili in questione.

Questa definizione è equivalente a dire che tra la testa e la coda del blocco vi è un direzionato caratterizzato da sufficienza forte, una relazione piuttosto rara. Si consideri, infatti, il seguente caso:

ESEMPIO 4.2.4

Durante un'assemblea viene avanzata, da due membri dell'assemblea, una proposta, che passerà solo se otterrà la maggioranza dei voti. L'assemblea è proposta da 5 membri. Si dà per scontato che chi avanza una proposta la vota. ■

Modellando questa situazione in maniera completa, avremo cinque votanti P_1, P_2, V_1, V_2, V_3 e una variabile binaria per l'esito E . Si veda che nessun voto è, di per sé, in grado di formare un blocco causale $\vec{R}_{E=1}$, in quanto nessun singolo voto può mai essere sufficiente quando gli altri quattro sono liberi di variare. Nonostante ciò, nella situazione attuale, in cui $\vec{u} = \{P_1 = 1, P_2 = 1\}$, ognuno dei votanti può formare un blocco causale. La nozione risulta, quindi, troppo forte per rappresentare correttamente situazioni in cui, essendo dato un contesto \vec{u} , qualche variabile è in grado di instaurare un blocco causale. Per affrontare adeguatamente questi casi abbiamo bisogno di una definizione di blocco causale più debole: una definizione *condizionale*, o *relativa*.

DEFINIZIONE 4.2.3 (DIPENDENZA DIRETTA CONDIZIONALE)

$X = x$ ha una dipendenza diretta condizionale con $Y = y$ in M se esiste un $\vec{W} \subseteq \mathcal{V} \setminus \{X, Y\}$ e un $\vec{w} \in \mathcal{R}(\vec{W})$, tali che \vec{w} non contiene blocchi causali per $Y = y$ e che, se $\vec{Z} = \mathcal{V} \setminus \{X, Y, \vec{W}\}$ allora per ogni $\vec{z} \in \mathcal{R}(\vec{Z})$ ed ogni $\vec{u} \in \mathcal{R}(\mathcal{U})$, abbiamo:

$$(M, \vec{u}) \models [X \leftarrow x, \vec{W} \leftarrow \vec{w}, \vec{Z} \leftarrow \vec{z}] Y = y.$$

DEFINIZIONE 4.2.4 (BLOCCO CAUSALE CONDIZIONALE)

Un blocco causale condizionale per $Y = y$, denominato $\vec{R}_{Y=y|\vec{w}}$, in un modello causale M è una tupla $\langle I, <, \vec{w} \rangle$, con $I = \{V_0: X_0 = x_0, \dots, V_n: Y = y\}$, con $<$ rappresentante la relazione di dipendenza diretta condizionale in M rispetto alle condizioni \vec{w} , I tale che, per tutti i V_i , $V_{i-1} < V_i$ e \vec{w} è minimale, ovvero nessun $\vec{w}' \subset \vec{w}$ soddisfa la presente definizione.

Il blocco causale condizionale è un blocco causale in cui ogni elemento ha dipendenza diretta condizionale con il precedente nel medesimo \vec{w} , con \vec{w} minimale. Questo \vec{w} può essere considerato l'insieme delle condizioni abilitanti per blocco causale.

Ora presentiamo le seguenti definizioni:

²³ In questo contesto definirò un evento come l'ottenere un determinato valore di una variabile, ad esempio $X = x$.

DEFINIZIONE 4.2.5 (SALVAGUARDIA)

$X = x$ svolge azione di salvaguardia rispetto ad $Y = y$ se $X = x$ fa parte di un blocco causale $\vec{R}_{Y=y}$ e $(M, \vec{u}) \models [X \leftarrow x']Y = y$.

Questa è una definizione formale del concetto di salvaguardia. Nella vasta letteratura sulla causalità e i modelli causali, sono estremamente sporadiche²⁴ le tracce di questa nozione, che pur mi sembra fondamentale a livello cognitivo e funzionale, oltre che ad essere una nozione chiaramente afferente al dominio della causalità. Naturalmente, questo non significa che salvaguardare sia una forma di causa. Questo è, ad esempio, fortemente negato da Hall²⁵.

Nonostante questo, è evidente che questa nozione sia connessa alla causalità in quanto un'omissione di salvaguardia viene classificata come una causa di un esito oltre che dalla maggior parte delle persone, anche dalle nostre definizioni di causalità²⁶. Si pensi all'esempio **Esempio Carestia**: l'omissione, da parte del governo, di fare scorte di derrate alimentari, è una causa della carestia²⁷. Secondo la nostra definizione di salvaguardia, e assumendo che il modello preveda che in questo modo la carestia venga evitata, il fare scorta di derrate alimentari è un caso di salvaguardia.

Se, in certi contesti, le omissioni di salvaguardia sono delle cause, l'ottenere delle salvaguardie deve essere causa di un effetto differente, facente parte della classe di contrasto della variabile che rappresenta l'effetto²⁸. Di seguito la definizione di causa per salvaguardia:

DEFINIZIONE 4.2.6 (CAUSA EFFETTIVA PER SALVAGUARDIA)

Sia $X = x$ una salvaguardia di $Y = y$ in un contesto $\vec{u}' \in \mathcal{U}$. $X = x$ è una causa effettiva per salvaguardia di $Y = y$ sse:

1. $(M, \vec{u}) \models X = x \wedge Y = y$
2. $(M, \vec{u}) \models [X \leftarrow x'] \wedge Y \neq y$

Che, a parole, significa che una salvaguardia viene promossa a causa solo nel momento in cui un processo causale rivale è attivo. Questa definizione collima con le nostre intuizioni, ovvero con il fatto che, quando non contrastata, una salvaguardia non ha alcun ruolo causale nella situazione, ma rappresenta un processo che, causalmente, ha «sparato a vuoto». Invece, in quei casi in cui il colpo va a segno, ovvero quando la salvaguardia tutela un esito minacciato da un processo avverso, allora in quel caso essa viene promossa al rango di causa. Definite queste nozioni, procediamo alla discussione.

4.2.1.3 Sull'utilità dei concetti di blocco causale e di causa attuale per blocco

I concetti introdotti nella sezione precedente si dimostrano utili nel catturare aspetti cognitivi delle nostre interazioni rispetto alle situazioni che i modelli nei quali questi concetti si verificano

²⁴ In tutta la bibliografia consultata per questo scritto appare 36 volte di cui 16 solo in R. W. Wright, «Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts», *Iowa L. Rev.*, vol. 1001, n. 73 (1988), https://scholarship.kentlaw.iit.edu/fac_schol/698 e R. W. Wright, «The NESS Account of Natural Causation: A Response to Criticisms», in *Critical Essays on "Causation and Responsibility"*, DE GRUYTER, 2013, pp. 13-66, DOI: 10.1515/9783110302295.13.

²⁵ Hall, «Structural Equation and Causation» cit.: 120.

²⁶ Si veda anche, per una discussione dal punto di vista morale: Sartorio, «A New Asymmetry Between Actions And Omissions» cit.

²⁷ Naturalmente, questo vale se si considerano le omissioni cause. Io sono di questa opinione, e ritengo che la letteratura suggerisca che questa sia la posizione più ragionevole. Si veda, in ogni caso, la nota 83.

²⁸ Se la variabile rappresenta un evento binario, la salvaguardia sarà una causa dell'evento opposto.

rappresentano. Ritengo, soprattutto, che la definizione di salvaguardia, che, a quanto mi risulta, non ha precedenti in letteratura, sia specialmente importante: essa infatti può rendere conto, all'interno della cornice metodologica dei modelli causali, di tutta una serie di decisioni e manipolazioni che i soggetti più disparati compiono quotidianamente.

Ad esempio, ci risulta normale, nelle situazioni in cui si verificano, effettuare scelte di manipolazione volte al far ottenere salvaguardie. Oltre che ad impegnarci attivamente per l'ottenersi di queste situazioni, ci aspettiamo che gli altri soggetti compiano questo tipo di scelte, e troviamo moralmente reprimibile mancanze in questo senso.

Torniamo per un attimo al nostro **Esempio Il Miglior Impiegato Del Mese**. Il nostro concetto di blocco causale ci fornisce una chiara risposta riguardo all'intervento da effettuare nel **Caso IC₂**: il caporeparto fa parte di un blocco causale che determina l'esito finale della decisione, di conseguenza il nostro intervento andrà in questa direzione: consiglieremmo ad Impiegato di corrompere²⁹ il caporeparto, e non di corrompere i partecipanti all'assemblea.

Oppure, si immagini di essere un *policy-maker* di fronte ad una decisione complessa rispetto ad una situazione che, una volta modellata, è composta da un gran numero di variabili: non sarebbe un buon obiettivo adoperarsi al fine di individuare un blocco causale che realizzi il nostro scopo? Oppure, non sarebbe equamente importante individuare una situazione che produce un blocco causale nel senso opposto? Si immagini la situazione descritta dall'**Esempio Virus**, solo che al posto del **VVZ**, ci troviamo di fronte ad un virus differente, potenzialmente letale, che, invece che rimanere dormiente nei neuroni, li danneggia permanentemente, anche nel caso il paziente riesca a guarire. Se l'infezione fosse parte di un blocco causale e, quindi, determinasse con sicurezza un danneggiamento del sistema nervoso, il *policy-maker* saprebbe di dover investire sulla ricerca di un vaccino in misura maggiore rispetto ad una cura, in quanto un vaccino, al netto di altre considerazioni, permetterebbe di evitare danni neurologici permanenti prevenendo l'infezione.

Si noti che, in tutte le mie definizioni, non ho utilizzato notazione vettoriale per X . Questo per due motivi: il primo è che, in questo modo si enfatizza il fatto che con un solo intervento siamo in grado di propagare una grande influenza lungo il modello causale. In un certo senso, l'utilizzo della singola variabile X rappresenta la potenza di questo tipo di variabili e la loro grande efficacia, e, quindi, evidenzia il motivo per il quale siamo interessati alla manipolazione di quest'ultime in particolare.

L'altro motivo è che le definizioni sono state pensate sempre riferendosi alla testa del blocco causale. In realtà questa non è una necessità, dato che un intervento che crei una situazione in cui uno qualsiasi degli eventi $c \in \mathcal{I}$ è realizzato comporta l'efficacia immediata del blocco causale. In realtà, ritengo che, quando possibili, interventi più prossimi alla variabile che rappresenta il nostro obiettivo siano da preferire in quanto, in un certo senso, più stabili³⁰.

Presenterò ora, più rapidamente, alcune nozioni utili per affrontare casi simili rispetto a quello che l'Impiegato incontrerebbe nel **Caso IC₃**.

²⁹ Nuovamente, astenendoci dal valutare gli aspetti morali della situazione.

³⁰ Huber, ad esempio, è di questa mia stessa idea: Huber, «*Structural Equations and Beyond*» cit.

4.2.2 Più cause: cause per scambio

La nozione di causalità di scambio cattura la corretta modalità di interazioni con quei casi e quelle strutture causali nelle quali non vi è un singolo intervento in grado di far accadere l'esito a cui si mira. Questi casi sono estremamente discussi in letteratura e hanno fornito controesempi a diverse teorie della causa. In questo scritto, ne sono riportati molti.

Si tratta di quei casi, come l'**Esempio Plotone di Esecuzione** e l'**Esempio Rock Throwing**, in cui vi sono, simultaneamente attivi, o, nel caso di Esempio, potenzialmente attivi, più processi causali che realizzano o producono l'effetto. Diamo una definizione formale al fenomeno:

DEFINIZIONE 4.2.7 (CAUSE PER SCAMBIO EFFETTIVE)

Siano $\vec{C}_S \subset \mathcal{V}$, con $|\vec{C}_S| > 1$ e $\vec{c}_s \in \mathcal{R}(\vec{C}_S)$ e \vec{c}_s^{neg} tale per cui per ogni $c \in \vec{c}_s$, $c^{\text{neg}} = c'$, allora $\vec{C}_S = \vec{c}_s$ viene detto un insieme di cause per scambio sse, per ogni $X \in \vec{C}_S$ e $x \in \vec{c}_s$:

$$CS1 \quad (M, \vec{u}) \models [X \leftarrow x']\varphi$$

$$CS2 \quad (M, \vec{u}) \models [X \leftarrow x', \vec{C}_S \leftarrow \vec{c}_s^{\text{neg}}]\neg\varphi$$

$$CS3 \quad \vec{C}_S \text{ è minimale.}$$

Ogni elemento $X \in \vec{C}_S$ e $x \in \vec{c}_s$ viene definito una causa per scambio effettiva.

Questa definizione è, a vista, più complessa di quelle presentate sin ora, ma non così a parole: un insieme di cause di scambio è detto tale se esso è minimale, e se l'evento effetto (φ) si modifica solo nel caso in cui si imponga un valore differente a tutte le variabili all'interno dell'insieme. Un singolo elemento di questo insieme è detto causa per scambio effettiva.

Questo concetto causale ci è utile per trattare situazioni in cui vi sono molteplici percorsi causali che influenzano l'effetto che ci interessa e, per tanto, correggerne uno non è sufficiente ad evitare il verificarsi dell'evento che ci interessa. È, tra l'altro, in grado di catturare efficacemente casi di *switching* simili all'**Esempio Switching** (Switching appunto):

ESEMPIO 4.2.5

Una lampadina si accende se uno dei due interruttori I_1, I_2 è attivo. Se si disattiva I_1, I_2 si attiva. ■

In questo caso, $C_S = \{I_1, I_2\}$ e $c_s = \{1, 1\}$. Di conseguenza, sappiamo che un mero intervento su I_1 non sarà sufficiente a modificare l'esito e far spegnere la lampadina. Invece, fissare entrambi gli interruttori a 0 ci farà ottenere il nostro scopo.

Quello che la **Definizione 4.2.7** dice è che, quando ci troviamo di fronte a questi casi, sono necessari una molteplicità di interventi al fine di intervenire in maniera efficace. Quindi, nel caso in cui Impiegato ci rispondesse che il premio viene assegnato come descritto dal **Caso IC3**, la **Definizione 4.2.7** ci permetterà di individuare il corretto target di intervento, ovvero procedere con la corruzione sia del caporeparto, che del caposezione.

4.3 UNA CONCLUSIONE

Ritengo che la formalizzazione di questo apparato concettuale all'interno della cornice metodologica dei modelli causali la arricchisca parzialmente, oltre che a fornire un contributo, seppur piccolissimo,

alla proposta di approccio funzionale rispetto alla causalità. Senza dubbio, ci fornisce uno strumento con cui valutare in maniera più precisa le situazioni e ci facilita nell'individuazione di obiettivi di intervento fattibili ed efficaci.

Naturalmente, da questa pluralità di concetti causali non segue che il nostro concetto di causa *per sé* sia plurale o che abbia delle sfumature. Quello che viene qui perseguito è un esercizio di chiarificazione concettuale rispetto all'utilizzo cognitivo e pragmatico che del concetto di causa facciamo. Questo significa impegnarsi a catturare la complessa costellazione concettuale che, attorno al concetto di causa si addensa nelle nostre menti e nei nostri comportamenti.

La strada della descrizione del nostro comportamento rispetto alle cause ed alle manipolazioni si incontra, in questo capitolo, con quella della precisazione e formalizzazione di questi concetti ai fini di dargli una struttura ed una applicabilità metodologica. La cornice metodologica dei modelli causali ci fornisce un elegante formalismo che ci permette di ottenere delle definizioni precise ed agevoli, il cui raggio di applicazione può, potenzialmente, essere estremamente vasto: dalle dottrine legali, dove il concetto di blocco causale potrebbe essere rilevante per modellare omissioni di salvaguardie o simili; alle questioni di *policy-making*, dove individuare istanze di causalità per scambio può essere utile al fine di produrre politiche di intervento più efficaci e, quindi, migliori.

5

UNA NUOVA DEFINIZIONE DI CAUSA EFFETTIVA

Questo capitolo, inizialmente non previsto, tratterà brevemente una nuova definizione di causa effettiva che ho formalizzato. Nella **Sezione 5.1** presenterò la definizione ed esporrò, in maniera rapida, le connessioni con le definizioni già discusse. Indirizzerò, inoltre, un problema con un esempio proposto da Halpern (*Incendio^{mod}*), scoperto al seguito dell'applicazione della definizione. Nella **Sezione 5.2** applicherò la mia nuova definizione a qualche esempio, in modo da mostrarne l'efficacia.

5.1 UNA NUOVA CAUSA EFFETTIVA

Durante alcune sperimentazioni con il concetto di blocco causale e, in particolare, con il concetto di blocco causale condizionale (**Definizione 4.2.4**), ho notato, testando alcuni esempi riportati in letteratura, che dalla **Definizione 4.2.4** fosse possibile ricavare un'ottima nozione di causa effettiva differente in spirito da quelle della famiglia **HP**, che risultano, tutto sommato, un po' controintuitive e non immediatamente comprensibili. Nonostante questa leggera differenza, si notano somiglianze con tre proposte: quella di *sustenance* di Pearl¹, quella di *causa forte* di Halpern e Pearl² e quella di causa secondo il test causale *Necessary Element of a Sufficient Set (NESS)*³ di Halpern⁴. Tutte queste strade sono state abbandonate dagli autori, probabilmente perché ritenute troppo forti. Soprattutto per quanto riguarda *Causality: Models, Reasoning, and Inference*⁵, credo che una svista riguardo alle potenzialità della nozione di *sustenance* ne abbia causato una svalutazione nella letteratura successiva, causandone la caduta nel dimenticatoio.

La definizione è la seguente:

DEFINIZIONE 5.1.1 (CAUSA EFFETTIVA)

$X = x$ è una causa effettiva di φ in (M, \vec{u}) sse:

$$CE_1 \quad (M, \vec{u}) \models X = x \wedge \varphi$$

$$CE_2 \quad \text{Sia } \vec{W} \subset \mathcal{V} \text{ e } (M, \vec{u}) \models \vec{W} = \vec{w}^*, \text{ allora se } X = x \in \vec{R}_{Y=y|\vec{w}^*}, \text{ e } \vec{Z} \subseteq \mathcal{V} \setminus \{\vec{W} \cup \vec{R}_{Y=y|\vec{w}^*}\} \text{ e } \vec{Z}' \in \mathcal{R}(\vec{Z}) \text{ tali che:}$$

$$(M, \vec{u}) \models [X \leftarrow x', \vec{W} \leftarrow \vec{w}^*, \vec{Z} \leftarrow \vec{Z}'] \neg \varphi.$$

¹ Pearl, *Causality: Models, Reasoning, and Inference* cit.:def. 10.2.1.

² Halpern e Pearl, «Causes and Explanations: A Structural-Model Approach. Part II: Explanations» cit.: 855.

³ Wright, «Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts» cit.; Wright, «The NESS Account of Natural Causation: A Response to Criticisms» cit.

⁴ Halpern, «Defaults and Normality in Causal Structures» cit.: 206.

⁵ Pearl, *Causality: Models, Reasoning, and Inference* cit.

Che, a parole, significa che $X = x$ è una causa effettiva di $Y = y$ se, oltre che accadere nel contesto effettivo, fa parte di un blocco causale condizionale su di un sottoinsieme (proprio) del contesto e, congelando almeno le variabili che abilitano il blocco causale, esistono delle imposizioni di valore ad alcune variabili tali per cui, se x' ottiene, allora l'effetto non si verifica. La controparte positiva è garantita dal fatto che, se $X = x$ fa parte di un blocco causale condizionale al contesto attuale, allora per forza $Y = y$ si verifica.

L'idea di fondo è estremamente semplice: se, nelle condizioni attuali, vi è un evento che garantisce che l'effetto si verifichi, ma una volta che questo evento non accade (con le giuste condizioni al contorno), anche l'effetto non accade, allora l'evento è una causa. Le giuste condizioni al contorno, rappresentate da \vec{z} , servono a fare in modo che l'effetto sia controfattualmente dipendente dalla causa in condizioni ideali. Questa modifica ai valori del modello serve, semplicemente, a permetterci di disconnettere alcune cause spurie nel momento in cui si controlla la parte negativa del test lewisiano, $\neg C \square \rightarrow \neg E$. Naturalmente, il fatto che $X = x$ faccia parte di un blocco causale, nuovamente, ci assicura che, se l'evento si verifica, allora anche l'effetto si verifica, indipendentemente dal contesto (escluse, ovviamente, le condizioni da cui il blocco dipende).

Questa definizione è piuttosto differente dalle definizioni **HP**, che sono, forse, più complesse a causa della loro elaborazione, fatta su di una base estensiva, ovvero testandole caso per caso e apportando le necessarie modifiche affinché la definizione catturi tutti gli esempi. Questa definizione, è, invece, nata attraverso l'applicazione di una semplice idea, ovvero la sufficienza della causa per l'effetto, date alcune condizioni al contorno, sommata all'effettivo verificarsi della causa. Essa è più simile, ad un test **NESS**⁶, anche se, in realtà, involontariamente.

Passiamo, ora alla discussione di alcuni esempi al fine di mostrare come la definizione sia, a livello estensivo, un'ottima definizione. Iniziamo con la discussione dell'**Esempio Incendio**^{mod}. L'applicazione della definizione mi ha permesso di realizzare che, in questo caso, abbiamo di fronte una situazione che ricorda estremamente da vicino il *trumping*⁷. Si consideri, infatti, il seguente esempio, proposto da **Schaffer**:

ESEMPIO 5.1.1 (Trumping)

Un Soldato Semplice siede in trincea, in attesa di ordini. Il Maggiore e il Sergente urlano, al tempo: «AVANZARE!». Il Soldato, nel determinare le sue azioni, segue la gerarchia degli ufficiali che impartiscono l'ordine. In questo caso avanza. ■

Immaginando che i superiori possano dare due ordini (avanzare e ritirarsi), l'equazione che descrive il comportamento del soldato, data la storia, è la seguente, con $M = 0$ se il Maggiore non impartisce ordini:

$$C = \begin{cases} M, & \text{se } M \neq 0 \\ S, & \text{se } M = 0 \end{cases}$$

In questo caso la nostra intuizione è che è siano gli ordini del maggiore a causare il comportamento del Soldato Semplice. In questo caso, sia $M = 1$ (avanzare), che $M = 2$ fanno parte di un blocco

⁶ Wright rende verbalmente il test **NESS** nella seguente maniera: una particolare condizione è stata una causa di (ha contribuito a) un risultato specifico se e solo se era un elemento necessario di un insieme di antecedenti condizioni effettive che era sufficiente per l'occorrenza del risultato. Wright, «Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts» cit.: 1019. Si noti la somiglianza con la condizione **INUS** di Mackie.

⁷ Schaffer, «Trumping Preemption» cit.

causale non condizionale, in quanto sono sufficienti a garantire l'evento. Non si può dire lo stesso per gli ordini impartiti dal Sergente: questi fanno parte di un blocco causale condizionale su $M = 0$. Di conseguenza, possono essere considerati una causa solo nei contesti $\vec{u}: M = 0 \in \vec{u}$.

Si ricordi ora la discussione dell'**Esempio Incendio^{mod}**. In questo caso **Halpern** sosteneva che $F = 1$ fosse da considerarsi una causa, in quanto ci troviamo di fronte ad un caso di sovradeterminazione causale, e che il fallimento delle definizioni **HP** era problematico. Anche per questa nuova definizione è problematico: F , infatti, non fa parte di alcun blocco causale, nemmeno condizionale, escluso quello in cui $\vec{w} = P_1 = P_2 = 0$. Di conseguenza, non vi è modo che esso sia una causa, in quanto non è mai sufficiente a produrre $FF = 1$, se non nel caso in cui i piromani non facciano cadere i fiammiferi. Infatti, si noti che non possiamo creare un blocco causale condizionale su $P_1 = P_2 = 1$, in quanto è esso stesso un blocco causale, inoltre, non possiamo creare altri blocchi causali condizionali in quanto, se $P_1 \neq P_2$, allora abbiamo $FF = 2$. Questo pattern di dipendenze è estremamente simile ai casi di trumping.

Il problema è, quindi, non che la nostra definizione di causa non funziona, bensì che il fulmine, nella situazione, non è sufficiente a far accadere nulla, proprio come il nostro sergente, e che, di conseguenza è l'intuizione ad essere errata. La definizione di causa effettiva proposta (e anche quelle della famiglia **HP**) performa benissimo, in quanto produce la corretta ascrizione di causalità in quanto il fulmine non partecipa alla produzione di nessun aspetto della situazione causale.

Per evidenziare che l'intuizione di **Halpern** è scorretta, sfrutterò nuovamente l'argomento di **Blanchard e Schaffer**: il modello è inadatto, ma questa volta non nel senso che non produce l'output corretto, bensì nel senso che non aiuta la nostra intuizione a comprendere correttamente la situazione. Questo avviene, ironicamente, perché non ci rendiamo che il modello è isomorfo ad un caso di trumping, quando, con la modellazione corretta, l'isomorfismo è chiarissimo. Ciò che è necessario fare in questo caso è interpolare una variabile tra i due piromani e l'incendio (IP) e modificare l'equazione di FF :

$$FP = \begin{cases} 1, & \text{se } P_1 = P_2 = 1 \\ 2, & \text{se } P_1 \neq P_2 \\ 0, & \text{se } P_1 = P_2 = 0 \end{cases}$$

$$FF = \begin{cases} FP, & \text{se } FP \neq 0 \\ F, & \text{se } FP = 0 \end{cases}$$

In **Figura 5.1**, i grafi delle due situazioni: si noti l'isomorfismo del sottografo $IP \rightarrow FF \leftarrow L$ rispetto a $M \rightarrow SS \leftarrow S$. Fondamentalmente, quello che facciamo per mettere in luce l'isomorfismo è trattare le variabili P_1 e P_2 come se fossero parte del contesto \vec{u} .

5.2 ALCUNI ESEMPI

In questa sezione presenterò alcuni esempi, alcuni dei quali già trattati, e li discuterò, al fine di mostrare le potenzialità della nuova definizione.

Iniziamo con un caso di *late preemption*, l'**Esempio Rock Throwing**. In questo caso è semplice vedere che il lancio di Suzy fa parte di un blocco causale non condizionale e, quindi, dato che si

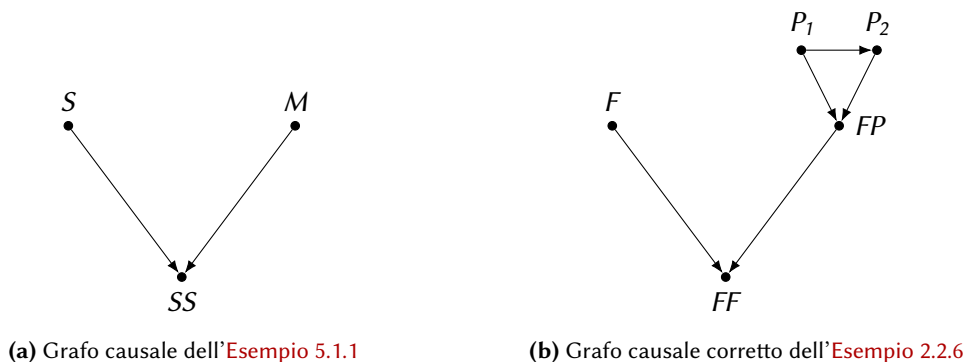


Figura 5.1: Grafi causali degli Esempi **Trumping** e **Incendio^{mod}**.

verifica, esso è una causa (è sufficiente considerare la contingenza in cui Billy non colpisce per vederlo, incorporata attraverso Z , che in questo caso è un singoletto). Non così, invece, per il lancio di Billy: esso fa parte di un blocco causale, ma condizionale. In particolare, condizionale rispetto al contesto $w = \{SC = 0\}$. Non verificandosi questo caso, non può essere considerato una causa in quanto non è condizionale su di un sottoinsieme del contesto effettivo.

Per quanto riguarda l'Esempio **Incendio Simultaneo**, chiaramente entrambi i casi fanno parte di un blocco causale. Infatti, sia il fulmine che il piromane sono in grado di offrire la garanzia dell'esito, ovvero sono, singolarmente, sufficienti al verificarsi dell'incendio. Vale, quindi, quanto detto per il tiro di Suzy: si imposti z in modo da fare in modo che l'evento avverso rispetto ai casi che stiamo valutando come causa non accada.

Per l'Esempio **Voting Machine**, ovvero il caso che ci ha spinti lontani dalla prima definizione di **Pearl**, la situazione è la medesima, in quanto si tratta semplicemente di un caso di sovradeterminazione (come Incendio Simultaneo), ma mediato da una variabile interpolata. Nuovamente, semplicemente selezioniamo z in modo che dia l'esito corretto nel momento in cui valutiamo V_1 e V_2 come cause. Si noti quanto la possibilità di modificare dei valori presenti nella situazione attraverso l'utilizzo dell'insieme \vec{Z} sia fondamentale per catturare correttamente tutti questi casi in cui siamo di fronte a cause per scambio.

Backup, pur essendo diverso (si tratta, infatti, di un caso di early preemption), è, per la **Definizione 5.1.1**, identico agli esempi trattati sin ora, con la sola differenza che non dobbiamo nemmeno settare z , ma ci è sufficiente mantenere il contesto $w^* = \{M = 0\}$, in quanto la dipendenza controfattuale, una volta congelata la variabile, è mantenuta.

Plotone di Esecuzione, invece, rappresenta un problema. Non nella situazione rappresentata, dove $\vec{u} = A = 1, B = 0, C = 1$, in cui, naturalmente, $C = 1$ è parte di un blocco causale effettivo e, quindi, è una causa semplicemente mantenendo fisso w^* ; bensì nel caso in cui $C = 0$ e $A = B = 1$. Infatti, né A né B fanno parte di un blocco causale e ci troviamo, quindi, privi di causa. La **Definizione 5.1.1** non prevede, infatti, che le cause possano essere congiunti.

La soluzione è piuttosto semplice: modificare la definizione di dipendenza diretta (forte e condizionale) in modo da accettare anche congiunzioni di eventi, purché minimali. La modifica è mera materia di sostituire X e Y con \vec{X} e \vec{Y} . Lo stesso vale per la definizione di blocco causale: semplicemente permettere agli elementi dell'insieme \mathcal{I} di essere congiunti⁸. Infine, si corregge la definizione

⁸ Questa modifica non è stata applicata nel capitolo precedente in quanto si voleva porre enfasi sull'utilizzo di interventi su singole variabili. Nonostante ciò, è possibile fare uso dei concetti presentati anche con questa variazione. Semplicemente, si considereranno interventi congiunti.

di causa affinché accetti cause congiunte, purché minimali. Ora, anche questo caso, è correttamente analizzato.

Un altro esempio, si presenta la trattazione che questa definizione offre di un caso considerato estremamente problematico per la definizione **HP** e, ad oggi, ancora irrisolto dalla teoria nella sua forma genuina⁹ è il seguente caso, presentato da Glymour *et al.*:

ESEMPIO 5.2.1

Un ranch ha 5 persone: Cowboy, Ranger, Wrangler e due Aiutanti. Tutti quanti votano se rimanere o meno al falò ($O = 0$), oppure perlustrare la zona ($O = 1$). Per decidere, si usa una regola complicata:

1. Se $R = C$, allora R .
2. Se $C = W = H_1 = H_2$, allora R .
3. Altrimenti maggioranza.

Si supponga che $\vec{u} = \{R = C = 1, W = H_1 = H_2 = 0\}$. $W = 0$ è una causa effettiva di $O = 1$?¹⁰ ■

La nostra definizione considera, a differenza delle varie definizioni della famiglia **HP**¹¹, che considerano tutti i voti come causa, riconoscono solo $R = 1$ e $C = 1$ come cause. Si noti, infatti, che, innanzitutto $R = 1$ è un blocco causale in questo contesto, in quanto, dato $C = 1$, abbiamo la garanzia che $O = 1$. Inoltre, se $R = 0$, allora $O = 0$. Per quanto riguarda $C = 1$, essa è una causa perché, anche se imponendo esclusivamente $C = 0$, avremmo comunque $O = 1$ (per la regola 2), possiamo selezionare, per un sottoinsieme delle variabili W, H_1, H_2 , uno z . In questo modo, ovvero impostando $C = 0$ e, ad esempio, $H_2 = 0$ otterremo $O = 0$ per la regola 3, soddisfacendo la definizione. Si noti, invece, che non possiamo fare lo stesso ragionamento per congiunzioni di variabili (ovvero, il problema che incontra la **Definizione 2.2.4**), in quanto tutte dovrebbero contenere R o C , contro la condizione di minimalità aggiunta per le congiunzioni.

Con questo ultimo esempio abbiamo esplorato un discreto campo di esempi. Molto altro resta, naturalmente, da testare, ma per ora la definizione proposta sembra essere in grado di catturare un gran numero di casi. Questo suggerisce che il fatto di *andare a caccia* dei vari concetti causali che impieghiamo in diversi campi e situazioni, possa aiutarci nel nostro progetto di esplorazione della causalità e della formalizzazioni di criteri estremamente utili per orientarci nelle nostre decisioni e nell'ottenimento dei nostri obiettivi.

⁹ Si veda Halpern, *Actual Causality* cit.: 112-113 per una trattazione. La soluzione proposta da Halpern richiede l'introduzione di variabili extra.

¹⁰ Glymour *et al.*, «Actual causation: a stone soup essay» cit.: 185.

¹¹ Si vedano: *ivi* per una critica della **Definizione 2.2.1**, Halpern, *Actual Causality* cit. analizza le Definizioni **Causa effettiva**^a e **Causa effettiva**^m, riconoscendone la non sufficienza.

CONCLUSIONE

Nel corso di questo scritto è stata esplorata la cornice metodologica dei modelli causali. C'è stato modo di mostrare le grandi potenzialità di questo approccio, che coglie gli spunti della teoria controfattuale, reinterpretandoli in una nuova maniera, al fine di rendere la trattazione della causalità più precisa ed efficace. La forte matematizzazione, inoltre, permette ai modelli causali un'applicazione quasi trasversale nelle diverse discipline che con la causalità hanno a che fare. **Blanchard e Schaffer**, ci ricordano, infatti:

Forse, il principale punto forte di questa cornice è lo sviluppo di 'algoritmi di scoperta' che permettono di inferire la struttura causale da dati correlazionali (qualcosa che gli statistici un tempo screditavano come un'impresa impossibile).

Il potere e la precisione di questa cornice metodologica è, a nostro parere, senza rivali. Non per nulla tutto il lavoro recente sulla causa attuale è formulato nei suoi termini. Un account formulato in altri termini—senza uno sviluppo della cornice metodologica rivale a livelli di sofisticazione comparabili—diventa difficile da prendere seriamente.¹²

I modelli causali sono, probabilmente, lo strumento più potente che possediamo per analizzare, inferire e valutare le relazioni causali. Questo scritto ha voluto estendere alcune delle linee di ricerca che sono recentemente emerse all'interno della discussione sui modelli causali.

È stata analizzata la proposta di una semantica dei controfattuali da parte di **Hiddleston**. Sono stati mostrati i punti interessanti di questa proposta, che fanno luce anche sulle grandi potenzialità dei modelli causali in campo filosofico. Dopo aver discusso, invece, alcune delle mancanze della semantica, ho discusso e proposto degli emendamenti che renderebbero più precisa. Nonostante ciò, abbiamo notato che la semantica non può valutare controfattuali che derivano situazioni che non seguono le equazioni del modello: una limitazione, questa, che denota una certa crudezza della proposta, in quanto questo tipo di controfattuali sono utilizzati su base praticamente quotidiana. Vi è, quindi, necessità di ulteriore lavoro su questa problematica al fine di produrre una semantica veramente completa. Infine, il problema principale della semantica proposta è che più che collimare con quelle che sono le nostre intuizioni riguardo ai controfattuali, descrive la modalità con cui ragioniamo quando si tratta di ricostruire il passato.

Successivamente, mi sono concentrato sul **PI**, una problematica estremamente pericolosa per i modelli causali. La discussione del problema da parte di **Blanchard e Schaffer** ha offerto l'occasione di presentare quella che è una certa *malafede* riposta nei modelli causali da parte di alcuni filosofi che si occupano di causalità, quasi fossero troppo efficaci per poter essere corretti.

Questo ci ha dato l'occasione di notare che, anche dopo vent'anni dal monumentale lavoro di **Pearl**, ci troviamo di fronte a grossolani errori di modellazione, che si mascherano da controesempi quando, in realtà, sono modelli errati che non rispecchiano la struttura intrinseca della situazione che vogliono rappresentare. Ciò ci ha permesso, attraverso una ricognizione della letteratura, di tentare di fornire alcune linee guida per una modellazione più corretta e consapevole.

¹² **Blanchard e Schaffer**, «Cause without Default» cit.: 178.

Quella di fare modelli causali resta, al momento, un'arte; ma procedendo a piccoli passi dei progressi si stanno facendo. Probabilmente, in un futuro non troppo remoto, riusciremo a formulare dei criteri esatti attraverso i quali produrre ottimi modelli. Nel frattempo, possiamo provare ad appoggiarci a procedure per raggiungere l'accordo su quale modello utilizzare, come quella proposta nel **Capitolo 3**, sperando che possa aiutarci a chiarire in che modo il nostro modello rispecchia le nostre assunzioni e in che modo, invece, non rispecchia quelle proposte dal modello rivale.

Rifacendosi alla posizione funzionalista propugnata da **Woodward** e **Danks** (ma non solo), abbiamo esplorato il pluralismo causale e abbiamo proposto dei concetti causali volti a catturare il modo in cui ci confrontiamo con il mondo e il modo in cui ragioniamo riguardo a situazioni causali. Tralasciando il fatto che tutto l'apparato di supporto, come, ad esempio, la nozione di salvaguardia, meriterebbe una trattazione a sé dedicata, ci siamo concentrati su due concetti ritenuti di grande interesse, oltre che descrittivo, anche normativo: i due concetti fondamentali sono quello di blocco causale e di cause per scambio. Questi due concetti catturano, a mio parere, in maniera molto efficace il modo in cui ragioniamo sulle cause.

Per quanto riguarda il concetto di blocco causale: oltre alle considerazioni fatte in chiusura della **Sezione 4.2.1.3**, si pensi a quanto spesso, avendo la possibilità di bloccare un esito, lo facciamo anche nei casi in cui non vi sia alcun pericolo che esso non si verifichi. Abbiamo un'innata preferenza verso i blocchi causali in quanto ci permettono di agire con la (relativa) sicurezza che non può non avvenire l'evento per cui stiamo agendo. Ho il sospetto che questo sia il primo concetto di causalità che ci formiamo, nella sua versione ad una singola variabile, in quanto il più semplice ed immediato. Il blocco causale è, nella sua assenza: azione, reazione; senza dubbi di sorta. Probabilmente, è anche il concetto che informa la risposta che si ottiene se si chiede a qualcuno che cosa sia una causa.

Per quanto riguarda, invece, il concetto di cause per scambio, credo che esso funga da guida in situazioni in cui le acque sono più torbide, ed è necessaria una maggiore cautela nell'agire: è, infatti, lo strumento attraverso il quale pensiamo una vasta gamma di situazioni causali in cui diversi eventi vanno a contribuire alla produzione dell'effetto. Questa contribuzione va intesa in senso deterministico¹³. Il concetto di cause per scambio ci fornisce gli strumenti necessari per effettuare manipolazioni consapevoli in contesti in cui abbiamo diversi processi in atto, e ci aiuta a determinare quali vadano selezionati come buoni obiettivi di intervento.

Infine, quest'analisi funzionale di alcuni dei concetti afferenti alla nozione di causalità ha permesso lo sviluppo di concettualità ulteriore: nella fattispecie, una nuova definizione di causa effettiva. Una rapida esplorazione degli esempi presenti in letteratura ha rivelato che, effettivamente, vi è del potenziale. Questa nuova definizione, infatti, riesce a rendere conto anche di almeno un caso con il quale le definizioni **HP** si trovano in difficoltà, riuscendo a restituire il verdetto corretto oltre che in questo caso, anche negli altri analizzati. Oltre ad essere estensivamente potente, la nuova definizione ha anche, dal suo lato, la semplicità: essa afferma, in poche parole, che una causa deve essere sufficiente per il proprio effetto dato un determinato contesto. Questa è una caratteristica promettente: essa deriva, infatti, dal concetto di blocco causale che, come affermato precedentemente, è estremamente primitivo ed intuitivo, perlomeno nella sua versione minimale (ovvero $|\vec{R}| = 2$). Appoggiandosi, a questo concetto quasi naturale, la definizione è estremamente comprensibile rispetto alle artificiali

¹³ Il concetto in versione probabilistica è già stato sviluppato da **Hitchcock** (C. Hitchcock, «A Tale of Two Effects», *The Philosophical Review*, vol. 110, n. 3 (2001), p. 361, DOI: [10.2307/2693649](https://doi.org/10.2307/2693649)) e **Pearl** (J. Pearl, «Direct and Indirect Effects», in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, Morgan Kaufmann Publishers Inc., Seattle, Washington 2001, pp. 411-420; **Pearl**, *Causality: Models, Reasoning, and Inference* cit.).

ed oscure definizioni della famiglia **HP**. Lo sviluppo di questa definizione, che naturalmente è ancora incompleta e necessita di *labor limae* al fine di avere garanzie di correttezza e applicabilità, offre un'ulteriore prospettiva di ricerca che spero, in futuro, di poter perseguire.

BIBLIOGRAFIA

- ARISTOTELE, *Metafisica*, trad. da G. REALE, Bompiani, 2000.
- BALKE, A. e J. PEARL, «Counterfactuals and Policy Analysis in Structural Models», in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, Morgan Kaufmann Publishers Inc., Montréal, Qué, Canada 1995, pp. 11-18.
- BARKER, S., «Counterfactuals, probabilistic counterfactuals and causation», *Mind*, vol. 108, n. 431 (1999), pp. 427-469, DOI: [10.1093/mind/108.431.427](https://doi.org/10.1093/mind/108.431.427).
- BEEBEE, H., «Causing and Nothingness», in J. COLLINS, N. HALL e L. A. PAUL (a cura di), *Causation and Counterfactuals*, MIT Press, 2004, pp. 291-308.
- BEEBEE, H., C. HITCHCOCK e P. MENZIES (a cura di), *The Oxford Handbook of Causation*, Oxford University Press, 2010, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001).
- BEEBEE, H., C. HITCHCOCK e H. PRICE (a cura di), *Making a Difference*, Oxford University Press, 2017, DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001).
- BENNET, J., «Event Causation: The Counterfactual Analysis», in *Causation*, a cura di E. SOSA e M. TOOLEY, Oxford Readings in Philosophy, Oxford University Press, 1993.
- BJÖRNSSON, G., «How Effects Depend on their Causes, Why Causal Transitivity Fails, and Why we Care About Causation», *Philosophical Studies*, vol. 133, n. 3 (2007), pp. 349-390, DOI: [10.1007/s11098-005-4539-8](https://doi.org/10.1007/s11098-005-4539-8).
- BLANCHARD, T. e J. SCHAFER, «Cause without Default», in H. BEEBEE, C. HITCHCOCK e H. PRICE (a cura di), *Making a Difference*, Oxford University Press, 2017, DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001), pp. 175-214, DOI: [10.1093/oso/9780198746911.001.0001](https://doi.org/10.1093/oso/9780198746911.001.0001).
- BRIGGS, R., «Interventionist counterfactuals», *Philosophical Studies*, vol. 160, n. 1 (2012), pp. 139-166, DOI: [10.1007/s11098-012-9908-5](https://doi.org/10.1007/s11098-012-9908-5).
- BROADBENT, A., *Philosophy of Epidemiology*, Palgrave Macmillan UK, London 2013, DOI: [10.1057/9781137315601](https://doi.org/10.1057/9781137315601).
- CAMPBELL, J. K., M. O'ROURKE e H. S. SILVERSTEIN (a cura di), *Causation and Explanation*, Topics in Contemporary Philosophy, Bradford Books, Cambridge, MA 2007.
- CARRARA, M., C. DE FLORIO, G. LANDO e V. MORATO, *Introduzione alla metafisica contemporanea*, il Mulino, 2021.
- CARTWRIGHT, N., *Nature's Capacities and Their Measurement*, Oxford University Press, 1994, DOI: [10.1093/0198235070.001.0001](https://doi.org/10.1093/0198235070.001.0001).
- «Against Modularity, the Causal Markov Condition, and Any Link between the Two: Comments on Hausman and Woodward», *The British Journal for the Philosophy of Science*, vol. 53, n. 3 (2002), pp. 411-453, <http://www.jstor.org/stable/3541889>.
 - «Two Theorems on Invariance and Causality», *Philosophy of Science*, vol. 70, n. 1 (2003), pp. 203-224, DOI: [10.1086/367876](https://doi.org/10.1086/367876).
 - *Hunting Causes and Using Them. Approaches in Philosophy and Economics*, Cambridge University Press, 2007.

- CHENG, P. W., «From covariation to causation: A causal power theory.» *Psychological Review*, vol. 104, n. 2 (1997), pp. 367-405, DOI: [10.1037/0033-295X.104.2.367](https://doi.org/10.1037/0033-295X.104.2.367).
- COLLINS, J., N. HALL e L. A. PAUL (a cura di), *Causation and Counterfactuals*, MIT Press, 2004.
- COLLINS, J., N. HALL e L. A. PAUL, «Counterfactuals and Causation: History, Problems and Prospects», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit., pp. 1-58.
- CUSHMAN, F., J. KNOBE e W. SINNOTT-ARMSTRONG, «Moral appraisals affect doing/allowing judgments», *Cognition*, vol. 108, n. 1 (2008), pp. 281-289, DOI: [10.1016/j.cognition.2008.02.005](https://doi.org/10.1016/j.cognition.2008.02.005).
- DANKS, D., «The psychology of causal perception and reasoning», in H. BEEBEE, C. HITCHCOCK e P. MENZIES (a cura di), *The Oxford Handbook of Causation*, Oxford University Press, 2010, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001), pp. 447-470, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001).
- «Functions and Cognitive Bases for the Concept of Actual Causation», *Erkenntnis*, vol. 78, n. S1 (2013), pp. 111-128, DOI: [10.1007/s10670-013-9439-2](https://doi.org/10.1007/s10670-013-9439-2).
- DANKS, D., D. ROSE e E. MACHERY, «Demoralizing causation», *Philos Stud*, vol. 171, n. 2 (2013), pp. 251-277, DOI: [10.1007/s11098-013-0266-8](https://doi.org/10.1007/s11098-013-0266-8).
- DENG, D. e K. Y. LEE, «Indicative and counterfactual conditionals: a causal-modeling semantics», *Synthese*, vol. 199, n. 1-2 (2021), pp. 3993-4014, DOI: [10.1007/s11229-020-02966-9](https://doi.org/10.1007/s11229-020-02966-9).
- DOWE, P., *Physical Causation*, Cambridge University Press, 2000.
- EHRING, D., «Causal Relata», in [BEEBEE et al.](#) (a cura di), *The Oxford Handbook of Causation* cit., pp. 387-413, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001).
- EITER, T. e T. LUKASIEWICZ, «Complexity results for structure-based causality», *Artificial Intelligence*, vol. 142, n. 1 (2002), pp. 53-89, DOI: [10.1016/S0004-3702\(02\)00271-0](https://doi.org/10.1016/S0004-3702(02)00271-0).
- FENTON-GLYNN, L., «Of Miracles and Interventions», *Erkenntnis*, vol. 78, n. S1 (2013), pp. 43-64, DOI: [10.1007/s10670-013-9436-5](https://doi.org/10.1007/s10670-013-9436-5).
- «A Proposed Probabilistic Extension of the Halpern and Pearl Definition of ‘Actual Cause’», *The British Journal for the Philosophy of Science*, vol. 68, n. 4 (2017), pp. 1061-1124, DOI: [10.1093/bjps/axv056](https://doi.org/10.1093/bjps/axv056).
- FISCHER, E., «Causation and the Problem of Disagreement», *Philos. of Sci.*, vol. 88, n. 5 (2021), pp. 773-783, DOI: [10.1086/714852](https://doi.org/10.1086/714852).
- FRIEDMAN, N. e J. Y. HALPERN, «Plausibility Measures: A User’s Guide», in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI’95, Morgan Kaufmann Publishers Inc., Montréal, Qué, Canada 1995, pp. 175-184.
- GERSTENBERG, T., C. BECHLIVANIDIS e D. LAGNADO, «Back on track: Backtracking in counterfactual reasoning», in *Proceedings of the Annual Meeting of the Cognitive Science Society*, n. 35, 2013, vol. 35, DOI: [10.13140/2.1.1473.1209](https://doi.org/10.13140/2.1.1473.1209).
- GLYMOUR, C., D. DANKS, B. GLYMOUR *et al.*, «Actual causation: a stone soup essay», *Synthese*, vol. 175, n. 2 (2009), pp. 169-192, DOI: [10.1007/s11229-009-9497-9](https://doi.org/10.1007/s11229-009-9497-9).
- GLYMOUR, C. e F. WIMBERLY, «Actual Causes and Thought Experiments», in J. K. CAMPBELL, M. O’ROURKE e H. S. SILVERSTEIN (a cura di), *Causation and Explanation*, Topics in Contemporary Philosophy, Bradford Books, Cambridge, MA 2007, pp. 43-68.
- GOLDBERGER, A. S., «Structural Equation Methods in the Social Sciences», *Econometrica*, vol. 40, n. 6 (1972), p. 979, DOI: [10.2307/1913851](https://doi.org/10.2307/1913851).

- GOPNIK, A., C. GLYMOUR, D. M. SOBEL *et al.*, «A Theory of Causal Learning in Children: Causal Maps and Bayes Nets», *Psychological Review*, vol. 111, n. 1 (2004), pp. 3-32, DOI: [10.1037/0033-295X.111.1.3](https://doi.org/10.1037/0033-295X.111.1.3).
- HALL, N., «Causation and the Price of Transitivity», in *COLLINS et al.* (a cura di), *Causation and Counterfactuals* cit., pp. 181-203.
- «Two Concepts of Causation», in *COLLINS et al.* (a cura di), *Causation and Counterfactuals* cit., pp. 225-276.
 - «Structural Equation and Causation», *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 132, n. 1 (2007), pp. 109-136.
- HALL, N. e L. A. PAUL, *Causation: A User's Guide*, Oxford University Press, 2013.
- HALPERN, J. Y., «Axiomatizing Causal Reasoning», *Journal of Artificial Intelligence Research*, vol. 12 (2000), pp. 317-337, DOI: [10.1613/jair.648](https://doi.org/10.1613/jair.648).
- «Defaults and Normality in Causal Structures», in *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, KR'08, AAAI Press, Sydney, Australia 2008, pp. 198-208.
 - «A Modification of the Halpern-Pearl Definition of Causality», in *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, AAAI Press, Buenos Aires, Argentina 2015, pp. 3022-3033.
 - *Actual Causality*, MIT Press, 2016.
 - «Appropriate Causal Models and the Stability of Causation», *The Review of Symbolic Logic*, vol. 9, n. 1 (2016), pp. 76-102, DOI: [10.1017/s1755020315000246](https://doi.org/10.1017/s1755020315000246).
- HALPERN, J. Y. e C. HITCHCOCK, «Actual Causation and the Art of Modeling», in *Heuristics, Probability and Causality. A Tribute to Judea Pearl*, a cura di R. DECHTER, H. GEFFNER e J. Y. HALPERN, College Publications, 2010.
- «Compact Representations of Extended Causal Models», *Cognitive Science*, vol. 37, n. 6 (2013), pp. 986-1010, DOI: [10.1111/cogs.12059](https://doi.org/10.1111/cogs.12059).
 - «Graded causation and defaults», *The British Journal for the Philosophy of Science*, vol. 66, n. 2 (2015), pp. 413-457.
- HALPERN, J. Y. e J. PEARL, «Causes and Explanations: A structural-model approach. Part I: Causes», in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001)*, 2001, pp. 194-201.
- «Causes and Explanations: A structural-model approach. Part I: Causes», *British Journal for the Philosophy of Science*, vol. 56, n. 4 (2005), pp. 843-887, DOI: [10.1093/bjps/axi147](https://doi.org/10.1093/bjps/axi147).
 - «Causes and Explanations: A Structural-Model Approach. Part II: Explanations», *The British Journal for the Philosophy of Science*, vol. 56, n. 4 (2005), pp. 889-911.
- HART, H. L. A. e A. M. HONORÉ, *Causation in the law*, Oxford University Press, 1959.
- HAUSMAN, D. M., *Causal Asymmetries*, Cambridge University Press, 1998, DOI: [10.1017/cb09780511663710](https://doi.org/10.1017/cb09780511663710).
- HAUSMAN, D. M. e J. WOODWARD, «Independence, Invariance and the Causal Markov Condition», *The British Journal for the Philosophy of Science*, vol. 50, n. 4 (1999), pp. 521-583, DOI: [10.1093/bjps/50.4.521](https://doi.org/10.1093/bjps/50.4.521).

- HERNÁN, M. A. e S. L. TAUBMAN, «Does obesity shorten life? The importance of well-defined interventions to answer causal questions», *Int J Obes*, vol. 32, n. S3 (2008), S8-S14, DOI: [10.1038/ijo.2008.82](https://doi.org/10.1038/ijo.2008.82).
- HERNÁN, M. A., D. CLAYTON e N. KEIDING, «The Simpson's paradox unraveled», *International Journal of Epidemiology*, vol. 40, n. 3 (2011), pp. 780-785, DOI: [10.1093/ije/dyr041](https://doi.org/10.1093/ije/dyr041).
- HESSLOW, G., «Two Notes on the Probabilistic Approach to Causality», *Philos. of Sci.*, vol. 43, n. 2 (1976), pp. 290-292, DOI: [10.1086/288684](https://doi.org/10.1086/288684).
- HIDDLESTON, E., «A Causal Theory of Counterfactuals», *Noûs*, vol. 39, n. 4 (2005), pp. 632-657, DOI: [10.1111/j.0029-4624.2005.00542.x](https://doi.org/10.1111/j.0029-4624.2005.00542.x).
- «Causal Powers», *The British Journal for the Philosophy of Science*, vol. 56, n. 1 (2005), pp. 27-59, DOI: [10.1093/phisci/axi102](https://doi.org/10.1093/phisci/axi102).
- HITCHCOCK, C., «A Tale of Two Effects», *The Philosophical Review*, vol. 110, n. 3 (2001), p. 361, DOI: [10.2307/2693649](https://doi.org/10.2307/2693649).
- «The Intransitivity of Causation Revealed in Equations and Graphs», *The Journal of Philosophy*, vol. 98, n. 6 (2001), pp. 273-299, DOI: [10.2307/2678432](https://doi.org/10.2307/2678432).
- «Of Humean Bondage», *The British Journal for the Philosophy of Science*, vol. 54, n. 1 (2003), pp. 1-25, DOI: [10.1093/bjps/54.1.1](https://doi.org/10.1093/bjps/54.1.1).
- «How to be a Causal Pluralist», in *Thinking About Causes. From Greek Philosophy to Modern Physics*, a cura di P. MACHAMER e G. WALTERS, University of Pittsburgh Press, Pittsburgh 2007.
- «On the Importance of Causal Taxonomy», in *Causal Learning. Psychology, Philosophy and Computation*, a cura di A. GOPNIK e L. SCHULZ, Oxford University Press, 2007.
- «Prevention, Preemption, and the Principle of Sufficient Reason», *The Philosophical Review*, vol. 116, n. 4 (2007), pp. 495-532.
- «What's Wrong with Neuron Diagrams?», in [CAMPBELL et al.](#) (a cura di), *Causation and Explanation* cit., pp. 69-92.
- «Causal Models», in *The Stanford Encyclopedia of Philosophy*, a cura di E. N. ZALTA, Summer 2020, Metaphysics Research Lab, Stanford University, 2020, <https://plato.stanford.edu/archives/sum2020/entries/causal-models/>.
- HITCHCOCK, C. e J. KNOBE, «Cause and Norm», *The Journal of Philosophy*, vol. 106, n. 11 (2009), pp. 587-612.
- HOPKINS, M. e J. PEARL, «Clarifying the usage of structural models for commonsense causal reasoning», in *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 2003, pp. 83-89.
- HUBER, F., «Structural Equations and Beyond», *The Review of Symbolic Logic*, vol. 6, n. 4 (2013), pp. 709-732, DOI: [10.1017/s175502031300018x](https://doi.org/10.1017/s175502031300018x).
- ICARD, T. F., J. F. KOMINSKY e J. KNOBE, «Normality and actual causal strength», *Cognition*, vol. 161 (2017), pp. 80-93, DOI: [10.1016/j.cognition.2017.01.010](https://doi.org/10.1016/j.cognition.2017.01.010).
- KHOO, J., «Backtracking Counterfactuals Revisited», *Mind*, vol. 126 (503 2016), pp. 841-910, DOI: [10.1093/mind/fzwo05](https://doi.org/10.1093/mind/fzwo05).
- KIM, J., «Causation, Nomic Subsumption, and the Concept of Event», *The Journal of Philosophy*, vol. 70, n. 8 (1973), p. 217, DOI: [10.2307/2025096](https://doi.org/10.2307/2025096).
- «Events as Property Exemplifications», in *Action Theory*, a cura di M. BRAND e D. WALTON, Springer Netherlands, 1976, pp. 159-177, DOI: [10.1007/978-94-010-9074-2_9](https://doi.org/10.1007/978-94-010-9074-2_9).

- «Events: Their Metaphysics and Semantics», *Philosophy and Phenomenological Research*, vol. 51, n. 3 (1991), p. 641, DOI: [10.2307/2107887](https://doi.org/10.2307/2107887).
- KNOBE, J. e B. FRASER, «Causal Judgement and Moral Judgement: Two Experiments», in *Moral Psychology, vol.2: The Cognitive Science of Morality: Intuition and Diversity*, a cura di W. SINNOT-ARMSTRONG, MIT Press, 2008, pp. 441-448.
- LEWIS, D., «Causation», *Journal of Philosophy*, n. 70 (1973), pp. 556-67.
- *Counterfactuals*, Harvard University Press, 1973.
- «Postscripts to 'Causation'», in *Philosophical Papers*, Oxford University Press, 1986, vol. II.
- «Causation as Influence», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit., pp. 75-106.
- «Void and Object», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit., pp. 277-290.
- LIVENGOD, J., «Actual Causation and Simple Voting Scenarios», *Noûs*, vol. 47, n. 2 (2013), pp. 316-345, DOI: [10.1111/j.1468-0068.2011.00834.x](https://doi.org/10.1111/j.1468-0068.2011.00834.x).
- LOMBROZO, T., «Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions», *Cognitive Psychology*, vol. 61, n. 4 (2010), pp. 303-332, DOI: [10.1016/j.cogpsych.2010.05.002](https://doi.org/10.1016/j.cogpsych.2010.05.002).
- MACKIE, J., «Causes and Conditions», *American Philosophical Quarterly*, n. 2 (1965), pp. 245-264.
- MCGRATH, S., «Causation By Omission: A Dilemma», *Philosophical Studies*, vol. 123, n. 1-2 (2005), pp. 125-148, DOI: [10.1007/s11098-004-5216-z](https://doi.org/10.1007/s11098-004-5216-z).
- MENZIES, P., «Difference-making in Context», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit.
- «Causation in Context», in *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, a cura di H. PRICE e R. CORRY, Oxford University Press, 2007.
- «Platitudes and counterexamples», in [BEEBEE et al.](#) (a cura di), *The Oxford Handbook of Causation* cit., pp. 341-367, DOI: [10.1093/oxfordhb/9780199279739.001.0001](https://doi.org/10.1093/oxfordhb/9780199279739.001.0001).
- «The causal structure of mechanisms», *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n. 4 (2012), pp. 796-805, DOI: [10.1016/j.shpsc.2012.05.008](https://doi.org/10.1016/j.shpsc.2012.05.008).
- «The Problem of Counterfactual Isomorphs», in [BEEBEE et al.](#), *Making a Difference* cit., DOI: [10.1093/os0/9780198746911.001.0001](https://doi.org/10.1093/os0/9780198746911.001.0001).
- MILL, J. S., *Sistema di logica deduttiva e induttiva*, a cura di M. TRINCHERO, Classici della Filosofia, UTET, 1988.
- MOORE, M. S., *Causation and Responsibility. An Essay in Law, Morals, and Metaphysics*, Oxford University Press, Oxford 2009.
- PAUL, L. A., «Aspect Causation», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit., pp. 205-224.
- PEARL, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., 1988.
- «Direct and Indirect Effects», in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, Morgan Kaufmann Publishers Inc., Seattle, Washington 2001, pp. 411-420.
- «Reply to Woodward», *Economics and Philosophy*, vol. 19, n. 2 (2003), pp. 341-344, DOI: [10.1017/S0266267103001196](https://doi.org/10.1017/S0266267103001196).

- PEARL, J., *Causality: Models, Reasoning, and Inference*, 2. Cambridge University Press, 2009.
- «Comment: Understanding Simpson's Paradox», *The American Statistician*, vol. 68, n. 1 (2014), pp. 8-13, DOI: [10.1080/00031305.2014.876829](https://doi.org/10.1080/00031305.2014.876829).
 - «Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes», *Journal of Causal Inference*, vol. 6, n. 2 (2018), DOI: [10.1515/jci-2018-2001](https://doi.org/10.1515/jci-2018-2001).
- PETTIT, D. e J. KNOBE, «The Pervasive Impact of Moral Judgment», *Mind & Language*, vol. 24, n. 5 (2009), pp. 586-604, DOI: [10.1111/j.1468-0017.2009.01375.x](https://doi.org/10.1111/j.1468-0017.2009.01375.x).
- POPA, E., «Getting counterfactuals right: the perspective of the causal reasoner», *Synthese*, vol. 200, n. 1 (2022), DOI: [10.1007/s11229-022-03492-6](https://doi.org/10.1007/s11229-022-03492-6).
- RUSSO, F., *Causality and Causal Modelling in the Social Sciences*, Springer Netherlands, 2009, DOI: [10.1007/978-1-4020-8817-9](https://doi.org/10.1007/978-1-4020-8817-9).
- SARTORIO, C., «A New Asymmetry Between Actions And Omissions», *Noûs*, vol. 39, n. 3 (2005), pp. 460-482, DOI: <https://doi.org/10.1111/j.0029-4624.2005.00509.x>.
- «The Prince of Wales Problem for Counterfactual Theories of Causation», in *New Waves in Metaphysics*, Palgrave Macmillan UK, 2010, pp. 259-276, DOI: [10.1057/9780230297425_14](https://doi.org/10.1057/9780230297425_14).
- SCHAFER, J., «Causation by Disconnection», *Philosophy of Science*, vol. 67, n. 2 (2000), pp. 285-300, DOI: [10.1086/392776](https://doi.org/10.1086/392776).
- «Trumping Preemption», in [COLLINS et al.](#) (a cura di), *Causation and Counterfactuals* cit., pp. 59-74.
 - «Contrastive Causation», *The Philosophical Review*, vol. 114, n. 3 (2005), pp. 327-358, DOI: [10.1215/00318108-114-3-327](https://doi.org/10.1215/00318108-114-3-327).
 - «Grounding in the image of causation», *Philosophical Studies*, vol. 173, n. 1 (2016), pp. 49-100, DOI: [10.1007/s11098-014-0438-1](https://doi.org/10.1007/s11098-014-0438-1).
- SLOMAN, S. A. e D. A. LAGNADO, «Do We "do"?», *Cognitive Science*, vol. 29, n. 1 (2005), pp. 5-39, DOI: [10.1207/s15516709cog2901_2](https://doi.org/10.1207/s15516709cog2901_2).
- SPIRITES, P., C. GLYMOUR e R. SCHEINES, *Causation, Prediction, and Search*, MIT Press, 2000.
- STATHAM, G., «Causes as Deviations from the Normal: Recent Advances in the Philosophy of Causation», in *Perspective on Causation. Selected Papers from the Jerusalem 2017 Workshop*, a cura di E. A. B.-A. SIEGAL e N. BONEH, Jerusalem Studies in Philosophy and History of Science, Springer International Publishing, 2020, pp. 445-462, DOI: [10.1007/978-3-030-34308-8_14](https://doi.org/10.1007/978-3-030-34308-8_14).
- SWANSON, E., «Lessons from the Context Sensitivity of Causal Talk», *The Journal of Philosophy*, vol. 107, n. 5 (2010), pp. 221-242, <http://www.jstor.org/stable/25764443>.
- SYTSMA, J., J. LIVENGOD e D. ROSE, «Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions», *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n. 4 (2012), pp. 814-820, DOI: [10.1016/j.shpsc.2012.05.009](https://doi.org/10.1016/j.shpsc.2012.05.009).
- TVERSKY, A. e D. KAHNEMAN, «Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment», in *Reasoning*, Cambridge University Press, 2008, pp. 114-135, DOI: [10.1017/cb09780511814273.008](https://doi.org/10.1017/cb09780511814273.008).
- VAN FRAASSEN, B. C., *The Scientific Image*, Oxford, England: Oxford University Press, 1980.
- *Laws and Symmetry*, Oxford University Press, 1989.
- WESLAKE, B., «A Partial Theory of Actual Causation», *British Journal for the Philosophy of Science* (2015), in uscita.

- WOODWARD, J., «What Is a Mechanism? A Counterfactual Account», *Philosophy of Science*, vol. 69, n. S3 (2002), S366-S377, DOI: [10.1086/341859](https://doi.org/10.1086/341859).
- «Critical Notice: Causality by Judea Pearl», *Economics and Philosophy*, vol. 19, n. 2 (2003), pp. 321-340, DOI: [10.1017/S0266267103001184](https://doi.org/10.1017/S0266267103001184).
 - *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, 2004, DOI: [10.1093/0195155270.001.0001](https://doi.org/10.1093/0195155270.001.0001).
 - «A Functional Account of Causation: or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)», *Philos. of Sci.*, vol. 81, n. 5 (2014), pp. 691-713, DOI: [10.1086/678313](https://doi.org/10.1086/678313).
- WOODWARD, J. e C. HITCHCOCK, «Explanatory Generalizations, Part I: A Counterfactual Account», *Noûs*, vol. 37, n. 1 (2003), pp. 1-24, DOI: [10.1111/1468-0068.00426](https://doi.org/10.1111/1468-0068.00426).
- WRIGHT, R. W., «Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts», *Iowa L. Rev.*, vol. 1001, n. 73 (1988), https://scholarship.kentlaw.iit.edu/fac_schol/698.
- «The NESS Account of Natural Causation: A Response to Criticisms», in *Critical Essays on "Causation and Responsibility"*, DE GRUYTER, 2013, pp. 13-66, DOI: [10.1515/9783110302295.13](https://doi.org/10.1515/9783110302295.13).
- WRIGHT, S., «Correlation and causation», *Journal of agricultural research.*, vol. 20, n. 3 (1921).
- «The Method of Path Coefficients», *The Annals of Mathematical Statistics*, vol. 5, n. 3 (1934), pp. 161-215, DOI: [10.1214/aoms/1177732676](https://doi.org/10.1214/aoms/1177732676).
- ZHANG, J., «A Lewisian Logic of Causal Counterfactuals», *Minds and Machines*, vol. 23, n. 1 (2013), pp. 77-93, DOI: [10.1007/s11023-011-9261-z](https://doi.org/10.1007/s11023-011-9261-z).
- ZHU, F., «A New Halpern-Pearl Definition of Actual Causality by Appealing to the Default World», *Axiomathes* (2022), DOI: [10.1007/s10516-021-09613-z](https://doi.org/10.1007/s10516-021-09613-z).

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst’s seminal book on typography “*The Elements of Typographic Style*”. `classicthesis` is available for both L^AT_EX and L^yX:

<https://bitbucket.org/amiede/classicthesis/>

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>

Thank you very much for your feedback and contribution.

Versione Finale dell’11 Novembre 2022 (1.0)