

Causalità

Vittorio Morato
Dipartimento di Filosofia
vittorio.morato@unipd.it

Corso di filosofia della scienza - a.a. 06-07

Versione del 6 giugno 2007

Attenzione: queste pagine sono degli appunti per le lezioni e sono pensate per i frequentanti. Sebbene siano (abbastanza) autosufficienti, non possono essere considerate come sostitutive delle letture indicate nella bibliografia del corso. Nemmeno i testi indicati in bibliografia, però, dovrebbero essere affrontati senza leggere queste pagine (sia perchè esse servono a segnalare quello che dei testi mi interessa che voi sappiate sia perchè, a volte, ho inserito qui del materiale che nei testi non c'è). Dopo le lezioni, qualcosa potrebbe essere cambiato o aggiunto. L'eventuale versione aggiornata di questo documento si potrà trovare qui:

- <http://www.filosofia.lettere.unipd.it/analitica/morato/lezione.pdf>

Prima di sostenere l'esame, sarebbe bene scaricare ed usare la versione più aggiornata disponibile.

Indice

1	Hume e la teoria regolarista della causalità	3
1.1	La teoria delle idee	3
1.2	La relazione di causa-effetto	4
1.3	L'idea di connessione necessaria	6
2	Gli sviluppi del regolarismo: cause come condizioni	10
2.1	Cause come condizioni INUS	11
2.2	Il problema degli effetti congiunti di una causa comune	13
2.3	Le critiche di Kim	15
3	La teoria controfattuale della causalità	17
3.1	Mondi possibili e controfattuali	18
3.2	Dipendenza controfattuale e dipendenza causale	21
3.3	Casi di causalità a vuoto	25
4	Letture per l'esame	29

1 Hume e la teoria regolarista della causalità

1.1 La teoria delle idee

Il dibattito contemporaneo sulla causalità è fortemente influenzato dall'impostazione datagli dal filosofo inglese David Hume (1711 - 1776).

La caratteristica principale della filosofia di David Hume è la sua *epistemologia empirista*. Nel caso di Hume, tale epistemologia può essere riassunta da quella che si potrebbe chiamare *Tesi di Hume*:

Tutte le *idee* corrispondono ad *impressioni* e viceversa

Idee e impressioni appartengono alla classe delle *percezioni*:

Impressioni: “Col termine impressione intendo tutte le nostre percezioni più vivide quando udiamo, o vediamo o sentiamo, o amiamo, o odiamo, o vogliamo”

Idee: “le idee sono le percezioni meno vivide, di cui siamo coscienti quando riflettiamo su una delle sensazioni o degli atteggiamenti sopra ricordati”¹

Visto che le idee sono, per Hume, i costituenti dei pensieri (sono quello che oggi chiameremmo *concetti*), la massima di Hume sembra implicare che non vi possa essere alcun pensiero i cui costituenti non corrispondano ad impressioni.

Messa in questo modo, però, la tesi sembra essere troppo forte e si presta a facili controesempi: infatti sembra possibile immaginare dei casi di:

- *idee senza impressioni*. Ad es.: posso immaginare una montagna dorata ma certo non posso dire di aver avuto l'impressione di una montagna dorata.
- *impressioni senza idee*. Ad es.: la vista di Roma dal Gianicolo genera in me un'impressione sensoriale ma non un'idea.

La soluzione è qualificare la massima distinguendo tra idee e impressioni *semplici* e *complesse*. Qui Hume riprende la distinzione espressa da un altro filosofo empirista a lui precedente, ossia J. Locke (1632 - 1704):

Le percezioni, le impressioni e le idee semplici sono tali da non ammettere distinzione o separazione. Quelle complesse, invece, sono opposte a queste e possono essere distinte in parti. Sebbene un particolare colore, gusto e odore siano tutte qualità presenti insieme in una mela, è facile percepire che esse non sono la stessa e che sono almeno distinguibili l'una dall'altra²

In base a questa distinzione, la tesi di Hume può essere qualificata come segue:

¹Vedi [3, p. 25]; i numeri di pagina si riferiscono alla traduzione italiana.

²Vedi [8]

Tutte le idee semplici corrispondono ad impressioni semplici e viceversa

I casi della montagna dorata e della vista di Roma sono casi, rispettivamente, di idee e impressioni complesse e quindi non rappresentano più controesempi alla tesi di Hume.

Ho caratterizzato la relazione tra impressioni ed idee usando il termine neutro “corrispondenza”. Le definizioni di impressione ed idea date sopra suggeriscono che le idee siano delle entità che risultano da una sorta di “impoverimento cognitivo” delle impressioni. Le idee, quindi, sono *prodotte* dalle impressioni.

È il fatto che le idee siano versioni impoverite delle impressioni che permette di utilizzare le idee al fine, ad esempio, di ricordare un oggetto di cui abbiamo un'impressione; sarebbe cognitivamente troppo impegnativo richiamare alla mente l'impressione e tutta l'informazione che essa codifica³.

La teoria delle idee di Hume dà luogo ad una precisa metodologia di indagine filosofica:

Metodologia di Hume: Comprendere la mente umana significa comprendere perchè pensiamo in un certo modo. Per comprendere perchè pensiamo in un certo modo dobbiamo scoprire l'*origine* dei nostri modi di pensare. Visto che, per la teoria delle idee, pensare implica la presenza di percezioni e le percezioni, a loro volta, la presenza (in un tempo presente o passato) di impressioni sensoriali, comprendere la mente umana significa ricercare nel mondo dell'esperienza sensibile le impressioni che hanno originato le idee con cui pensiamo.

1.2 La relazione di causa-effetto

Per Hume ci sono due tipi di ragionamenti:

relazioni di idee: “qualsiasi affermazione che sia certa intuitivamente o dimostrativamente. Proposizioni di questa specie si possono scoprire con una semplice operazione del pensiero, senza dipendenza alcuna da qualche cosa che esiste nell'universo”⁴

questioni di fatto: “le questioni di fatto non si possono accertare nella stessa maniera [delle relazioni di idee], nè l'evidenza della loro verità è della stessa natura della precedente. Il contrario di ogni questione di fatto è sempre possibile, poichè non può mai implicare contraddizione e viene concepito dalla mente colla stessa facilità e distinzione che se fosse del pari conforme a realtà”⁵

³L'influenza dell'approccio humeano è ben riconoscibile nelle moderne teorie rappresentazionali dei concetti; si veda ad esempio [1]

⁴Cfr. [3, p. 37]

⁵*Ibidem*

Il test per decidere se una certa proposizione P appartiene all'una o all'altra categoria è quindi quello di chiedersi se la sua negazione, $\neg P$, implica contraddizione logica.

Ad esempio, la proposizione "il sole sorgerà domani" appartiene alle questioni di fatto perchè la sua negazione "il sole *non* sorgerà domani" non implica contraddizione logica (è semmai contrario alle leggi della fisica).

Per Hume tutti i ragionamenti che riguardano questioni di fatto sono basati sulla *relazione causa-effetto*.

Tutti i ragionamenti relativi a materie di fatto sembrano fondati sulla relazione di *causa ed effetto*. Soltanto per mezzo di questa relazione possiamo andare al di là dell'evidenza della memoria e dei sensi. Se chiedete ad una persona *perchè crede a qualche fatto che non è presente*, per esempio che un suo amico si trova in campagna o in Francia, vi darà una ragione; e questa ragione sarà qualche altro fatto, come una lettera ricevuta da parte dell'amico o la conoscenza di sue risoluzioni o promesse precedenti. Un uomo che trovasse un orologio o un'altra macchina in un'isola deserta, concluderebbe che in quell'isola una volta vi siano stati degli uomini. [...] Se dunque vogliamo metter capo a una spiegazione sufficiente intorno alla natura dell'evidenza che ci assicura dei fatti dobbiamo ricercare come arriviamo alla conoscenza di causa ed effetto ([3, p. 39], corsivo mio)

Per Hume tutti i ragionamenti basati sulla relazione causa-effetto sono basati sull'esperienza.

[...] la conoscenza di questa relazione non si consegue in alcun caso mediante ragionamenti *a priori*; ma nasce interamente dall'esperienza quando troviamo che certi particolari oggetti sono costantemente congiunti tra loro. Presentiamo un oggetto ad una persona di capacità ed abilità forti quanto si voglia; *se quell'oggetto le è del tutto nuovo*, essa non riuscirà, coll'esame più accurato delle qualità sensibili di esso, a scoprire qualcuna delle sue cause e dei suoi effetti ([3, p. 41], corsivo mio)

Hume ritiene, però, che la relazione causa-effetto abbia una *doppia natura*; essa può essere considerata come:

- **Una relazione filosofica:** quando la si intende come una relazione tra oggetti ed eventi nel mondo (ogni relazione tra oggetti è una relazione filosofica. La relazione di causa-effetto intesa come relazione filosofica è una relazione oggettiva, indipendente dalla mente.
- **Una relazione naturale:** quando la si intende come una relazione tra due pensieri (che riguardano questioni di fatto). La relazione di causa-effetto tra A e B (dove A e B sono eventi) è una relazione naturale quando l'idea di A produce nella mente l'idea di B .

Tutte le relazioni sono relazioni filosofiche; non sono molte invece le relazioni filosofiche che sono anche relazioni naturali: oltre alla relazione di causa-effetto, vi sono la relazione di *somiglianza* e di *contiguità spaziale*.

Il problema adesso è capire perchè Hume si sia sentito in dovere di caratterizzare la causalità in questa maniera duale.

Per farlo è bene ricordare come a Hume si presenti il problema della causalità alla luce della metodologia menzionata sopra e della sua teoria delle idee:

I ragionamenti di tipo causale sembrano essere una parte fondamentale del nostro modo di pensare (tutti i ragionamenti che riguardano questioni di fatto sono causali). Questo significa che abbiamo dei pensieri, ad esempio, della forma “ A è la causa di B ” in cui l’idea di causa è un costituente. L’idea di causa può essere un’idea semplice o composta. Nel caso sia semplice, si dovrà scoprirne l’origine individuando l’impressione semplice che la ha generata; nel caso sia complessa, bisognerà scoprire da quali idee essa sia composta (e a quali impressioni tali idee costitutive corrispondano).

Il *Problema di Hume* è quindi il seguente:

Problema di Hume Quando pensiamo qualcosa come “ A è la causa di B ”, qual’è l’impressione (nel caso sia semplice) o le idee (nel caso sia composta) che ha generato in noi l’idea di causalità?

Hume non ritiene che l’idea di causalità sia un’idea semplice e sceglie quindi un’*approccio riduzionista*, ossia tenta di spiegare una nozione nei termini di altre che non la implicano.

Per escludere che l’idea di causalità sia semplice, è necessario escludere che quando pensiamo “ A causa B ” l’idea di causa derivi:

- da una proprietà di A
- da una proprietà di B
- da una proprietà della relazione tra A e B

Vedremo tra un po’ perchè Hume escluda la terza ipotesi. Le prime due sono escluse dal secondo pezzo citato sopra: nessuna impressione che ci derivi da proprietà di A o di B può aiutarci a determinare se A sia una causa (posso osservare un oggetto per ore ma non riuscire a determinare, solo sulla base delle sue proprietà, di cosa sia l’effetto o di cosa sia la causa).

1.3 L’idea di connessione necessaria

Uno degli esempi paradigmatici che Hume usa per discutere la natura della relazione causa-effetto è un caso di comunicazione di movimento per mezzo di un impulso ed è il seguente:

Esempio del biliardo: Si consideri un tavolo da biliardo in cui all'istante t_1 una palla P_1 si muova lungo una traiettoria d . Si consideri una palla P_2 immobile in un punto di d . Ad un istante t_2 , successivo a t_1 , P_1 colpisce P_2 , si ferma, e imprime un impulso a P_2 che si mette in movimento lungo d .

Questo sembra essere un tipico caso di relazione causale. Per Hume gli elementi "essenziali" che caratterizzano questa situazione sono i seguenti:

- *L'idea di contiguità spaziale*
- *L'idea di successione temporale*

Presumibilmente, tali idee sono idee semplici poichè corrispondono a proprietà direttamente percepibili di relazioni tra oggetti.

Si noti che fintantochè siamo interessati ad una descrizione completa di una *particolare sequenza causale di eventi*, la menzione degli oggetti coinvolti e delle loro relazioni spazio-temporali nei vari istanti di tempo sembra essere più che sufficiente. È proprio per questo motivo che Hume esclude che quando pensiamo "A causa B" l'idea di causa sia prodotta da qualche relazione tra A e B.

Diremmo, quindi, che l'idea di causa è prodotta da relazioni di contiguità spaziale e successione temporale?

Se così fosse definiremmo come causale ogni sequenza di eventi i cui costituenti siano in relazioni spaziali di contiguità e temporali di successione.

Il problema è che non riusciremmo a distinguere sequenze causali da mere *coincidenze*:

- **Coincidenza 1:** entrate in casa di amici che cominciano ad esultare. L'esultanza però non è causata dal vostro arrivo ma dal fatto che proprio in quel momento la nazionale ha segnato.
- **Coincidenza 2:** nel momento in cui chiudete un libro una matita cade. La caduta della matita proprio in quel momento è, però, semplicemente l'esito di un suo progressivo rotolamento dovuto all'inclinazione del piano.

Hume è perfettamente consapevole che manca almeno un ingrediente per qualificare una certa successione di eventi come causale e per giustificare nei nostri pensieri la comparsa di tale idea; ritiene però anche che, *se si restringe l'attenzione a singole sequenze di eventi*, la causalità *intesa come relazione filosofica*, ossia come una relazione nel mondo, non sia distinguibile da coincidenze. Se ci si limita all'osservazione di singole sequenze, non vi è nessuna proprietà metafisica che possa distinguere coincidenze da sequenze causali.

Perchè allora siamo disposti a sostenere che alcune sequenze sono causali mentre altre non lo sono?

L'ingrediente mancante per Hume è il seguente:

- *L'idea di connessione necessaria*

Riusciamo a distinguere una sequenza causale da una coincidenza perchè se due eventi A e B sono causalmente connessi saremmo disposti a dire che esiste una sorta di legame, di forza necessitante che lega A a B e che ci sia qualcosa *in virtù di cui* A causa B .

I predecessori di Hume pensavano che questa forza necessitante fosse in natura, ossia che una sequenza causale fosse composta dagli oggetti coinvolti, dalle loro relazioni spazio-temporali e da questa forza necessitante.

Per Hume non esistono forze necessitanti in natura. Del resto è fermo nel sostenere che *l'idea* di connessione necessaria sia essenziale per caratterizzare i nostri ragionamenti causali. Il problema di Hume può essere quindi riformulato come segue.

Problema di Hume: come spiegare nei termini della teoria delle idee la comparsa dell'idea di connessione necessaria?

Quali sono le impressioni che generano questa idea?

La comparsa dell'idea di connessione necessaria tra A e B ha a che fare con

- la congiunzione costante tra A e B .

Ovvero: una sequenza tra due eventi A e B è causale non solo perchè A e B sono in relazione di contiguità e successione temporale ma anche perchè *tutte le sequenze dello stesso tipo* soddisfano tali requisiti. Il caso del biliardo è una sequenza causale poichè ogni volta che due palle si scontrano in quel modo, la prima trasmette il movimento alla seconda; il caso del libro *non* è invece una sequenza causale perchè non ogni volta che un libro si chiude, una matita cade.

È l'appello alla congiunzione costante che permette di definire la teoria di Hume come una *teoria regolarista della causazione*.

Intesa come relazione filosofica (ossia da un punto di vista metafisico), quindi, si può dire che per Hume

A causa B se e solo se

- A è contiguo a B
- A precede B
- tutti gli eventi dello stesso tipo di A sono regolarmente seguiti (o sono costantemente congiunti con) da eventi dello stesso tipo di B

Definizione di causa come relazione filosofica We may define a CAUSE to be 'An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter (Vedi [2])

La caratterizzazione della causalità come relazione naturale (ossia da un punto di vista epistemico) non è ancora conclusa. Hume da una parte ritiene che le regolarità in natura siano sufficienti per spiegare la comparsa dell'idea di

connessione necessaria, dall'altra l'idea di connessione necessaria non è semplicemente il risultato dell'impressione provocata dall'osservazione di congiunzioni costanti tra eventi.

Questo sostanzialmente si deve a due motivazioni:

- Hume non ritiene che anche un gran numero di osservazioni della congiunzione tra A e B sia sufficiente per concludere che A è B sono necessariamente congiunti.
- Hume ritiene che per quanto alto sia il numero delle osservazioni che testimoniano la congiunzione tra A e B , il contenuto percettivo rimane invariato. Nessuna nuova impressione si genera solo per il fatto che osservo regolarmente la congiunzione tra due eventi. Le proprietà percepibili di una sequenza di eventi vista una volta *sono le stesse* tutte le volte che essa si ripete. La prima volta che ho osservato due palle da biliardo che si scontravano ho ricevuto esattamente le stesse impressioni della centesima volta che le ho viste scontrarsi.

Questo brano della *Ricerca* è rivelatorio:

Quest'idea di una connessione necessaria fra eventi sorge da un numero di casi simili in cui si verifica la *costante congiunzione* di eventi, mentre quell'idea non può mai essere suggerita da qualcuno solo di questi casi, anche se considerato in tutte le luci e le posizioni possibili. *Ma in un certo numero di casi non c'è nulla di diverso da quello che c'è in ciascun caso singolo* che si suppone sia esattamente simile agli altri, eccetto soltanto che dopo il ripetersi di casi simili, la mente viene spinta dall'*abitudine* in base al presentarsi di un evento, ad attendere l'evento che di solito lo accompagna ed a credere che esso si verificherà (Cfr. [3, p. 117], corsivi miei)

e subito dopo:

In tutti i singoli casi dell'operare dei corpi o delle menti, non c'è nulla che produca qualche impressione, nè che possa di conseguenza suggerire qualche idea di potere e connessione necessaria. Ma quando si mostrano molti casi uniformi e lo stesso oggetto è sempre seguito dallo stesso evento, allora incominciamo ad accogliere la nozione di causa e di connessione. Allora noi *sentiamo* un nuovo sentimento o una nuova impressione, cioè una connessione abituale nel pensiero e nell'immaginazione fra un oggetto e quello che di solito lo accompagna

A generare l'idea di connessione necessaria è quindi un'impressione generata dalle operazioni della mente (e non dagli oggetti fisici nel mondo) ossia dall'abituale connessione nell'immaginazione di due eventi.

L'opinione secondo la quale vi siano forze necessitanti in natura dipende, secondo Hume, dalla tendenza della mente a "proiettarsi" fuori di essa e di

attribuire agli oggetti delle proprietà che sono in verità le proprietà con cui la mente pensa ad essi.

Definizione di causa come relazione naturale A CAUSE is an object precedent and contiguous to another and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other (vedi [2])

Da un punto di vista metafisico, uno dei problemi che rimangono aperti dell'approccio di Hume è che il mero appello alle regolarità non sembra essere sufficiente per distinguere sequenze causali (tra tipi di eventi) e *generalizzazioni vere*.

Ad esempio:

- **Generalizzazione 1:** il giorno è costantemente congiunto con la notte (nel senso che c'è una costante successione di giorno e notte) ma non diremmo che il giorno/notte è la causa della notte/giorno
- **Generalizzazione 2:** ogni mattina mi sveglio alle 7:49, accendo la radio e alle 7:50 inizia il giornale radio; il mio risveglio è costantemente congiunto all'inizio del giornale radio ma non diremmo che il mio risveglio è la causa del giornale radio.

Qui la natura della duale della relazione di causalità torna utile: ciò che per Hume distingue regolarità causali da quelle non causali è che non associamo alle seconde alcuna idea di connessione necessaria.

Come vedremo buona parte del dibattito successivo consisterà nel distinguere le regolarità "buone" da quelle "cattive".

2 Gli sviluppi del regolarismo: cause come condizioni

Il cuore dell'approccio regolarista consiste nel sostenere che un evento a causa un evento b se un evento del tipo a cui appartiene a è seguito regolarmente da un evento a cui appartiene b .

Un modo per chiarire cosa vuol dire che ad un tipo di evento "segue regolarmente" un altro tipo di evento è sostenere che il primo tipo di eventi è, in qualche modo, una *condizione* per il secondo tipo di eventi.

Ma cosa vuol dire che un evento è una condizione per un altro evento? Due sono i casi interessanti al fine di caratterizzare la causalità:

- **Condizione sufficiente** Un evento a è una condizione sufficiente per un evento b se e solo se ogni volta che si verifica a si verifica anche b
- **Condizione necessaria** Un evento a è una condizione necessaria per un evento b se e solo se ogni volta che si è verificato b si è verificato anche a

2.1 Cause come condizioni INUS

J. Mackie in [9] ha presentato una teoria, diventata molto influente, col fine di chiarire quale possa essere il ruolo delle nozioni di condizione necessaria e sufficiente per chiarire la natura della causalità.

Partiamo con un esempio:

Esempio della casa Si supponga che in un'abitazione si sviluppi un incendio che però viene estinto prima che l'abitazione venga completamente distrutta. Gli esperti, dopo accurate indagini, concludono che l'incendio è stato causato da un corto-circuito

In che senso gli esperti stabiliscono questa connessione causale?

- Il corto circuito **non è una condizione sufficiente**: come da definizione sopra, per falsificare la tesi che qualcosa sia una condizione sufficiente bisogna falsificare un enunciato della forma “tutti gli A sono B ”. Per falsificare un tale tipo di enunciato basta presentare un caso di un A che non è B . Nel caso della casa, questo vuol dire sostenere che un corto circuito poteva anche accadere ma senza provocare un incendio (nel caso, ad esempio, ci fosse stato un salvavita o di un sistema anti-incendio funzionante). La condizione sufficiente per l'incendio, in quel caso, sarebbe il corto circuito, l'assenza di un salva-vita, l'assenza del sistema anti-incendio ma non il solo corto-circuito.
- Il corto circuito **non è una condizione necessaria**: sono concepibili casi in cui si verifica un incendio nella casa in questione ma nessun corto circuito (il corto circuito non era l'unico modo in cui poteva verificarsi un incendio in quella casa). Ad esempio, nel caso in cui un'incendio sia doloso o provocato, ad esempio, da una sigaretta lasciata accesa.

Il corto circuito, quindi, pur essendo la causa dell'incendio, non è nè la condizione necessaria nè la condizione sufficiente. Da questo segue che, in generale:

se A è la causa di B da questo non segue che A è una condizione necessaria e sufficiente per B .

Attenzione: quello che questo esempio mostra è che c'è almeno un caso in cui la causa di un evento non è la sua condizione necessaria e sufficiente e quindi confuta la tesi che *in generale* le cause possano essere così definite (confuta la tesi che tutte le cause siano condizioni necessarie e sufficienti). Questo è compatibile con l'esistenza di altri casi in cui la causa di un evento sia proprio la sua condizione necessaria e sufficiente di un evento.

Analizziamo meglio la situazione:

Si consideri l'insieme C formato dai seguenti elementi:

{ corto circuito, \neg salva-vita, \neg sistema anti-incendio }

L'insieme C è una *condizione non necessaria ma sufficiente* per l'incendio.

Quali sono i rapporti tra le parti di C ?

Tra i membri di C , il corto circuito è una parte indispensabile poichè senza il corto circuito la combinazione degli altri elementi non avrebbe provocato l'incendio. Se tolgo il corto circuito da C , C non sarebbe più una condizione sufficiente mentre se tolgo, ad esempio, l'assenza di salva vita magari l'incendio si sarebbe sviluppato lo stesso.

Questo si deve al fatto che quando diciamo che una certa condizione X è sufficiente non intendiamo anche dire che X è una *condizione minimale*, ossia che tutti gli elementi di X sono indispensabili; qualche elemento di X potrebbe essere superfluo; ad esempio, l'insieme:

{ corto circuito, \neg salva-vita, \neg sistema anti-incendio, metodo di costruzione }

è comunque una condizione sufficiente ma si può presumere che il metodo di costruzione non giochi un ruolo indispensabile. Si chiamano *condizioni sufficienti minimali* quelle condizioni sufficienti in cui nessuno degli elementi è ridondante.

Da questo deriva che il corto circuito è una parte *insufficiente ma necessaria di una condizione non necessaria ma sufficiente*. Il termine "condizione INUS" è l'acronimo della traduzione inglese di questa espressione:

Insufficient but Necessary part of an Unnecessary but Sufficient condition.

Visto che il corto circuito è una condizione INUS, sappiamo che nel caso della casa il corto circuito è la causa dell'incendio e il caso della casa è sufficientemente paradigmatico, possiamo concludere che le cause sono condizioni INUS.

Mettiamola in termini generali:

Si consideri un particolare evento P . A P saranno associate una serie di condizioni complesse ciascuna delle quali è sufficiente minimale ma non necessaria. Le condizioni complesse sufficienti e minime saranno delle *coniunzioni* della forma:

- $A \wedge B \wedge \neg C$
- $D \wedge \neg E \wedge F$
- $\neg G \wedge H \wedge I$
- ...

La condizione necessaria e sufficiente per P sarà data dalla *disgiunzione* di tutte queste congiunzioni e quindi sarà della forma:

$$(A \wedge B \wedge \neg C) \vee (D \wedge \neg E \wedge F) \vee (\neg G \wedge H \wedge I) \vee \dots$$

Si noti che poichè abbiamo stipulato che i congiunti sono condizioni minimali sufficienti, ciascun membro della congiunzione sarà una condizione INUS.

Si consideri, ad esempio, A . Per semplificare e arrivare al massimo della generalità sostituiamo gli altri elementi con cui A è congiunto con una variabile X e sostituiamo tutti gli altri disgiunti con un'altra variabile Y . La formula sopra si trasforma in:

$$(A \wedge X) \vee Y$$

La definizione generale di condizione INUS è quindi la seguente:

Definizione di condizione INUS: A è una condizione INUS di P se e solo se, per qualche X e per qualche Y , $(A \wedge X) \vee Y$ è una condizione necessaria e sufficiente per P ma nè A nè X sono una condizione sufficiente per P .

La “forma logica” di una regolarità è quindi per Mackie la seguente:

$$(A \wedge X) \vee Y \leftrightarrow P$$

Una cosa importante da notare è che A sta per una specifica condizione per P mentre X e Y sono delle variabili e quindi possono prendere come valore tutte le condizioni per P . In particolare la definizione non esclude che:

- $X = A$, ossia che A sia *da solo* una condizione minimale sufficiente per P . In questo caso la regolarità avrebbe la forma $A \vee Y \leftrightarrow P$.
- $Y = A$, ossia che $(A \wedge X)$ sia l'unica condizione minimale sufficiente per P (cioè una condizione necessaria e sufficiente); in questo caso la regolarità avrebbe la forma $AX \leftrightarrow P$.
- $X = A$ e $Y = A$, ossia che A da sola sia una condizione necessaria e sufficiente per P ; in questo caso la regolarità avrebbe forma $A \leftrightarrow P$.

In tutti questi tre casi A è un po' di più che una condizione INUS per P . Nei confronti di P , A può essere anche nella relazione più forte possibile (dato il quadro generale) ossia una condizione necessaria e sufficiente.

Quindi quello che la teoria di Mackie suggerisce è che la causa di un evento P è, *come minimo*, una condizione INUS:

A è la causa di un evento P se e solo se A è **almeno** una condizione INUS per P .

2.2 Il problema degli effetti congiunti di una causa comune

Uno dei problemi della teoria di Mackie è che non riesce a distinguere regolarità causali da *effetti congiunti di cause comuni*. Per spiegare cosa siano questi ultimi, è bene usare un esempio.

Esempio delle fabbriche: Tutti gli operai delle fabbriche inglesi smettono di lavorare alle 5 del pomeriggio, *dopo* il suono di una sirena. Quindi, in un giorno qualsiasi, gli operai della fabbrica F_1 , che si trova nella periferia di Londra, smettono di lavorare al suono della loro sirena s_1 . Lo stesso giorno, alla stessa ora, anche gli operai della fabbrica F_2 , che invece si trova nella periferia di Manchester, smettono di lavorare al suono della loro sirena s_2 .

È plausibile sostenere che (in una giornata *standard* in cui la sirena non suona per errore all'ora sbagliata) il suono della sirena s_1 è la causa della fine della giornata lavorativa degli operai di F_1 mentre il suono di s_2 è la causa della fine della giornata lavorativa degli operai di Manchester. Quindi, il suono di s_1 è (almeno) la condizione INUS della fine del lavoro a Londra e il suono di s_2 è (almeno) la condizione INUS per la fine del lavoro a Manchester. Inoltre la congiunzione del suono della sirena alle 17:00 con altre condizioni di supporto (ad esempio che nessuno per scherzo disattivi la sirena prima della 17:00) è una condizione sufficiente (ma non necessaria) perchè i lavoratori di F_1 o di F_2 smettano di lavorare.

Purtroppo, sulla base di quanto detto, si può anche dimostrare che:

il suono di s_2 è la condizione INUS per la fine del lavoro in F_1 e viceversa.

ossia, il suono della sirena a Manchester può esser fatto risultare come la causa della fine della giornata lavorativa degli operai di Londra (e viceversa).

Vediamo come:

- sia A il suono di s_2 in F_2 (la fabbrica di Manchester)
- sia X la congiunzione di due qualsiasi condizioni tali che
 - la prima assicuri che i lavoratori di F_1 (i londinesi) smettano effettivamente di lavorare alle 17:00 (ad esempio, che la loro sirena s_1 non sia danneggiata o manomessa e che quindi non continuino a lavorare oltre le 17:00 o che proprio quel giorno non siano costretti a straordinari)
 - la seconda che assicuri che s_2 (la sirena di Manchester) sia perfettamente regolata e funzionante (e che quindi suoni proprio alle 17:00).
- sia Y una qualsiasi congiunzione di condizioni tale da essere sufficiente perchè i lavoratori di Londra, ad un'ora qualsiasi, smettano di lavorare (ad esempio, uno sciopero o un black out).

$(A \wedge X) \vee Y$ è una condizione necessaria e sufficiente perchè gli operai londinesi smettano di lavorare. $A \wedge X$ è una condizione sufficiente ma non necessaria perchè i londinesi smettano di lavorare (rappresenta la condizione sufficiente

secondo cui gli operai di F_1 smettono di lavorare alle 17:00 al suono della sirena). In $A \wedge X$, A è una parte *insufficiente ma necessaria* perchè è l'unica, per come è stato costruito l'esempio, che codifica l'informazione che sono le 17:00. Quindi A , il suono della sirena a Manchester, è una condizione INUS perchè i lavoratori di Londra smettano di lavorare alle 17:00.

2.3 Le critiche di Kim

J. Kim in [4] ha rivolto alla teoria di Mackie una serie di critiche che mettono in luce alcuni problemi di natura *logica* ed *ontologica*.

Problema categoriale: Per comprendere questo problema ci si deve innanzitutto chiedere:

Quando diciamo “ A causa B ” che tipo di entità sono A e B ?

ossia, che tipo di entità sono cause ed effetti? Fino ad ora abbiamo parlato (piuttosto disordinatamente) di *condizioni*, *stati*, *fatti*, *eventi*. A giudicare proprio dagli esempi di Mackie sembra, però, che le entità più adatte siano *eventi*; (ad esempio, quando diciamo “il corto circuito ha causato l'incendio” o “la pugnalata di Bruto causò la morte di Cesare”). Se sono gli eventi ad essere i costituenti delle relazioni causali, è bene chiedersi

Con che tipo di espressioni ci riferiamo ad eventi?

Normalmente, ci riferiamo ad eventi usando *termini singolari* (“l'incendio”, “il corto circuito”, “la pugnalata”), ossia espressioni che possono essere soggetto di frasi dichiarative. Il problema è che Mackie caratterizza le regolarità causali per mezzo di *connettivi enunciativi* (disgiunzioni e congiunzioni), ossia per mezzo di particelle che invece connettono *enunciati*, non termini singolari. Ma i connettivi enunciativi non hanno alcun senso per termini singolari:

- per quanto riguarda la *congiunzione*:
 - mentre $A \wedge B$ è un enunciato e può essere vero o falso, “la pugnalata di Bruto e la pugnalata di Cassio” non è un enunciato.
 - a volte usiamo il connettivo “e” tra due termini singolari come in “Giovanni e Maria hanno mangiato molto pesce” ma in questi casi quello che vogliamo dire è “Giovanni ha mangiato molto pesce e Maria ha mangiato molto pesce”; quando invece diciamo “la pugnalata di Bruto e quella di Cassio causarono la morte di Cesare” non intendiamo dire che “la pugnalata di Cassio causò la morte di Cesare e la pugnalata di Bruto causò la morte di Socrate”.
- per quanto riguarda la *negazione*:

- in generale, la negazione non sembra essere definita per termini singolari: cosa significa “non-Cesare”, “non-il marito di Francesca”? Nel caso degli eventi: se A è un termine singolare (“la pugnolata di Cassio”) cosa significa $\neg A$?, Ossia cosa significa “non-la pugnolata di Cassio”?
- possiamo anche assumere che la negazione sia definita per eventi ma se diciamo che “non-la pugnolata di Cassio” significa “l’assenza della pugnolata di Cassio” a che evento ci stiamo riferendo? Tutti gli eventi che non siano la pugnolata di Cassio possono essere l’assenza della pugnolata di Cassio. È cioè implausibile pensare che dato un evento A , esista un *unico* evento $\neg A$.

Il problema della definizione della nozione di condizione minimale e sufficiente. Come abbiamo visto tale nozione gioca un ruolo essenziale nella teoria di Mackie; una condizione $A \wedge B \wedge C$ è una condizione minimale sufficiente per P se e solo se (i) è una condizione sufficiente per P e (ii) non contiene alcun elemento ridondante.

In maniera più esplicita e generale di quanto fatto da Mackie si può definire tale nozione come segue:

L’evento $A_1 \wedge A_2 \wedge \dots \wedge A_n$ è una condizione minimale e sufficiente per un evento P se e solo se, per ciascun i (tra 1 e n), $A_1 \dots A_{i-1}, A_{i+1}, \dots A_n$ non è una condizione sufficiente per P

Non spaventatevi! Questa definizione dice (per mezzo della clausola complessa “per ciascun i (tra 1 e n), $A_1 \dots A_{i-1}, A_{i+1}, \dots A_n$ ”) che se tolgo un elemento a caso da $A_1, \dots A_n$ (nel nostro caso, quello che sta tra $i-1$ e $i+1$), allora essa non è più una condizione sufficiente; ovvero: ogni elemento di $A_1, \dots A_n$ è indispensabile per renderla una condizione sufficiente per P .

Il problema di questa caratterizzazione (che semplicemente esplicita quella di Mackie) è il seguente:

Data una qualunque condizione minimale $A_1, \wedge \dots \wedge A_n$ per P , si può mostrare che un qualunque evento C (con alcune limitazioni che adesso vedremo) è una condizione INUS per P

L’origine del problema ha sempre a che fare con il fatto che Mackie sceglie di caratterizzare le regolarità causali per mezzo di connettivi che appartengono alla logica enunciativa.

In particolare, si supponga che $A \wedge \dots \wedge B$ (un caso semplice di $A_1, \wedge \dots \wedge A_n$) sia una condizione minimale sufficiente per P . Da questo segue che:

$$C \wedge (\neg C \wedge A) \wedge B$$

è una condizione minimale e sufficiente per P . Vediamo perchè:

Dalla definizione data sopra segue che $C \wedge (\neg C \vee A) \wedge B$ è una C.M.S.⁶ per P se e solo se, nel caso si tolga o C , o $(\neg C \vee A)$ o B il risultato non è più una C.M.S. per P .

- *se tolgo C* : rimane $(\neg C \vee A) \wedge B$ ma non è più sufficiente per P poichè $(\neg C \vee A)$ potrebbe essere vera in una situazione in cui $\neg C$ è vera e A è falsa mentre abbiamo assunto che la presenza di A è essenziale per P . Per come è costruita la formula, solo la presenza di C mi garantisce che $(\neg C \vee A)$ sia vera perchè A sia vera; infatti, se C è vera $\neg C$ è falsa e quindi $(\neg C \vee A)$ è vera perchè A è vera.
- *se tolgo $(\neg C \vee A)$* : rimane $C \wedge B$ ma tale formula non può essere una C.M.S. poichè, per assunzione (ossia che $A \wedge B$ è una C.M.S.), abbiamo bisogno di A .
- *se tolgo B* : rimane $C \wedge (\neg C \vee A)$ ma, per i motivi di cui sopra, non è una C.M.S. perchè adesso manca B e B è essenziale per P .

Tenete presente che tutto il ragionamento assume due *restrizioni* ossia che che $\neg C \vee A = A$ e che $C \wedge B$ siano sufficienti per P .

Ne segue che $C \wedge (\neg C \vee A) \wedge B$ è una condizione sufficiente e minimale per P . Vista la sua forma logica abbiamo anche che C è una condizione INUS di $C \wedge (\neg C \vee A) \wedge B$; quindi in termini generali la teoria di Mackie implica che

Se $A_1 \wedge \dots \wedge A_n$ è una C.M.S. per P , allora per qualsiasi C (a meno delle due restrizioni dette sopra), C è una condizione INUS per P

3 La teoria controfattuale della causalità

La teoria controfattuale della causalità afferma che:

“(l’evento) A causa (l’evento) B ” va spiegato nei termini di *condizionali controfattuali* del tipo “se A non fosse accaduto, B non sarebbe accaduto”

Un condizionale controfattuale è un condizionale, spesso al tempo congiuntivo, il cui antecedente esprime una proposizione contraria al modo in cui le cose stanno di fatto (da cui il nome controfattuale)⁷. Esempi di enunciati controfattuali sono: “Se mi buttassi dalla finestra, mi farei molto male”, “Se Mario fosse andato al cinema, avrebbe incontrato Luisa”, “Se non avessi frequentato questo corso, non avrei mai imparato nulla di filosofia della scienza”, “Se tutti usassero i mezzi pubblici, le fabbriche di automobili fallirebbero”, “Se i canguri non avessero la coda, perderebbero l’equilibrio” *etc.*

⁶C.M.S. sta per “condizione minimale sufficiente”.

⁷Come vedremo in seguito, però, ci sono anche controfattuali per così dire “devianti” che hanno un antecedente vero.

Sebbene le analisi controfattuali della causalità siano diventate popolari solo di recente, colui che per primo usò dei controfattuali per definire la causalità fu proprio Hume.

In [3, p. 119], Hume definisce la causa nel modo seguente:

Possiamo così definire una causa: è un oggetto seguito da un altro oggetto e dove tutti gli oggetti simili al primo sono seguiti da oggetti simili al secondo. *In altre parole:* dove se il primo oggetto non fosse esistito, non sarebbe esistito nemmeno il secondo.

Le “altre parole” di Hume, però, non sono una mera riformulazione delle prime nè sono ad esse equivalenti. Hume, quindi, ha fornito due diverse definizioni di causa. La prima è quella che egli di fatto esplicita ed è all’origine della tradizione regolarista; la seconda o è stata del tutto tralasciata o è stata erroneamente considerata una riformulazione della prima. La ragione del primo atteggiamento era la difficoltà di collocare i ragionamenti controfattuali nell’ambito dell’epistemologia empirista che ha caratterizzato le discussioni sulla causalità da Hume ai neo-positivisti logici.

3.1 Mondi possibili e controfattuali

Il motivo per cui gli enunciati controfattuali suscitano scetticismo presso i filosofi di impostazione empirista è dovuto al fatto che tali condizionali sembrano fare riferimento a *possibilità non realizzate*. Difficilmente si può dire che noi conosciamo possibilità non realizzate per mezzo dei sensi e quindi tali entità non possono far parte di alcuna teoria.

Per evitare ogni riferimento a possibilità non realizzate, le prime teorie dei controfattuali furono teorie *meta-linguistiche* che tentarono di analizzare i controfattuali solo per mezzo di relazioni di implicazione tra enunciati.

Un controfattuale è vero, o per lo meno asseribile, secondo tali teorie, se e solo se il suo antecedente, insieme a una serie di altre premesse “di supporto”, implica il conseguente. Ad esempio, consideriamo il controfattuale “Se Mario fosse andato al cinema, avrebbe incontrato Luisa”. Se riteniamo che esso sia vero, non basterà assumere che Mario sia andato al cinema: il fatto che Mario sia andato al cinema non è sufficiente di per sè a garantirci che abbia incontrato Luisa e sicuramente non lo implica; è necessario che si verifichino un certo numero di condizioni di supporto che garantiscano che Mario e Luisa, una volta al cinema, si incontrino effettivamente (che si vedano, che si riconoscano, che il cinema non abbia due entrate o due uscite, *etc.*). Se vogliamo spiegare perchè il nostro controfattuale è vero nei termini della teoria metalinguistica, queste condizioni vanno quindi specificate.

In generale un controfattuale del tipo “se fosse A , allora sarebbe B ” va analizzato per mezzo di un argomentazione del tipo:

Date come premesse A e s_1, \dots, s_n , posso derivare B

dove s_1, \dots, s_n sono le premesse di supporto. Il controfattuale a questo punto o è l'enunciato il cui significato è che tale argomento esiste o è una forma abbreviata di un tale argomento (ossia non è un vero enunciato).

Il problema principale di questo approccio è quello di trovare un modo sistematico di specificare con precisione quale sia l'insieme delle premesse di supporto $s_1 \dots s_n$. Tali difficoltà non fecero che aumentare lo scetticismo verso l'utilità dei controfattuali in filosofia.

Negli anni '70 del '900, una volta caduti alcuni dogmi metodologici tipici dell'impostazione empirista, si decise di imboccare un'altra strada.

David Lewis in [6] sviluppa un'articolata teoria dei controfattuali (diventata, sostanzialmente, la teoria *standard*) basata su quest'idea:

il controfattuale “Se mi buttassi dalla finestra, mi farei molto male” significa qualcosa come: in ogni stato di cose possibile in cui mi butto fuori dalla finestra e che assomiglia il più possibile al modo in cui di fatto sono andate le cose, allora mi faccio molto male

Di questa caratterizzazione gli elementi prominenti sono quelli di:

- stato di cose possibile
- somiglianza tra stati di cose possibili (rispetto al modo in cui sono di fatto andate le cose)

Entrambi sono stati resi precisi (o se non altro “gestibili” da un punto di vista logico e filosofico) per mezzo della nozione di *mondo possibile*.

Un mondo possibile è *un modo alternativo in cui le cose sarebbero potute andare*. Per rendere l'idea, consideriamo un esempio “da laboratorio”:

considerate una scatola (chiamiamola w^*) ad un certo istante t in un cui ci siano tre oggetti a, b, c ; a gode della proprietà P , b gode della proprietà Q e c gode della proprietà R ; in w^* non ci sono altri oggetti nè altre proprietà. La scatola è completamente individuata dagli oggetti e dalle proprietà godute all'interno di essi. Il modo in cui le cose stanno dentro w^* è completamente descritto dal fatto che a gode di P , b di Q e c di R . Nulla vieta, però, che in w^* e proprio nell'istante t , le cose sarebbero potute andare diversamente; ad esempio:

- a avrebbe potuto godere della proprietà Q invece che di P (e tutto il resto rimanere invariato)
- sarebbe potuto esistere un quarto oggetto d con la proprietà Q (e tutto il resto rimanere invariato)
- b sarebbe potuto non esistere (e tutto il resto rimanere invariato)
- a avrebbe potuto godere della proprietà Q invece di P e b sarebbe potuto non esistere (e tutto il resto rimanere invariato)

- *etc.*

Ciascuno dei punti sopra rappresenta un modo in cui le cose sarebbero potute andare in w^* a t e ciascuno può essere identificato con un nome: w_1, w_2, w_3, \dots ; si noti che ogni minima differenza in w^* è sufficiente per essere un modo alternativo in cui le cose sarebbero potute andare in w^* . Ora: se con w^* intendiamo non una scatola ma il nostro mondo (in senso lato come il nostro universo o “la realtà”) ad un certo istante, i vari w_1, w_2, \dots , saranno i modi in cui le cose sarebbero potute andare diversamente nel nostro mondo a t : essi saranno, cioè, dei *mondi possibili*.

Una volta che abbiamo caratterizzato i mondi possibili come l’infinito numero di modi in cui le cose sarebbero potute andare diversamente, possiamo, secondo Lewis, *confrontare i gradi di somiglianza complessiva dei vari mondi possibili rispetto al mondo attuale* (per “mondo attuale” intendo il nostro mondo, la realtà). La somiglianza è complessiva perchè non si confrontano pezzi di mondi possibili con pezzi di mondi possibili ma interi mondi possibili con mondi possibili (o con il mondo attuale). Per esempio: un mondo possibile che differisce da quello attuale solo per il fatto che oggi in classe c’è uno studente in meno è più simile al mondo attuale di quello che differisce dall’attuale per il fatto che oggi è scoppiata una guerra tra Stati Uniti ed Iran.

La relazione di *somiglianza complessiva* tra mondi possibili deve soddisfare due requisiti:

- presi due mondi possibili a caso, essi devono essere comparabili.
- il mondo attuale è il più simile a sè stesso.

Si noti che non si è richiesto che, dato un insieme di mondi possibili, vi sia un unico mondo possibile più simile all’attuale.

I giudizi di somiglianza complessiva sono, a volte, un po’ vaghi ma non così vaghi da impedire la comunicazione: ad esempio, pur nella vaghezza, sembriamo essere capaci di confrontare entità abbastanza complesse come persone (“Mario è più simile a Luisa di quanto lo sia Antonia”) o anche città (“Vicenza è più simile a Padova di quanto lo sia Los Angeles”)

Visto che abbiamo, almeno caratterizzato, le nozioni di mondo possibile e somiglianza tra mondi possibili, possiamo ora riformulare le condizioni di verità di un condizionale controfattuale in maniera rigorosa come segue:

il controfattuale “se mi buttassi fuori dalla finestra, mi farei molto male” è vero (nel mondo attuale) se e solo se (i) non ci sono mondi possibili in cui io mi butto fuori dalla finestra (ii) esiste almeno un mondo possibile in cui mi butto dalla finestra e mi faccio molto male più simile al mondo attuale di un mondo possibile in cui mi butto dalla finestra e non mi faccio male.

La clausola (i) serve solo per rendere conto dei controfattuali che sono vacuamente veri (un controfattuale con l'antecedente che non è vero in nessun mondo possibile è un controfattuale con l'antecedente *impossibile* e si ritiene, di solito, che da un antecedente impossibile tutto segua; ad esempio: quando diciamo “se gli asini volassero, allora ... ci sentiamo autorizzati a sostituire ... con qualsiasi enunciato e riteniamo che il condizionale controfattuale risultante sia vero); *la clausola che interessa è solo* (ii) e corrisponde all'idea (espressa all'inizio) secondo cui un controfattuale è vero se ci si allontana di meno dalla realtà concependo una situazione in cui l'antecedente sia vero con il conseguente di una in cui l'antecedente sia vero ma non il conseguente.

Nel nostro caso, possiamo concludere che il controfattuale è vero poichè i mondi in cui mi butto fuori dalla finestra e mi faccio male sono più simili a quelli in cui mi butto e non mi faccio niente (un mondo possibile di questo tipo richiederebbe delle variazioni più significative rispetto al mondo attuale: ad esempio, che l'aula in cui stiamo facendo lezione sia al piano terra e non al secondo piano, che vi sia un camioncino con dei materassi proprio sotto la finestra, che vi sia una sospensione delle leggi fisiche proprio nel momento in cui mi butto, *etc.*

Prima di procedere, è bene specificare un po' meglio i criteri di confronto tra mondi possibili. Si consideri il seguente enunciato:

Se il presidente degli Stati Uniti avesse premuto il bottone rosso, sarebbe scoppiata una guerra nucleare

Intuitivamente saremmo disposti a dire che un tale controfattuale è vero. Dalle condizioni date sopra, quindi, ci aspetteremmo che il controfattuale sia vero perchè il mondo possibile in cui il presidente preme il bottone e scoppia una guerra è più simile al mondo attuale del mondo in cui il presidente preme il bottone ma non scoppia una guerra. Il problema, però, è che (almeno superficialmente) un mondo in cui è scoppiata una guerra nucleare è molto più diverso dal mondo attuale di quello in cui non è scoppiata una guerra nucleare. Il problema si risolve riconsiderando meglio l'evento: un mondo in cui il presidente schiaccia il bottone ma non succede niente è un mondo in cui sono accaduti vari *miracoli*, ossia sono state violate alcune leggi di natura. Quindi, nonostante le apparenze, un mondo in cui vengono violate le leggi di natura del mondo attuale è più diverso da quest'ultimo di un mondo in cui non vengono violate anche se il primo è più simile all'attuale “esteriormente”.

3.2 Dipendenza controfattuale e dipendenza causale

Lewis in [5] applica la sua teoria dei controfattuali alla causalità. La strada per giungere dalla caratterizzazione delle condizioni di verità dei condizionali controfattuali alla caratterizzazione della causalità nei termini di controfattuali è composta dai seguenti passi:

- **Primo passo:** definire la nozione di *dipendenza controfattuale*

- **Secondo passo:** definire la nozione di *dipendenza causale* usando la nozione dipendenza controfattuale
- **Terzo passo:** definire la causalità usando la nozione di dipendenza causale

Primo passo: dipendenza controfattuale La relazione di dipendenza controfattuale si ha tra due classi di proposizioni alternative possibili quando ciascun elemento della prima classe può formare l'antecedente di un controfattuale (vero) di cui il conseguente è un elemento della seconda classe.

Consideriamo un esempio:

Esempio del barometro: Nella terrazza della vostra casa c'è un barometro. Considerate tutte le possibili proposizioni r_1, \dots, r_n che specificano, per un dato istante di tempo t , diverse letture (del tipo "il valore del barometro a t è x_1 " "il valore del barometro a t è x_2 ", *etc*). p_1, \dots, p_n sono invece le proposizioni che specificano, sempre a t , le variazioni della pressione atmosferica circostante ("la pressione a t è y_1 ", "la pressione a t è y_2 , *etc*).

Se il barometro funziona correttamente, la verità delle proposizioni espresse dagli enunciati del primo gruppo (le letture) *dipende* dalla verità delle proposizioni espresse dal secondo gruppo di enunciati (le variazioni di pressione). Questa dipendenza può essere colta per mezzo di controfattuali. Tra tutti gli r e tutti i p si possono, infatti, formare i seguenti controfattuali:

- se fosse p_1 , allora sarebbe r_1
- se fosse p_2 , allora sarebbe r_2
- ...
- se fosse p_n , allora sarebbe r_n

Questo è un caso, secondo Lewis, in cui le varie possibili letture del barometro *dipendono controfattualmente* dai possibili valori della pressione circostante.

In generale, la dipendenza controfattuale può essere definita come segue:

Sia A_1, \dots, A_n una famiglia di possibili proposizioni nessuna delle quali sia compossibile con un'altra e sia C_1, \dots, C_n un'altra famiglia di proposizioni (della stessa dimensione e con la stessa proprietà di non essere compossibili). I C *dipendono controfattualmente* dagli A se e solo se tutti i controfattuali della forma "se fosse A_1 , allora sarebbe C_1 ", ..., "se fosse A_n , allora sarebbe C_n sono veri.

Secondo passo: dipendenza causale La relazione di dipendenza controfattuale è una relazione tra famiglie di proposizioni. La causalità, invece, è una relazione tra eventi.

La nozione di *dipendenza causale* è un'estensione della nozione di dipendenza controfattuale in modo che si possa applicare ad eventi (è la nozione di dipendenza controfattuale applicata agli eventi)

Ma come si fa ad estendere ad eventi la relazione di dipendenza controfattuale per proposizioni?

Se ad ogni evento corrispondesse una proposizione, potremmo definire la relazione di dipendenza causale tra due eventi come una relazione di dipendenza controfattuale tra le corrispondenti proposizioni.

In effetti, questa corrispondenza è abbastanza facile da trovare (almeno dato la cornice logica in cui Lewis si muove): basta associare a ciascun evento la proposizione che afferma che quell'evento avviene. Dato un evento e , ad esso si può associare la proposizione $O(e)$, ossia la proposizione che ci dice che e avviene. $O(e)$ è una proposizione che sarà vera in tutti i mondi possibili in cui e avviene e falsa in tutti i mondi possibili in cui e non avviene.

La relazione di dipendenza causale tra eventi distinti a e b sarà quindi una relazione di dipendenza controfattuale tra $O(a)$ e $O(b)$.

In generale, quindi, la relazione di *dipendenza causale* può essere definita come segue:

Siano c_1, \dots, c_n e e_1, \dots, e_n due famiglie di eventi (della stessa dimensione) tali che nessun c sia compossibile con un altro c e nessun e sia compossibile con un altro e . La famiglia e_1, \dots, e_n *dipende causalmente* dalla famiglia c_1, \dots, c_n se e solo se la famiglia $O(e_1), \dots, O(e_n)$ dipende controfattualmente dalla famiglia $O(c_1), \dots, O(c_n)$.

La definizione si può semplificare se invece di prendere in considerazione famiglie di eventi se ne prendono in considerazione solo due.

Dato un qualsiasi evento possibile e , ci saranno dei mondi possibili in cui e avviene e dei mondi possibili in cui e non avviene; ad e sarà quindi possibile associare due proposizioni che tra di esse non sono compossibili, ossia la proposizione che e avviene $O(e)$ e la proposizione che e non avviene $\neg O(e)$.

Dati due eventi possibili c ed e , quindi, avremo per c la coppia $\{O(c), \neg O(c)\}$ e per e la coppia $\{O(e), \neg O(e)\}$.

A questo punto possiamo affermare che:

l'evento e dipende causalmente dall'evento c se e solo se la famiglia di proposizioni $\{O(e), \neg O(e)\}$ dipende controfattualmente dalla famiglia $\{O(c), \neg O(c)\}$, ossia se sono veri i seguenti due controfattuali:

- Se fosse $O(c)$, sarebbe $O(e)$ (se si verificasse c , si verificherebbe anche e)
- Se non fosse $O(c)$, non sarebbe $O(e)$ (se non si verificasse c non si verificherebbe anche e)

Questa definizione è stata data, al fine di raggiungere il massimo della generalità, e quindi è stata data per eventi *possibili*. Ovviamente, noi siamo più interessati a capire cosa sia la dipendenza causale tra eventi che di fatto accadono. Se c ed e sono eventi che di fatto accadono, la definizione sopra si semplifica. Infatti se c ed e sono eventi attuali, le proposizioni $O(c)$ e $O(e)$ saranno vere e quindi il primo dei due controfattuali sopra è banalmente vero⁸; ma se il primo controfattuale è banalmente vero, segue che *nel caso c ed e siano eventi attuali, la relazione di dipendenza causale dipenderà solo dalla verità del secondo controfattuale*; quindi:

se c ed e sono eventi attuali, allora e dipende causalmente da c se e solo se è vero che se c non si fosse verificato, allora e non si sarebbe verificato.

Ma questa caratterizzazione delle dipendenza causale non è altro che la caratterizzazione della causalità data da Hume sopra. Lewis considera la definizione di Hume, quindi, come una buona definizione di dipendenza causale ma non ancora come una buona definizione di causalità.

Terzo passo: Causalità Per arrivare alla nozione di causalità da quella di dipendenza causale è necessario un altro (piccolo) passaggio.

La tesi di Lewis è che la dipendenza causale implica la causalità, ossia che tutte le volte che due eventi sono nella relazione di dipendenza causale sono anche nella relazione di causalità. Il problema è che non tutte le relazioni di causalità implicano quella di dipendenza causale (in altri termini: la dipendenza causale è *sufficiente* per la causazione ma *non è necessaria*).

In particolare si daranno casi di causalità che non sono casi di dipendenza causale. Vediamo un esempio:

Riconsideriamo l'esempio della casa: uno degli effetti dell'incendio è, magari, che il proprietario viene rimborsato dall'assicurazione. Visto che l'incendio della casa è stato causato da un corto circuito, diremmo anche che il corto circuito è la causa del fatto che il proprietario si è preso i soldi dell'assicurazione.

Questo esempio mostra che se e causa e' e c causa e , saremmo disposti a concludere che c causa e' ossia che *la causalità gode della proprietà transitiva*.

Il problema è che *la relazione di dipendenza causale non gode della proprietà transitiva*: il pagamento dell'assicurazione dipende causalmente dall'incendio (poichè se non ci fosse stato l'incendio, non ci sarebbe stato il pagamento dell'assicurazione) e l'incendio dipende causalmente dal corto circuito (se non ci

⁸Perchè un controfattuale della forma "se fosse A , sarebbe B " è vero nel caso l'antecedente ed il conseguente siano veri? Proviamo ad usare la definizione data sopra: perchè un controfattuale sia vero deve esistere un mondo possibile in cui A e B sono veri più simile al mondo attuale di un mondo possibile in cui A è vero e B è falso; ma visto che A e B sono di fatto veri, il mondo possibile che cerchiamo è proprio il mondo attuale; poichè non c'è alcun mondo più simile al mondo attuale del mondo attuale stesso, il nostro condizionale controfattuale è vero.

fosse stato il corto circuito non ci sarebbe stato l'incendio) ma il pagamento dell'assicurazione *non* dipende causalmente dal corto circuito poichè l'assicurazione avrebbe pagato anche se l'incendio si fosse sviluppato per altri motivi (compatibili con il contratto).

La soluzione di Lewis consiste nell'*estendere la relazione di dipendenza causale in modo da renderla transitiva* e questo per mezzo della nozione di *catena causale*.

Una *catena causale* è una sequenza di eventi c, d, e, \dots tali che d dipende causalmente da c , e da d , *etc.*

La causalità può quindi essere definita nei termini di catena causale (e quindi indirettamente nei termini della dipendenza causale):

L'evento c **causa** e se e solo se esiste una catena causale da c ad e .

In questo modo, la relazione di causalità è solo tra eventi legati da relazioni di dipendenza causale. Nel caso dell'incendio possiamo ora dire che il corto circuito ha causato il pagamento dell'assicurazione solo sulla base del fatto che tra l'evento del corto circuito e l'evento del pagamento vi sono eventi legati da dipendenza causale.

Il ruolo della nozione di catena causale è quello di rendere più esplicativa la teoria: perchè c causi e non è necessario infatti che ci sia un legame "diretto" di dipendenza controfattuale tra c ed e ma basta che ci sia almeno un evento intermedio d tale che esso dipenda controfattualmente da c ed e dipenda controfattualmente da d .

3.3 Casi di causalità a vuoto

I casi di causalità a vuoto sono casi tradizionalmente presentati come controesempi alle teorie controfattuali della causalità. Presenterò tre casi e vedremo come la teoria di Lewis sia stata modificata al fine di risolverli.

Causazione a vuoto anticipata

Esempio dell'attentato: Due attentatori sono appostati sul tetto di un edificio in attesa del passaggio del corteo presidenziale. Il loro scopo è sparare al presidente. I due attentatori si appostano uno vicino all'altro in un punto che considerano ottimo per sparare. Al passaggio del presidente, il primo attentatore preme il grilletto e spara un colpo che va a segno mentre il secondo attentatore, al momento in cui si rende conto che il suo collega ha premuto il grilletto, desiste dallo sparare anche lui un colpo.

Questo esempio è chiamato causalità a vuoto anticipata ("early pre-emption") poichè c' è un certo evento possibile e non realizzato (lo sparo del secondo attentatore) che se si fosse verificato avrebbe raggiunto un certo effetto e (la morte del presidente) se un'altro evento attuale c (l'attuale sparo del primo attentatore), che di fatto causa e , non lo avesse impedito.

Questo esempio è considerato un controesempio alla teoria controfattuale della causalità perchè saremmo disposti a dire che lo sparo del primo attentatore causa la morte del presidente ma è falso che se il primo attentatore non avesse sparato, il presidente sarebbe morto (in quel caso avrebbe sparato il secondo attentatore raggiungendo il suo scopo).

La teoria di Lewis riesce ad evitare questo contro-esempio, senza modifiche, per mezzo della nozione di *catena causale*.

Sembra esserci, infatti, una catena causale che lega il premere del grilletto del primo attentatore all'evento della morte del presidente. Si consideri, ad esempio, il proiettile a metà della sua traiettoria. Il proiettile a metà della sua traiettoria dipende dal fatto che il primo attentatore ha premuto il grilletto e la morte del presidente dipende dal fatto che il proiettile abbia avuto proprio quella traiettoria. C'è quindi una catena causale tra i tre eventi e quindi, per la definizione di causalità, data sopra è il fatto che il primo attentatore ha premuto il grilletto a causare la morte del dittatore. Non è possibile, invece, costruire alcuna catena causale che colleghi le azioni del secondo attentatore alla morte del presidente.

La nozione di catena causale serve a Lewis sia a rendere la nozione di dipendenza causale compatibile con la transitività della causalità sia a spiegare casi di causalità a vuoto anticipata.

Causalità a vuoto posticipata I casi di causalità posticipata (“late pre-emption”) sono invece un problema anche per la teoria di Lewis (almeno nella sua versione originale). Vediamone uno:

Esempio dei sassi: Mario e Luisa stanno tirando dei sassi contro una bottiglia di vetro. Luisa tira per prima (o più forte) e colpisce la bottiglia. Mario tira un attimo dopo (o più debolmente) e il suo sasso, quando raggiunge il luogo dov'era la bottiglia, passa sopra i pezzi di vetro rimanenti.

Questo è un caso di causalità a vuoto posticipata poichè, quando l'evento c causa di fatto e (ossia quando il tiro di Luisa colpisce la bottiglia) la causa mandata a vuoto (il tiro di Mario) è ancora in corso e si completa (passa sopra quello che rimane della bottiglia) solo *dopo* che c è giunto a termine.

Anche in questo caso ci troviamo di fronte ad un controesempio alla teoria controfattuale poichè diremmo che il tiro di Luisa è la causa dell'infrangersi della bottiglia ma se Luisa non avesse tirato, la bottiglia sarebbe stata colpita dal tiro di Mario.

Questa volta, però, l'appello a catene causali non può essere d'aiuto a Lewis.

Si consideri il sasso di Luisa a metà della sua traiettoria; sicuramente questo evento dipende dal tiro di Luisa ma l'infrangersi della bottiglia non dipende dall'essere il sasso di Luisa a metà della sua traiettoria poichè anche senza questo evento intermedio la bottiglia sarebbe stata colpita comunque.

Un modo per risolvere questo tipo di casi consiste nello sfruttare il fatto che *se fosse stato il sasso di Mario ad infrangere la bottiglia, tale evento sarebbe accaduto in un istante successivo a quello in cui di fatto è accaduto.*

Mentre il sasso di Luisa colpiva la bottiglia, il sasso di Mario era ancora in volo. L'effetto di questa interazione causale (l'infrangersi della bottiglia) è già avvenuto quando il sasso di Mario passa sopra il luogo dove la bottiglia era collocata. Ma un evento c non può essere la causa di un evento e se c si conclude *dopo* e .

Il lancio di Mario, quindi, non sarebbe stato la causa dell'evento che di fatto è accaduto (e che è essenzialmente individuato dal tempo del suo verificarsi) ma di un altro evento che differisce dal primo, tra le altre cose, per il tempo della sua realizzazione.

La teoria di Lewis può essere salvata da casi di causalità a vuoto posticipata se si assume una *tesi generale sulla natura degli eventi* ossia che essi siano essenzialmente individuati dal tempo della loro occorrenza (in letteratura, questa caratteristica degli eventi di essere essenzialmente determinati, oltrechè dal tempo del loro verificarsi anche dal tempo e/o altre proprietà è chiamata "fragilità")

Causalità a vuoto sconfitta C'è però un ennesimo caso di causalità a vuoto che potrebbe essere chiamata (in attesa di una traduzione migliore) "causalità a vuoto sconfitta" ("trumping pre-emption"). Questo tipo di casi si trova in un recente articolo di J. Schaffer (vedi [10]).

Esempio dei soldati: Nell'esercito vige la regola che vanno eseguiti gli ordini del soldato con il grado maggiore. Durante le operazioni di guerra, un sergente ed un maggiore entrambi ordinano simultaneamente ad un caporale "spara!". Il caporale esegue l'ordine obbedendo al maggiore

Questo caso è un contro-esempio alla teoria controfattuale perchè saremmo disposti a dire che è l'ordine del maggiore (il più alto in grado) la causa dello sparo del caporale. Tuttavia se anche il maggiore non avesse ordinato di sparare, lo avrebbe fatto il sergente a cui il caporale avrebbe comunque obbedito.

In questo caso le strategie precedenti non servono:

- non c'è alcun evento intermedio che possa distinguere il processo causale che di fatto si realizza (quello tra l'ordine del maggiore e lo sparo del caporale) dal processo che viene inibito da quest'ultimo (quello dall'ordine del sergente allo sparo del caporale). Nei casi di questo tipo, infatti, o l'evento intermedio semplicemente non c'è (perchè il processo causale è diretto: che evento intermedio c'è tra l'ordine e l'esecuzione?) o semplicemente perchè il processo causale inibito si realizza completamente e si sovrappone perfettamente a quello rilevante.
- i due processi causali (almeno nel caso dei soldati) sono perfettamente sovrapposti: se anche fosse risultato determinante il processo che di fatto viene inibito (l'ordine del sergente), l'effetto (lo sparo del caporale) avrebbe avuto le medesime proprietà temporali, spaziali *etc.*

La soluzione a questo tipo di casi è stata proposta da Lewis in un recente articolo (vedi [7]).

L'idea è (grosso modo) questa:

per distinguere da un punto di vista controfattuale il processo causale reale (dal maggiore al caporale) da quello andato a vuoto (dal sergente al caporale) si potrebbero considerare gli effetti di *possibili alterazioni* all'uno o all'altro. In una situazione in cui l'ordine del maggiore fosse stato un po' differente (se, per esempio, avesse ordinato solo di caricare il colpo o di battere in ritirata) mentre quello del sergente fosse rimasto identico, le azioni del caporale sarebbero state diverse (ossia il caporale avrebbe solo caricato il colpo o avrebbe battuto la ritirata anche se il sergente gli ordinava di fare diversamente). Non è vero invece il contrario: in una situazione in cui gli ordini del maggiore fossero rimasti identici ma non quelli del caporale, il processo causale sarebbe rimasto inalterato (il caporale avrebbe eseguito comunque gli ordini del più alto in grado indipendentemente da quello che gli ordinava il sergente).

Proprio sulla base di questi casi, Lewis decide di riformulare la sua teoria in termini di *influenza*:

Siano c ed e due eventi distinti; c influenza e se e solo se esiste un *range* di possibili alternative di c , c_1, \dots, c_n e possibili alternative di e e_1, \dots, e_n tali che sono veri i seguenti controfattuali:

- se fosse c_1 , sarebbe e_1
- se fosse c_2 , sarebbe e_2
- ...
- se fosse c_n , sarebbe e_n

Sulla base della nozione di influenza, Lewis ridefinisce la causalità in questo modo:

c causa e se e solo se c'è una catena di influenze tra c ed e .

Da notare che questa strategia è sufficiente anche per risolvere i casi di causalità a vuoto posticipata *senza* assumere la fragilità degli eventi, ossia la tesi che un evento sia essenzialmente individuato dal luogo ed il tempo (o altri fattori) del suo verificarsi. Nel caso del lancio di Luisa, le possibili alternative avrebbero avuto delle conseguenze sull'infrangersi della bottiglia, mentre ciò non è vero delle possibili alternative del lancio di Mario.

4 Letture per l'esame

Per sostenere l'esame, studiate approfonditamente queste pagine e leggete i seguenti testi:

- D. Hume, *Ricerca sull'Intelletto Umano*, sezione IV e VII
- J. Mackie, "Causes and Conditions" in [11], da pagina 33 a pagina 39 (la prima sezione)
- D. Lewis, "Causation" in [11], da pagina 191 a pagina 200 (fino al capitolo intitolato "causation" compreso)

Riferimenti bibliografici

- [1] Jerry Fodor. *Concepts: where cognitive science went wrong*. Oxford University Press, 1989.
- [2] David Hume. *A Treatise of Human Nature*. 1739-40.
- [3] David Hume. *An Enquiry Concerning Human Understanding*. A. Millar, London, 1748. trad. it. Laterza, Roma-Bari, 1996.
- [4] Jaegwon Kim. Causes and events: Mackie on causation. *Journal of Philosophy*, 1965. Reprinted in [11, pp. 60-74],.
- [5] David K. Lewis. Causation. *Journal of Philosophy*, 70:556-567, 1973. Reprinted in [11, pp. 193-204].
- [6] David K Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
- [7] David K. Lewis. Causation as influence. *Journal of Philosophy*, 97:182-197, 2000.
- [8] John Locke. *An Essay Concerning Human Understanding*. 1690.
- [9] John L. Mackie. Causes as conditions. *American Philosophical Quarterly*, 2(4):245-264, 1965. Reprinted in [11, pp. 33-55].
- [10] Jonathan Schaffer. Trumping pre-emption. *Journal of Philosophy*, 97(4):165-181, 2000.
- [11] Michael Sosa, Ernest Tooley, editor. *Causation*. Oxford University Press, 1993.

5 Domande

Riporto qui sotto una serie di “domande-tipo” che potrebbero essere fatte durante l’esame; dovrebbero esservi utili per capire cosa mi interessa che voi sappiate.

Tenete presente che quelle qui riportate hanno solo un valore indicativo. Potrebbe succedere che non vengano fatte proprio queste domande o che ne vengano fatte di diverse.

Se avete dubbi o perplessità, contattatemi pure all’indirizzo e-mail riportato sopra.

Su Hume:

- Quali sono le caratteristiche principali dell’epistemologia di Hume (che lui chiama Teoria delle Idee)?
- perchè, per Hume, ci possono essere idee senza impressioni o impressioni senza idee? Puoi fare un esempio (diverso da quello riportato negli appunti)?
- Perchè, per Hume, tutti i ragionamenti su questioni di fatto sono basati sulla relazione causa-effetto?
- Cosa si intende per relazione filosofica e relazione naturale? Come si applica questa distinzione alla causalità?
- Alla luce della sua teoria delle idee, come si presenta a Hume il problema della causalità?
- Perchè Hume ritiene che la causalità non sia un’idea semplice?
- Cosa vuol dire che l’approccio di Hume alla causalità è riduzionista?
- Quali sono per Hume i tre aspetti che caratterizzano la relazione causa-effetto?
- Perchè i primi due non sono sufficienti a distinguere nessi causali da coincidenze (fare un esempio, possibilmente diverso da quello degli appunti)?
- Qual’è il terzo aspetto? Perchè per mezzo di esso riusciamo a distinguere nessi causali da coincidenze?
- Qual’è la relazione tra l’idea di connessione necessaria e la congiunzione costante di due eventi?
- Perchè la teoria di Hume è una teoria regolarista della causalità?
- Perchè l’appello alle regolarità non è sufficiente a distinguere nessi causali da generalizzazioni (fare degli esempi)?

Su Mackie:

- Che cosa vuol dire che un evento è la condizione necessaria per un altro evento?
- Che cosa vuol dire che un evento è la condizione sufficiente per un altro evento?
- Perché, in molti casi, la causa di un evento non è la sua condizione necessaria e sufficiente (fare esempio)?
- Che cos'è una condizione INUS?
- Qual'è la “forma logica” di una regolarità per Mackie?
- Perché la causa di un evento è *almeno* una condizione INUS?
- Perché il problema degli effetti congiunti di una causa comune è un problema per l'approccio di Mackie?
- In termini generali, perché, per Kim, la teoria di Mackie ha un problema categoriale?

Su Lewis:

- Qual'è l'idea che sta alla base della teoria controfattuale della causalità?
- Che cos'è un condizionale controfattuale (anche facendo esempi)?
- Qual'è l'idea su cui si basa il trattamento di Lewis dei controfattuali (anche con un esempio)?
- Che cos'è un mondo possibile?
- Quali sono i due requisiti che deve soddisfare la relazione di somiglianza tra mondi possibili?
- Qual'è la caratterizzazione “rigorosa” (ossia in termini di mondi possibili) delle condizioni di verità dei controfattuali data da Lewis?
- Perché l'enunciato “Se il presidente avesse premuto il bottone rosso, sarebbe scoppiata una guerra nucleare” è un controesempio alla teoria di Lewis? Come risolve Lewis il problema?
- Che cos'è la relazione di dipendenza controfattuale (anche facendo un esempio)?
- Cosa vuol dire che la nozione intuitiva di causa gode della proprietà transitiva (fare un esempio)?
- Perché la relazione di dipendenza controfattuale non gode della proprietà transitiva (fare un esempio)?
- Perché l'esempio degli attentatori è un controesempio alla teoria controfattuale della causalità? Perché è un caso di causalità a vuoto anticipata?

- Perché l'esempio dei sassi è un controesempio alla teoria controfattuale della causalità? Perché è un caso di causalità a vuoto posticipata?