Causal Explanation and Causal Models

Antony Eagle

Choices, Models and Morals » Lecture 5

Contents

Causation and Causal Explanation

Causes and Reasons

- We've now reviewed two models of rational choice: individual decision theory, and game theory.
- Abstracting from the details, they are both models of how **reasons cause individual actions**, fitting with Davidson's idea that rationalizing explanations are a species of causal explanation.
  - In decision theory, it is the fact that a given action maximises expected utility that is a reason to perform it; in game theory, that a given action is part of a Nash equilibrium is a reason to perform it. So when those actions are performed, those reasons can be cited as (partial) causes.
- In both cases, the reasons need not be **sufficient**: if there are two utility maximising acts, or two Nash equilibria, then we need some further reasons to explain why this action rather than another is performed.
- But note that these reasons are still causes that contribute to the resulting action, even if they aren't sufficient causes of it.

Selecting Causes

- But we haven't said **how** reasons cause actions.
  - Davidson says only 'Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason' (1963: 691). He doesn't say anything about what makes a reason and an act stand in this 'because' relationship.
  - To put it another way, given an action, what is the basis for its being explained by one reason the agent has rather than another?
    - We may suppose that the agent had both reasons, and was aware of both, prior to the act; that both reasons would rationalize the act; that both reasons are psychologically plausible motivators – still, it may be that only one was effective, the other standing idly by.
    - E.g., suppose the agent is aware of both game-theoretic and decision-theoretic treatments of the stag hunt; they hunt stag because it is the risk dominant equilibrium, rather than because it maximises expected utility given their pessimism about their fellow hunters. Both are rationalizing explanations, but only one is the 'real' explanation.

Causation and Effective Strategies

- Distinguishing between the real cause, and merely potential causes, in the case of rationalizing explanations is an example of a broader phenomenon.
- Consider **life insurance**. It is, we may suppose, a fact that individuals who purchase life insurance live longer than those who do not (Cartwright 1979: 420).
- So if you want to live longer, should you purchase life insurance?
- Arguably not: that is merely an **association**, and cannot be exploited as an **effective means** of generating the desired conclusion.
  - Buying life insurance is probably **evidence** of the kind of person you are – prudent and sober. It is a symptom of factors that conduce to long life, not a factor that conduces to long life itself.
  - Think back to the Newcomb problem: buying life insurance is merely **acting like** someone who will have a long life. Unless it is accompanied by the other things of which a propensity to purchase life insurance is a typical sign, it will be ineffective.
- Cartwright argues: the distinctiom between effective and ineffective strategies rests on causal relations, not mere associations (Cartwright 1979: 429–33).

Effective Strategies and Interventionist Policy

- There are a lot of statistics in social science. Economics, political science, sociology – all collect vast arrays of data about people and their activities.
- But the design of good policy aims to **intervene** on the systems that produce this statistical data, to makes changes that promote improvements in the data.
  - Take the minimum wage case (Reiss 2013: 87): we have lots of data on minimum wage levels across societies, and on unemployment rates. Can we discern whether an increase in minimum wages will result in higher unemployment? That will depend on the causal structure of the situation: is unemployment a variable that depends on wages, or is an observed association between them the result of other intermediate variables?
- The flipside of designing effective interventions is **understanding existing associations**. If intervening on wages doesn't causally promote unemployment, then the current unemployment rate is typically not explained by the current wage levels.

Revisiting D-N

- We discussed in lecture 2 the flagpole example: there is a regular association between the position of the sun, the height of a certain flagpole, and the length of a shadow cast.
- But though the height of the flagpole and the length of the shadow allow us to **predict with certainty** the position of the sun, they do not **explain** the position of the sun, because they are not among its causes.
  - Reiss claims that 'the core of the problem seems to be that explanation is an asymmetric relation whereas deduction is symmetric' (2013: 88). This can't be quite right: consider **paresis**, a form of dementia that is an uncommon result of untreated syphilis. Paresis entails untreated syphillis, but not *vice versa*; yet it is the presence of untreated syphilis which causes, and thus explains, paresis. In this case deduction is asymmetric too, but it holds in the 'wrong way'.

Associations

- We've talked very generally of 'associations'. The standard measure of association is **correlation**, which measures how dependent two variables are on one another – how strongly one can **predict** the value of one variable, given the value of another.
- Different **measures** of correlation depend on different hypotheses about the structure of the dependence.
  - E.g., the Pearson correlation coefficient (Reiss 2013: 89) measures the extent to which the association between variables can be captured by a linear relationship $Y = aX + b$; if $Y = X^2$, for example, then Y is completely dependent on X but their correlation coefficient will be 0 on this measure.
  - But if $\Pr(X \mid Y) = \Pr(X)$, i.e., X and Y are independent, then they will be measured to be uncorrelated by any measure using the real probabilities.
- Association in this sense is symmetric, as is probabilistic dependence – again, unlike causation.
- So when X causes Y, and the structure of that relationship is captured by a linear equation, there will be a **non-causal linear association** between Y and X.

Producing Patterns of Association

- An association needn't indicate a causal relationship at all. There may be an association between X and Y because both are effects of a **common cause** Z.
  - Recall the smoking gene case from Lecture 3, where a gene predisposes its bearers to lung cancer and to smoking to relieve the symptoms.
- An association might be **constitutive** – my location, and that of my hand, are tightly associated but not causal, since they are not wholly **distinct** events.
- An association might be merely accidental, or **spurious**.

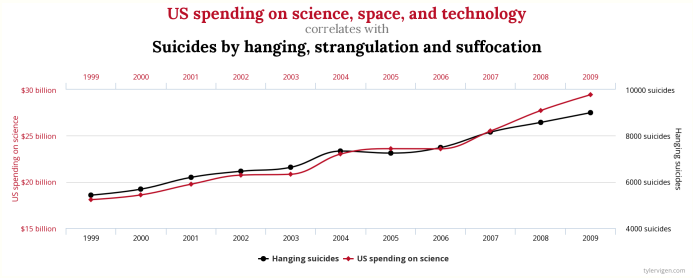

Figure 1: A spurious correlation, from https://www.tylervigen.com/spurious-correlations.

Theories of Causation

Humean Regularity Theory

> Here is a billiard-ball lying on the table, and another ball moving towards it with rapidity. They strike; and the ball, which was formerly at rest, now acquires a motion. This is as perfect an instance of the relation of cause and effect as any which we know, either by sensation or reflection. Let us therefore examine it. 'Tis evident, that the two balls touched one another before the motion was communicated, and that there was no interval betwixt the shock and the motion. *Contiguity* in time and place is therefore a requisite circumstance to the operation of all causes. 'Tis evident likewise, that the motion, which was the cause, is prior to the motion, which was the effect. *Priority* in time is therefore another requisite circumstance in every cause. But this is not all. Let us try any other balls of the same kind in a like situation, and we shall always find, that the impulse of the one produces motion in the other. Here therefore is a third circumstance, viz. that of a *constant conjunction* betwixt the cause and effect. Every object like the cause, produces always some object like the effect. Beyond these three circumstances of contiguity, priority, and constant conjunction, I can discover nothing in this cause. (Hume 1740: ¶9)

Hume's Later Habituation Theory

- Hume retreats from this in his typical way, identifies the concept of causation with the empirical experiences which generate the idea of causation in us:

> It appears, then, that this idea of a necessary connexion among events arises from a number of similar instances which occur, of the constant conjunction of these events…. after a repetition of similar instances, the mind is carried by habit, upon the appearance of one event, to expect its usual attendant, and to believe, that it will exist. …
> Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be *an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or in other words, *where, if the first object had not been, the second never had existed.* (Hume 1777: ¶¶7.28–29)

Associations and Regularities; Causes without Regularities

- The problems for reading causal relations of associations and correlations strike Hume's regularity theory, and even his revised theory appeals to the same constancy of regular association.
- We've seen already examples of associations and patterns which might habituate our minds to certain inferences, but which are non-causal.
  - Actually perhaps Hume is in better standing here than he thought – we don't seem inclined to be habituated by just any association, so perhaps the mind is more discriminating in its ability to learn causal relations from correlational data.
- But there are also maby causal relations which don't produce associations, or not of the right sort.
  - Consider cases where X promotes Y, but inhibits Z, which is an even more potent cause of Y. We might see no association between X and Y, or even an negative association between X and Y, even though X causes Y. (This is Hesslow (1976)'s contraceptive example.)

The Counterfactual Theory

› Hume's 'other words' have prompted **counterfactual** theories of causation:

> Let c and e be two distinct possible particular events. Then *e depends causally* on c iff the family $O(e), \neg O(e)$ depends counterfactually on the family $O(c), \neg O(c)$. As we say it: whether e occurs or not depends on whether c occurs or not. The dependence consists in the truth of two counterfactuals: $O(c)\Box \rightarrow O(e)$ and $\neg O(c)\Box \rightarrow \neg O(e)$. … if c and e are actual events, then it is the first counterfactual that is automatically true. Then e depends causally on c iff, if c had not been, e never had existed. I take Hume's second definition as my definition not of causation itself, but of causal dependence among actual events. …
>
> Causal dependence among actual events implies causation. If c and e are two actual events such that e would not have occurred without c, then c is a cause of e. But I reject the converse. Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence. … We extend causal dependence to a transitive relation in the usual way. … one event is a cause of another if there exists a causal chain leading from the first to the second. (Lewis 1973: 563)

Causal Tendencies

- Mill's theory is in line with Lewis', at least in Reiss (2013)'s presentation.
- Causes contribute to effects against a background held fixed for the evaluation of a counterfactual.

Explanation and Models

Events and Phenomena

- Every physical **event** has some causal antecedents, according to causal realists.
- But as we noted in Lecture 2, not all events are **phenomena**.
- In Hacking's words, a phenomenon is 'an event or process of a certain type that occurs regularly under definite circumstances' (Hacking 1983: 221; see also Reiss 2013: 17–19).
  - A one-off unique event won't be a phenomenon amenable to general scientific theorizing, even if it can be explained by citing its prior causes, and even if those causes are themselves of a familiar type.
- A phenomenon is thus **abstracted** from the particularities of the various events of a given type.
  - So *financial crises* are a phenomenon; but the GFC is not, because it had many distinctive featues it doesn't share with other crises. To explain the phenomenon is to explain the general features all financial crises share; to explain the GFC is to explain the what makes it distinctive of its kind.

Models and Theories

- Scientists use **models** of many sorts, in many ways. But the key use of models in economics is **representational**:

> Many scientific models are representational models: they represent a selected part or aspect of the world, which is the model's target system. Standard examples are the billiard ball model of a gas, the Bohr model of the atom, the Lotka–Volterra model of predator–prey interaction, the Mundell–Fleming model of an open economy, and the scale model of a bridge. (Frigg and Hartmann 2020: §1)

- The line between 'model' and 'theory' is very thin here. A theory might be no more than a collection of models (Frigg and Hartmann 2020: §4.1) – e.g., a theory of celestial mezchanics might include a model of the actual solar system along with models of various merely possible solar systems, all obeying the same mechanical laws.
  - A model is an abstract representation of a particular **instantiation** of a theory, making the laws of the theory true in a specific case. A theory is a collection of all models that make the same general laws true.

Models and explanation

- **Theories** certainly seem to explain: a true theory encompasses all the causes of an event; providing a theory that encompasses an event is to provide information about its causal history, describing **how** it could be caused.
  - Moreover, showing that a certain event is in the scope of a true theory, and that the theory predicts its occurence, thereby renders the event less surprising – contributing to a sense of **understanding** that is so often linked to explanation.
  - showing that a certain event is in the scope of a true theory, and that the theory predicts its occurence, thereby renders the event less surprising.
- One natural view of how models might explain is by being linked to theories, as on Cartwright's 'simulacrum account of explanation' (1983: ch. 8):

> she suggests that we explain a phenomenon by constructing a model that fits the phenomenon into the basic framework of a grand theory. On this account, the model itself is the explanation we seek. (Frigg and Hartmann 2020: §3.3)

- On this view, subsuming an event into a causal model **is** explaining it; it is showing how to represent the phenomenon in such a way that a theory applies to it.
- Since a model includes the relevant prior causes, it concretely establishes a way the event could be brought about, in line with the theory.

Mathematical Models

- A paradigm example is the **Hardy-Weinberg model** in evolutionary genetics (Hardy 1908), which is an idealized mathematical representation of the distribution of traits in a sexually reproducing population.
- Assume that there are two alleles a and A determining three genotypes, AA, aa, and Aa. Initial allele frequencies are $f_0(A)$ and $f_0(a) = 1 - f_0(A)$. Mating is random, each parent contributes one allele to its offspring independent of sex, the population is infinite and closed, and generations do not overlap. Then the model entails that
  1. The distribution of genotypes in the subsequent generation is $f_1(AA) = f_0(A)^2$; $f_1(Aa) = 2f_0(A)f_0(a)$; $f_1(aa) = f_0(a)^2$.
  2. The distribution remains constant (in 'Hardy-Weinberg equilibrium') for all subsequent generations.
- This 'base' model isn't very interesting, because the conditions constraining the model are clearly not instantiated.
- Illumination arises when we **liberalise** those conditions: e.g., what if mating is selective rather than random? What if mutations introduce new alleles or change the frequency of existing alleles? What if the population is finite, so that genetic drift plays a role? Removing or weakening these conditions and running **simulations** of the population over time can yield representations of actual populations.

‘All models are wrong’ (Box 1976: 792)

- If the H-W model is exemplary, then models may be **inaccurate** in many ways – ‘representation’ may be somewhat impressionistic (Wimsatt 2007: 101–2).
  - Models might be **idealized** – involving such things as infinite populations or completely random mating – but approximated in various real cases (large populations, unselective mating, e.g., coral spawning).
  - Models might be **incomplete** – if mating isn’t random, it has causal structure that is omitted from the H-W model – e.g., there is no **spatial structure** to the population, yet the location of individuals is surely relevant to the probability of reproduction.
  - Models might, by imcompleteness, misdescribe the relations between their mathematical components; e.g., omitting a common cause might lead to a spurious association. In the H-W case, including selective mating but neglecting spatial structure, and noting the increase over time in *AA* and relative to *aa* may lead to the hypothesis that *A* has a selective advantage, where the real answer might be the spatial clustering of *AA*s in a particular region.
  - A model might be totally wrong-headed – as, arguably, the H-W model is. This is just not at all how real populations evolve.

…but some are useful

> Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.…
> Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad. (Box 1976: 792)

- Box here argues that the 'wrongness' of a model isn't an obstacle: in fact, by focusing out attention on the core drivers of a phenomenon, a model can be wrong but right in everything important, and – crucially – be economical when compared to a complete depiction.
  - Recall again Lewis' idea that causal explanation is giving information about causal history - not all the information, necessarily, but the important information.

Models in Economics

- Reiss-2013 [pp. 121–123] gives the example of Hotelling's model of how geography influences vendors' pricing and/or location.
- The model assumes two vendors who can locate themselves where they want in a uniformly distributed population of consumers located along a single spatial dimension. Consumers care about price and transportation costs only; they buy from their closest vendor when prices are equal, buying from a more distant vendor only when a lower price offsets the increased transport costs, so the quantity of sales for vendor X is the proportion of the line where the total costs for consumers are lower when purchasing from X.
- The model entails that, when vendors are allowed to move, they will come to an equilibrium with both immediately adjacent to each other in the middle of the population.
  - This is robust; any move by a vendor A away from that location will prompt another move by the competitor B that cuts into A's profits. This is so even though consumers would be better off, having less travel costs on average, if vendors divided the line into equal regions and occupied the midpoints of those regions (Reiss 2013: 123)

Inaccuracy in Hotelling's model

- The model again involves inaccuracy (Reiss 2013: 124–26):
  - Idealization – geography is assumed to be one-dimensional.
  - Incompleteness – consumers care only about unit price and distance, whereas there are surely many more factors influencing real consumption.
  - Misdecribed interactions: neglecting other dimensions of consumer choice means that false causal relations are imputed, particularly considering non-equilibrium starting points – it seems unlikely that *A*'s being out of an equilibrium location will **cause** *B* to move to their immediate right. Vendors will, other things being equal, prefer to change their product offering in other ways than its location to preserve profit.
  - 'A wrong picture of reality' – the explanation of why vendors are nearby cites profit maximization - but what if the 'real' explanation is that land is cheap there? Or vendor *B* noted *A*'s success and simply wanted to copy it? The causal picture offered by the model would be wholly spurious.
- But the model is still taken as explanatory – we do see clustering of vendors, in physical space and in product space, and the model allows us to make sense of this.

Causal Models and Econometrics

- Hotelling's model is driven by *a priori* assumptions about rational consumers and vendors; in this sense, it is a model that implements an assumption of rational choice over locations and prices.
- Other economic models are more data-driven – e.g., econometric models that derive from observed correlations between economic statistics (Reiss 2013: ch. 9). As Hoover puts it, there is a divide

> between those who believe that economic logic itself gives privileged insight into economic behaviour (a priori approaches) and those who believe that we must learn about economic behaviour principally through observation and induction (the inferential approaches). (Hoover 2008: 11)

- The promise of the econometric program is the idea that we can derive and construct causal models simply by attending to **patterns in the statistical data** (Hoover 2008: 10).
  - Though there remain a priori assumptions, e.g., that observed independencies reflect causal independencies – although these may be weaker than the rationality assumptions that go into Hotelling's model.
- Regression analysis; Exogenous and endogenous variables. (Reiss 2013: 166)

False Models and the Paradox of Explanation (Reiss 2013: 127–41)

- Note that false models can be part of true theories.
- Contrastive explanation: the model better captures the phenomenon than a rival.

Idealized models in policy

https://philosophyofbrains.com/2023/07/11/cognitive-science-of-philosophy-symposium-idealized-models.aspx

References

References

Box, George E P (1976) 'Science and Statistics', *Journal of the American Statistical Association* **71**: 791–99. doi:10.1080/01621459.1976.10480949.

Cartwright, Nancy (1979) 'Causal Laws and Effective Strategies', *Noûs* **13**: 419–37. doi:10.2307/2215337.

Cartwright, Nancy (1983) *How the Laws of Physics Lie*. Clarendon Press.

Davidson, Donald (1963) 'Actions, Reasons, and Causes', *The Journal of Philosophy* **60**: 685–700. doi:10.2307/2023177.

Frigg, Roman and Stephan Hartmann (2020) 'Models in Science', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Hacking, Ian (1983) *Representing and Intervening*. Cambridge University Press.

Hardy, G H (1908) 'Mendelian Proportions in a Mixed Population', *Science* **28**: 49–50. doi:10.1126/science.28.706.49.

Hesslow, Germund (1976) 'Two Notes on the Probabilistic Approach to Causality', *Philosophy of Science* **43**: 290–92.

References (cont.)

Hoover, Kevin D (2008) 'Causality in Economics and Econometrics', in Steven N Durlauf and Lawrence E Blume, eds., *The New Palgrave Dictionary of Economics*: 1–13. Palgrave Macmillan UK.

Hume, David (1740/2022) *An Abstract of a Treatise of Human Nature*, Peter Millican and Amyas Merivale, eds. https://davidhume.org/texts/a/.

Hume, David (1777/2022) *An Enquiry Concerning Human Understanding*, Amyas Merivale and Peter Millican, eds. https://davidhume.org/texts/e/.

Lewis, David (1973) 'Causation', *Journal of Philosophy* **70**: 556–67. doi:10.2307/2025310.

Reiss, Julian (2013) *Philosophy of Economics*. Routledge.

Wimsatt, William (2007) *Re-Engineering Philosophy for Limited Beings*. Harvard University Press.