When correlation warrants a causal inference

An exploratory study of population effects and particular events

Suppose two devices can start the same pump. The primary starter succeeds nine times in ten. Selecting the auxiliary starter disconnects the primary and succeeds only once in ten. One morning an operator selects the auxiliary starter, the pump starts, and the electrical record shows current reaching the motor through the auxiliary circuit. What should the investigator conclude? An engineer who recommends leaving the auxiliary switch alone and an investigator who credits it with this morning's start can both be right. They are answering different questions about the same machine. The difficulty lies in explaining precisely how far their answers can come apart.

This example adapts a problem at the center of Phil Dowe and Paul Noordhof's edited volume Cause and Chance (2004). It supplies a demanding test for the familiar injunction about correlation and causation. Here the probabilities could have been established by ideal experiments. We could know which direction the causal influence runs, how large the effect is, and exactly which alternatives were available. Even then a question about the history of the individual start would remain. Conversely, following the current on one occasion gives little basis for putting numbers on the reliability of the two devices. The limits of inference run in both directions.

The working conclusion of this study is that a correlation warrants a causal claim to the extent that the circumstances that produced the evidence enable it to discriminate among histories or models that disagree about that claim. Here is the test I propose. Change the available backup while keeping the actual sequence fixed. Then change the actual sequence while preserving the population rates. If either alteration reverses the proposed conclusion without disturbing the evidence, ask what independent observation could detect it. The alternatives that matter for estimating the effect of a policy differ from those that matter for explaining a particular outcome. What one

knows about assignment can settle the first comparison. What one knows about an interrupted circuit can settle part of the second. Sometimes both bodies of evidence together still leave open what would have happened to this particular unit under a different assignment. Those remaining possibilities are scientifically meaningful, and in some settings one can put numbers on them.

Keep two people in view as we go. The maintenance supervisor can test a population of machines and must decide which starter to use. The investigator has the record of one start and must decide what it establishes about Tuesday. The same evidence will repeatedly change hands between them. Each time, ask what the second person still needs to learn before accepting the first person's conclusion.

I. What a correlation can give us

An observer might mean at least four things by saying that a pattern reveals its cause. First, the measurements may serve as a sign: the observed effect helps locate an otherwise hidden condition. Second, there may be an effect of changing one variable on the distribution of another. Third, there may be evidence that the outcome would have differed for the particular unit if the exposure had differed. Fourth, there may be evidence that a specified process actually brought about the outcome. It is useful to formulate each as a full sentence. "The alarm is good evidence of heat," "installing this detector increases the frequency of alarms," "without this detector there would have been no alarm on Tuesday," and "the current that rang the bell on Tuesday passed through this detector" make different commitments.

There is a corresponding ambiguity in probability. The frequency of pump starts across machines mixes differences in maintenance, design, and environment. The chance of a start for one particular machine in a specified physical state could have a different value. A historian's probability that the auxiliary starter caused yesterday's start expresses uncertainty about an event which has already happened. Moving between a population frequency, an objective chance, and a degree of belief requires premises in each direction. Much of Cause and Chance concerns objective chances and individual

causal relations. A study that treats the book as a manual for interpreting regression coefficients will miss its central problem.

A further ambiguity concerns implication. A population distribution plus explicit causal assumptions may logically determine an effect. This is identification. An observed pattern may instead make a causal hypothesis more credible while leaving the magnitude or direction unidentified. And the residual uncertainty may be small enough to make acting on it reasonable given the cost of a mistake. Someone who would replace a suspect component for ten dollars might hesitate to attribute a costly failure to its manufacturer on identical evidence. Their different decisions disclose different consequences of error. They need not disagree at all about the evidence.

II. What would count as finding the cause?

Return to Tuesday's pump. Imagine that the only record initially available is a light on the control panel. The light came on just before the motor started. The operator has seen the same sequence a hundred times. A maintenance supervisor wants to install the same panel at another station. The investigator wants to know how Tuesday's motor started. What should either of them learn from the light? The six philosophers considered here would recognize different difficulties in that question. The comparisons that follow are reconstructions. None of them was trying to supply a modern theory of statistical estimation.

Aristotle and Descartes on what an explanation owes us

Aristotle gives us a way to locate the first mistake. In Posterior Analytics I.13 he distinguishes knowing that something is so from knowing why it is so. Consider his astronomical example on its own ancient assumptions. Someone reasons from the planets' steady appearance to their nearness. Someone else reasons from their nearness to their steady appearance. The first inference can establish a fact. The second gives what Aristotle takes to be its explanation. The pair of facts permits reasoning in both

directions. Only one direction supplies Aristotle's reason why. Our confidence in the first inference can grow without turning it into the second.

Now suppose we establish that current in either starter reliably lights the panel. The light gives us reason to infer that current flowed. Disconnecting the bulb might leave the motor entirely unaffected. Aristotle's four causes help explain why it matters to phrase the next question carefully. For an artifact we might ask what it is made of, what organization makes it the kind of thing it is, what initiated a change in it, or what it was designed to accomplish. These are the material, formal, efficient, and final questions distinguished in Physics II.3. The purpose of the backup explains why it was installed. The wiring diagram explains how either branch can reach the motor. Tuesday's investigator still has to determine which branch energized it. The engineer deciding where to install the panel needs to know which features of that organization carry over to a second station.

Descartes puts pressure on the analyst who thinks a satisfactory answer to "how could it work?" has settled "how does it work?" His two clocks in Principles of Philosophy IV.203–206 can tell the same time with different internal arrangements. Descartes's mechanical inquiry also sets aside appeals to divine purposes in explaining physical effects (Principles I.28). There is a genuine difference from Aristotle over the place of ends in natural explanation. In our artifact case the purpose is known because people designed the apparatus. That makes the limit especially clear. Even an undisputed account of the designer's intention cannot establish that the device worked as intended on Tuesday.

Part VI of Descartes's Discourse on Method suggests what the investigator can do. Draw two possible circuits, one in which the panel light responds to current through the auxiliary branch and one in which it responds to either branch. Run the primary alone. The drawings now predict different observations. If the light comes on, Tuesday's light cannot by itself identify the auxiliary as the source. Notice what this test has achieved. We know more about the evidential value of the original observation even if we still cannot tell how Tuesday's start occurred. From Aristotle I take the need to specify the

"why." From Descartes I take the demand for a test whose outcome depends on which proposed explanation is right.

Locke and Hume on what repeated success establishes

Turn to the supervisor's question. Suppose she commissions trials in which a lottery assigns the starter to use and a separate instrument measures whether the motor starts. The difference in success rates persists. Must she wait for a complete theory of the relay before changing policy? Locke's treatment of power in the Essay Concerning Human Understanding gives a reason to proceed. We acquire the idea of active and passive powers through attending to changes and reflecting on agency. One can recognize that heat melts wax before understanding the constitution by which it does so. Recognizing an operation and knowing its internal manner are different achievements. A well-designed trial can likewise support using one starter in a specified setting while the relay remains sealed. (Essay II.xxi.1–4 and II.xxvi.1–2.)

Hume would ask what licenses the next sentence: "The same procedure will work at the second station." Past conjunction teaches expectation, yet an appeal to earlier successful projections already relies on the habit of projecting. Repetition also changes the observer. The hundredth start may feel inevitable because the light now brings the expected motion irresistibly to mind. Hume asks what impression of a necessary connection the observer gained on that last occasion that was missing on the first. Even voluntary action, which helps Locke articulate active power, reveals remarkably little about how the will moves a limb. The strength of the expectation cannot settle the physical range of the capacity. (Enquiry concerning Human Understanding IV–V and VII.)

Hume's challenge changes the scope of the Lockean conclusion. For the tested population we have independent grounds for attributing the contrast to assignment. Extending that contrast to another station requires evidence about what stays stable. Suppose cold weather stiffens the auxiliary relay and only the second station operates outdoors. Equal wiring diagrams leave open unequal success rates. Locke himself stresses how little a substance's familiar qualities disclose about all its powers (Essay

IV.iii.9–14). The two philosophers can therefore support the same practical caution for different reasons. Locke locates the uncertainty in our incomplete grasp of constitution. Hume also challenges our entitlement to project even a familiar sequence under new conditions. My conclusion is to grant the local population claim and state exactly what evidence an extension would require.

There is a consequence for Tuesday as well. In Enquiry VII.29 Hume moves from a formulation involving similar successions to a gloss involving what would have happened without the first event. A working backup separates those formulations. Across trials the auxiliary can reliably produce starts when it works. On Tuesday the primary might have produced a start in its absence. To establish that the auxiliary was necessary for Tuesday's start, we need evidence about the unused primary on that occasion. Its condition never appears in a record that reports only successful auxiliary starts.

Kant and James on what a witness can tell us

Now give the investigator a witness. The operator remembers pressing the auxiliary switch, feeling resistance give way, and hearing the motor engage. Hume has given us a reason to distrust an effortless inference from familiar succession. Kant asks a prior question. What allows the operator to describe the order of an event at all? In the Second Analogy, one may survey the parts of a house in different orders, while a ship moving downstream occupies its successive positions in an order belonging to the motion. For Kant, experiencing an objective alteration involves representing the succession as determined according to a rule. The particular rule governing a given machine still has to be discovered. (Critique of Pure Reason, A189–195/B232–240.)

There is a practical distinction here between establishing the order of two records and establishing the order of the changes they represent. Suppose one channel of the logger buffers its readings for half a second. A display could show switch closure followed by current even if a different relay had already energized the motor. Synchronizing the channels makes the records usable as evidence about the event. That is our application of Kant's distinction. It neither derives an empirical wiring claim from the Second

Analogy nor treats Kant's transcendental argument as a proposal about sensor calibration.

William James would ask what our description lost when it reduced the witness's experience to time stamps. In "The Experience of Activity," effort, resistance, and release give experienced activity a positive character. A technician who felt a jammed contact yield may offer evidence unavailable to a spectator who saw only the lamp. Hume's warning still bites. A second operator could have released the jam from the other side. James himself distinguishes the experience of striving from the further question of where effective activity lies among overlapping processes. His earlier discussion of association also distinguishes the connection between thoughts from the connection those thoughts represent. (A Pluralistic Universe, appendix B, and The Principles of Psychology, I, ch. XIV.)

These differences change how I would interview the witness. "What made you sure you started it?" invites a report of confidence. "Where did you feel resistance, when did it release, and could the other relay have moved that contact?" elicits features that rival histories predict differently. Kant supplies the reason to establish the event's objective order. James supplies the reason to ask what occurred within that order. Hume supplies the reason to test the witness's attribution against an alternative. If the other relay could mimic the whole sequence, the testimony warrants a narrower conclusion. If the felt release is specific to the auxiliary contact and a separate log shows that contact carrying current, two pieces of evidence jointly support a particular history. Neither observation tells the supervisor how often the starter will succeed at the second station.

We can now divide the work. To learn the population effect, vary the starter in comparable conditions, count starts, and establish where the comparison applies. To learn the history of Tuesday's start, establish the order and source of the current. To learn whether Tuesday's start required the auxiliary, investigate the primary that remained unused. The first task can succeed with a sealed relay. The second can succeed after a single well-recorded episode. The third may remain unresolved after

both of the others succeed. The next section asks how experimental design can earn the comparability that the first task requires.

III. How comparisons acquire causal force

The nineteenth-century confrontation between Mill and Whewell gives a useful pair of answers. In A System of Logic, book III, chapters VIII and X, Mill develops the methods of agreement, difference, and concomitant variations while expressly confronting the plurality of causes and the intermixture of effects. A comparison gains force when a difference between cases isolates a circumstance capable of making the observed difference. Whether two cases match in all relevant respects is itself a substantive claim. Whewell's consilience of inductions, developed in The Philosophy of the Inductive Sciences, gives a reason to credit an explanation when a conception reached through one class of facts succeeds with another class as well. A successful cross-check is particularly valuable when an error that would explain the first result would not explain the second.

John Snow's inquiry into South London's water supply illustrates both. In streets where the suppliers' pipes were intermingled, knowing which company served a dwelling made it possible to separate a water-supply prediction from a simple neighborhood-air prediction. The mortality figures alone could not reveal that property of the comparison. Knowledge of how the companies laid their pipes, where their water came from, and how Snow established household supply made the figures probative. Yet intermingling streets did not randomize every household attribute. A careful observer could accept that the comparison substantially favored waterborne transmission while asking how errors of classification or customer differences would affect the estimated risk.

Fisher's randomized allocation offers an unusually strong answer to Mill's difficulty. A lottery can give comparable assignment probabilities to units whose differences the researcher cannot enumerate. Under a specified sharp null hypothesis, the known assignment mechanism supplies a distribution against which to assess an observed

contrast. Even here one must distinguish assignment from the treatment received, check whether outcomes were recorded differentially, and define the population to which the result will be applied. Randomization gives a reason for a particular comparison. It has no automatic power to settle the subsequent medical history of an individual assigned to one group.

Russell's "On the Notion of Cause" (1912–13) and the covering-law account of Hempel and Oppenheim (1948) expose a difficulty that persists even with perfect equations. A functional relation may permit calculation in either direction. A shadow and solar angle permit a calculation of a pole's height, although the shadow supplies little explanation of how the pole came to be that height. Bromberger's later objection to the covering-law picture repeats the structure of Aristotle's planetary case. The direction in which a calculation runs can be chosen for convenience. An explanation commits its author to what the system would sustain under relevant alterations or to what produced the facts from which the calculation begins.

Reichenbach's common-cause principle (1956) gives positive work back to correlation. A population dependence that cannot be attributed to influence in either direction can, under the principle, license the search for a common causal history. This is a weaker inference than locating a common cause, and it presupposes that the association describes the intended population. A relationship between two reasons for referral may arise in a specialist's waiting room solely because having either condition makes referral likely. Suppes's probabilistic analysis (1970) and Cartwright's work on causal laws (1979) sharpen the associated question of background. Conditioning on a common source can remove a misleading association. Conditioning on a mediator can remove an effect one wished to estimate. Conditioning on a common effect can create an association between its causes. The instruction to consider all the evidence does not entail conditioning on every measured variable.

Modern work turns these observations into a testable claim about what data can determine. In a structural causal model, one can ask whether two models that satisfy the assumptions and agree on all observable probabilities can nevertheless give

different answers to the intervention question. If they can, the effect is unidentified by those observations under those assumptions. A researcher can then ask for a design feature, measurement, or external result that excludes one model. Pearl's Causality and the potential-outcome tradition give overlapping ways to do this. Woodward's interventionism addresses a related philosophical question about what an invariant dependence under interventions means. The mathematical result that an effect is identified settles neither the empirical credibility of the assumptions nor the metaphysics of individual production.

An explicit identification result clarifies how much one can claim. Write Y(1) and Y(0) for a unit's outcome under a specified exposure and its alternative. Suppose C records a set of pre-exposure conditions. If assignment within each level of C is independent of the two potential outcomes, observed outcomes correspond to the exposure actually received, and both exposure states occur at every relevant level of C, the average effect in the study population equals the average of the within-C differences, weighted by that population's distribution of C. In compact notation it is $\Sigma c\ P(c)\{E[Y \mid X=1,c] - E[Y \mid X=0,c]\}$. This expression assumes an appropriate unit and exposure definition, including how interference between units is handled. Its job is to tell us precisely which comparison would answer the question if those conditions hold. Asserting that C makes the two groups exchangeable supplies a causal premise. It is never a consequence of choosing to put C into the formula. (Hernán and Robins, What If, chs. 1–3.)

Here is a direct failure of that premise. In a greenhouse, sixty percent of plants exposed to a new light flower, compared with thirty percent in a control area. Suppose plants of a vigorous variety always flower in either location and the rest never do. If the gardener placed sixty percent vigorous plants under the new light and thirty percent in the control area, the true effect is zero. The same figures are compatible with random allocation and an effect of thirty percentage points. The gardener's allocation records, if independently reliable, do epistemic work that a larger count of flowers cannot do. In the randomized example below I deliberately grant an answer to this first question.

That is how we can isolate a second and quite different limitation on individual attribution.

The same logic yields a qualified positive answer for observational discovery. With three variables, two independent candidate inputs can become dependent when one selects on a common effect. Under the assumptions of an acyclic graph, causal sufficiency, the causal Markov condition, and faithfulness, a pattern of independences of this kind can orient an unshielded collider. By contrast, three observed variables arranged in a simple chain and in a fork can have the same conditional independences. The word “can” matters. Direction is sometimes inferable from observational patterns under explicitly defended assumptions. Graphs in the same equivalence class need other evidence to distinguish them. A study can therefore warrant an arrow somewhere in the system without identifying every arrow or the magnitude of the effect of intervening on it. (Stanford Encyclopedia of Philosophy, Causal Models, discussion of causal discovery.)

IV. The dispute inside Cause and Chance

The editors set a harder problem than the discovery of confounding. Imagine that the initial event does influence the outcome and that both routes are correctly understood. One route is usually effective. The initiating event disables that route and initiates a less reliable process which actually succeeds. The event has lowered the overall chance of the outcome. Yet the outcome can be traced through the process the event initiated. The editors distinguish early preemption, where the alternate process is blocked before completion, from late preemption, where arrival of the effect itself prevents the other process from completing. The distinction matters to an investigator. Finding that both processes were underway shortly before the outcome is consistent with only one having completed. (Dowe and Noordhof, introduction, pp. 1–11.)

Dorothy Edgington opens her chapter with someone who misses a flight because the car breaks down. Before departure it was reasonable to expect arrival in Paris conditional on catching the plane. Later it becomes known that the plane crashed. What should one

then believe would have happened if the traveler had caught it? The crash can inform that retrospective judgment insofar as the traveler's presence would have made no difference to the flight's fate. It is perfectly rational for relief to replace frustration when the news arrives. Some facts learned after a missed opportunity improve our judgment of the counterfactual. That happens because we have a reason to carry those facts across the hypothetical change. (Edgington, pp. 12–27.)

This argument bears directly on the apparent circularity of counterfactual analysis. If the rule for holding subsequent facts fixed already asks whether they are causally independent of the antecedent, the counterfactual cannot straightforwardly define causal independence from scratch. As a method of inquiry, however, it may be entirely legitimate to use independent evidence about one causal relation to learn about another. To investigate whether boarding would have changed the traveler's fate, one could use the passenger manifest, the failure investigation, and the seat distribution. That inquiry would beg the question only if the independence one had assumed amounted to the very claim one purported to establish. An account of evidential warrant can therefore succeed where a proposed reduction of causation to wholly noncausal terms fails.

Phil Dowe's chapter gives the issue physical form. In a schematic decay example, an atom can reach the same final state by two routes of different reliability. On the actual occasion it takes the less promising branch and reaches the final state. For Dowe, isolating the path through which production actually occurred can recover a sense in which taking that branch favored the outcome along that path. The appeal to the intrinsic character of the actual process explains why the presence of a more promising alternative need not cancel its claim to have produced the result. Dowe also considers refining the description of the effect to include the route. That maneuver alone is insufficient. One can still ask what caused the final product to exist, a question which does not stipulate its ancestry in the description of the outcome. (Dowe, pp. 28–38, especially 33–37.)

Helen Beebee asks what constrains this abstraction. Spraying a plant with defoliant may harm it and it may survive. Imagine eliminating enough ordinary routes to survival, or adding a disease that could only be transmitted through leaves, and a comparison can be constructed in which loss of the leaves helps the plant survive. But the plant in the original case was never exposed to that disease. An account that permits arbitrary removal of alternatives can turn an ordinary history of survival despite damage into a history of rescue. Beebee takes hindering seriously as a relation in its own right. A process can run from an intervention to features of the eventual state even though, in her analysis, that intervention hindered the outcome being explained. This preserves a chance-increase requirement for calling something a cause at a cost to ordinary judgments in some switching cases. (Beebee, pp. 39–57, especially 48–55.)

Christopher Hitchcock supplies the most useful response for an investigator. His distinction between an independently variable contributor and alternative values of one variable limits what can count as a separate route. In the backup case the operation of one device can causally disable a different device. Whether the second operated is a distinct question. By contrast, picturing a plant as sprayed and unsprayed gives two incompatible values of one variable. Those values cannot be treated as independent physical channels just because separating them would yield a convenient answer. A causal graph makes that modeling choice visible. It does not certify that the choice is correct. Evidence about how the system is built must still decide whether the supposed routes can vary independently in the relevant respect. (Hitchcock, pp. 138–151, especially 144–149.)

This disagreement gives the pump example a sharper moral. The electrical trace is valuable because current in one circuit can be evidenced independently of the operator’s label and of eventual motor rotation. A historian who defines “the successful auxiliary route” as whatever happened between selecting auxiliary and seeing the pump run has explained nothing. A historian who can show that a monitored contact closed, current crossed that contact, and the primary contact remained open has excluded alternatives. The first account chooses a description after the outcome is known. The

second takes risks about what independent records ought to show. Beebee's objection thus supplies a discipline even to someone who ultimately sides with Dowe about production.

Douglas Ehring moves the issue from the description of routes to the identity of what persists along one. In his cases, arranging dependable substitutes puts pressure on attempts to distinguish the producer entirely by patterns of counterfactual dependence. He explores a singularist component involving a persisting particular, such as an instance of a property. The philosophical advantage is the possibility of distinguishing two qualitatively similar histories by what continued through them. The evidential cost is that establishing persistence itself requires criteria. A technician can trace a tagged sample because the label was attached before the split and its behavior has been checked under the relevant conditions. A metaphysical commitment to persistence by itself supplies no trace. (Ehring, pp. 58–76, especially 72–74.)

Michael Tooley provides the volume's most extensive attack on treating chances as a metaphysical starting point. He distinguishes relative frequencies, objective chance, and logical probability, and asks what each could contribute to an analysis of a causal relation between particular states of affairs. He argues that treating objective chances as ultimate creates difficulties, that probabilities and noncausal facts can leave singular relations underdetermined, and that relations between causation and logical probability are better captured by theoretical postulates. One need not accept his alternative to appreciate his challenge. If investigators call a fitted population proportion the objective chance for this exact specimen, they have moved from an estimate to an ontological commitment. Tooley asks what justifies that move even in an ideal science. (Tooley, pp. 77–119, especially 116–118.)

Stephen Barker asks whether the repair to chance-raising has conceded too much to the counterfactual theorist. Once enough facts can be varied to isolate the route, dependence itself can sometimes be recovered. His account uses embedded counterfactuals to examine what would depend on what in a suitably altered situation. He also considers cases where each successive link appears causal yet the initial event

seems a poor candidate for causing the final result. For our inquiry the implication is methodological. Replacing one difficult probability with a more elaborate hypothetical comparison can relocate the difficulty into the selection of permissible changes. It is progress when the investigator can say which additional observations bear on that selection. (Barker, pp. 120–137, especially 120–123 and 129–131.)

Murali Ramachandran and Igal Kvart offer two increasingly fine-grained responses to the temporal problem. Ramachandran distinguishes the chance of an effect assessed early from the chance assessed just before the effect, attending also to whether it has occurred too early. A process can make an event likely in the long run while having little to do with its occurrence at the observed time. Kvart examines whether probabilistic relevance survives specification of further actual circumstances. He uses the ideas of a stable increaser and a stable screener to address cases in which an initial probability decrease obscures the operative route. His discrete transition examples are especially instructive: demanding a continuous trail of observable intermediates would make the answer depend on a questionable physical assumption. A failure to observe an intermediate stage has evidential force only if the proposed process should have left one at the resolution available. (Ramachandran, pp. 152–162. Kvart, pp. 163–187, especially 177–180.)

Paul Noordhof closes the volume by defending a counterfactual account that permits specified competitor events to be excluded while comparing chances at relevant times. His reply is neither a retreat to simple probability-raising nor an unrestricted permission to subtract inconvenient facts. The choice of admissible events and the demand that certain dependencies concern actual events do substantive work. Cases of trumping, alternative routes, and direct action put pressure on the details. The disagreement with Edgington turns partly on whether the admissible counterfactuals can themselves be given semantics independent of causal notions. For the inquiry here, the important division is between proving that a causal judgment is justified and defining causation in a vocabulary that contains no causal concepts. Success at the

former may require considerable prior causal knowledge and still be a genuine intellectual achievement. (Noordhof, pp. 188–202.)

What has this debate decided for our two readers? The supervisor needs the overall chance of a start, including both routes. The investigator needs an independent reason to single out the branch that carried current. Descartes’s second possible clock is the right objection to a diagram inferred from appearances alone. James’s felt release becomes a reply only if the other clockwork could not have produced the same sensation. Beebee’s challenge explains why one must establish that last premise independently. Three consequences follow. The sign of the net change in chance can conceal opposing processes. Evidence about an actual path earns weight when the path can be distinguished independently from a convenient redescription of the result. And an event can make a difference to the time, manner, or particular route of an outcome while leaving its eventual occurrence unchanged. The last point makes the wording of the question consequential. A pulse might cause the pump to start at 9:01 while a timer would have started it at 9:02. “Started before 9:05” and “started at 9:01” are both legitimate outcomes. One cannot switch between them midway through the inference.

V. How far a population result reaches into one case

Consider a binary exposure and a binary response, both well defined. Imagine a randomized study large enough that estimation error can be set aside. Sixty percent respond under exposure and thirty percent respond without it. Assignment is independent of each unit’s potential responses, outcomes are measured consistently, and one unit’s assignment does not change another unit’s response. These assumptions identify a thirty-percentage-point increase in average response in the study population. They are generous assumptions. Their purpose is to expose an uncertainty that survives an otherwise ideal design.

Give each unit two potential outcomes, one under each assignment. In this illustration they are fixed binary values. A unit can respond either way, respond only with exposure, respond only without it, or never respond. “Only with” is the category of units

for which exposure was necessary for the response and would suffice in the specified background. For an exposed responder whose identity tells us nothing further relevant, the retrospective question is what share of exposed responders belongs to that category. This is a probability of necessity. It is already a narrower question than asking which biochemical or mechanical process produced the observed response. (Pearl, "Causes of Effects and Effects of Causes," 2015.)

Response type	Population A per 100	Population B per 100
Responds either way	30	0
Responds only with exposure	30	60
Responds only without exposure	0	30
Never responds	40	10

Read the table one column at a time. Each column describes a whole population in which we can stipulate how every unit would respond under either assignment. In a real trial we would see only one of those responses per unit. First check what the trial could see. Population A gives sixty responses per hundred with exposure and thirty without. Population B gives exactly the same sixty and thirty. An ideal randomized trial observing only those two response rates has no way to tell the populations apart. In A, half of the exposed responders needed the exposure for their response. In B, every exposed responder needed it. The difference between the population effects in the two worlds is zero. The difference between the two probabilities of necessity is fifty percentage points. Increasing the trial's sample size would improve the precision of the observed margins while leaving that discrepancy intact.

One can derive what the experiment does establish. Call the proportion who respond either way a. The control response rate sets a ceiling of 0.30 on a, and the lower limit here is zero. Of the 0.60 who respond under exposure, 0.60 minus a respond only with exposure. The probability of necessity for an otherwise unselected exposed responder is therefore (0.60 − a)/0.60. Its range is one-half to one. The two populations in the table attain the endpoints, so these bounds are sharp. The positive lower bound is a substantive answer. Even without further observations about the particular individual,

population evidence can rule out entire ranges of retrospective credence under the assumptions of this model.

Suppose independent biological or engineering evidence establishes monotonicity in this population: exposure cannot prevent a response that would otherwise have occurred. Then everyone who would respond without exposure belongs to the either-way category. The population is forced into the mixture represented in column A and the probability is one-half. It may seem strange that strengthening a causal assumption fixes the probability at its lower bound. The reason is simple. Allowing the exposure to prevent some responses means it must also produce more responses elsewhere to preserve the same net gain. The average is a balance of gains and losses. Removing the losses fixes the number of gains.

These are claims about a randomly drawn or otherwise exchangeable member of a specified group. Imagine learning that the individual's unit bears a manufacturing mark concentrated in one response type. The aggregate probability may now be a poor guide. To use a mark as evidence, one needs independent information about its association with the potential response types or its connection to a process. Retrospectively dividing responders into those called "susceptible" because they responded and the rest adds no information. The reference class problem is therefore both statistical and causal. We must identify a group whose measured risks apply and explain why the information available about this individual does or does not warrant further differentiation.

The example also makes clear why "risk doubled" can be such a misleading shortcut. In this randomized illustration the treated-to-untreated risk ratio is two. Under exchangeability, consistency, our binary specification, and monotonicity, precisely half the exposed responders would have responded without exposure. Remove monotonicity and one-half is a lower bound in this example. Remove exchangeability between the relevant groups and even the interpretation of the observed doubling as a causal risk ratio is in jeopardy. Change the question to which of two sufficient processes actually completed and the same numerical calculation answers a different question. No

general verdict about a named person, much less a rule for adjudicating individual disputes, follows from a risk ratio alone.

There is a metaphysical qualification. The response-type construction pairs outcomes across alternatives for each unit. It is particularly transparent in worlds where a machine’s complete state determines its response under each switch setting. Under genuine indeterminism there may be several ways to couple the random histories in the two hypothetical worlds. What it means to hold “the same luck” fixed can then become contentious. The calculation shows an identification gap even in a favorable model. It does not establish that every probabilistic theory of actual causation must posit a uniquely correct pair of outcomes for each individual. Edgington’s lesson about what may be carried into a counterfactual has a mathematical counterpart here.

A second qualification follows from preemption. In the table, an either-way unit counts as someone whose response did not require the exposure. Yet the exposure might initiate a new process that produces the response before an otherwise effective backup can operate. In that case it could be the actual producer while falling in the either-way category. A trace of the process could raise confidence in that productive judgment even if independent evidence made the probability of necessity very low. Here Locke’s question about a power exercised on an occasion and Hume’s question about the absence of its putative cause come apart within the very same row of the table. This is why the two questions Alan wants to preserve generate a third useful distinction inside the singular question itself: necessity for this outcome and participation in its production can have different answers.

VI. What gives a particular observation leverage

Maureen Christie’s The Ozone Layer supplies a case where an actual historical process, an experimental tradition, and population-like patterns of measurements come together. Antarctic observations revealed a regional and seasonal loss of ozone. A spatial association between elevated chlorine monoxide and depleted ozone favored a chemical account. A moving aircraft, however, samples parcels of air with histories.

Mixed or transported parcels can generate correlations. Concentration measured at one point in an evolving system need not reveal which reactions produced the concentration there. Taken by itself, a map is poor evidence for a rate of chemical destruction. (Christie, The Ozone Layer, especially the discussion of the airborne expedition and pp. 140–143.)

Christie describes how the comparison became more demanding. A particular circulation account proposed upward movement of ozone-poor air. Evidence of descent in the vortex bore directly against that proposed direction. Measurements bearing on chlorine chemistry supplied positive evidence for a process expected under a rival account. These are different accomplishments. Disconfirming one circulation story does not eliminate the contribution of atmospheric transport. Evidence favoring chemistry need not show that transport played no role. Christie explicitly notes that the alternatives were not logically exclusive. The productive question asks which processes contributed to this depletion. The intervention question might ask how much change in a defined class of atmospheric conditions would follow from a specified change in chlorine loading. Each uses parts of the same evidence. Each requires additional assumptions distinctive to its target.

This example explains why convergent evidence sometimes warrants a historical conclusion surprisingly well. A time-resolved profile can constrain where a reaction occurred. A tracer can constrain the provenance of a parcel. Laboratory work can constrain feasible reaction rates. The explanations that accommodate one constraint but fail another are discoverable in advance. By contrast, collecting many readings with the same miscalibrated instrument merely reproduces the same vulnerability. Independence of error and difference in what the instruments physically measure matter at least as much as the count of observations. Whewell's consilience becomes a practical requirement to specify which alternative can survive each line of evidence.

Marcel Weber's Philosophy of Experimental Biology gives a subtle warning about this very argument. In discussing Perrin's independent determinations of Avogadro's number, Weber considers Salmon's common-cause reconstruction. Different

laboratories working with different samples converged on roughly the same value. What individual event could be their common cause? No particular collection of atoms was in every laboratory. Convergence here speaks for a shared structure or kind of entity manifested across experiments. Treating that structure as one common cause-token misdescribes the argument. Weber explores an inference to the best explanation of agreement across physically different methods. The example reverses a familiar mistake. One can seek a singular common producer even where the evidence primarily bears on the reality of a general structure. (Weber, pp. 281–283.)

The recovered anthology chapter by Nancy Cartwright gives an illuminating second reading of the very same episode. In "The Reality of Causes in a World of Instrumental Laws," she emphasizes Perrin's ability to choose effects sensitive to the exact character of the cause being investigated. Her question is what makes an inverse inference from effect to cause experimentally successful given a model of the connection between them. Weber's question is what kind of common cause one could be crediting across all the samples. Together they yield a useful refinement. One can have local warrant for one physical process in each apparatus and cross-experimental warrant for a common structural property. The premises that support these two inferences can be challenged separately. A calibration error confined to one instrument would weaken the first for that apparatus. A shared mistaken modeling assumption could damage the second across many instruments at once. (Cartwright, in Boyd, Gasper and Trout, eds., The Philosophy of Science, pp. 379–386, especially 385.)

The same issue arises when experimental intervention earns confidence in a mechanism. Hacking's discussion of using electrons as tools in Representing and Intervening puts weight on the practical reliability of a property in a new experimental arrangement. Successful manipulation supplies evidence unavailable to someone who has only fitted a model to familiar observations. Weber asks how far the result supports realism about an entity independently of the theory specifying its powers. A scientist who uses a component to charge another apparatus has evidence that something with the required capacity is present. Choosing among all rival accounts of that component's

constitution remains a larger task. Locke's distinction between establishing an operation and knowing its internal manner survives contact with twentieth-century experimental practice.

Tversky and Kahneman supply the complementary case in "Judgment under Uncertainty" (1974). Flight instructors observed improvement after criticism of poor performances and deterioration after praise of unusually good ones. A noisy performance measure creates these contrasts when feedback is triggered by extremes. Part of the selection criterion was a transient deviation which is unlikely to be repeated on the next attempt. The comparison mixes the effect of the feedback with regression toward the mean. This does not settle whether the feedback had any influence. It explains how precisely the observations that made the instructors confident could arise under a rival account. One can compare different feedback after comparable performances, or introduce a contemporaneous group subject to the same selection. Both give the rival more difficulty. (Local reprint, PDF p. 8.)

J. D. Trout's "Scientific Explanation and the Sense of Understanding" helps explain the individual version of the mistake. A retrospective sequence can feel compelling as soon as one learns its outcome. Familiarity and hindsight may produce that satisfaction even when a comparison would reveal poor predictive discrimination. This qualifies the use of James's phenomenology of activity. Experienced resistance can be evidence of contact. The satisfying feeling of having correctly identified the cause is a different phenomenon. One way to test the latter is to ask what records the account predicts that have yet to be inspected. The causal explanation gains independent support if it predicts a trace on the previously unopened circuit log. (Trout, 2002, pp. 212–215.)

The folder also contains Darrin Belousek's analysis of Bell's theorem. A particularly relevant passage separates a technical independence condition from a capacity to signal. Statistical dependence conditional on a specified state does not by itself give an observer controllable access to that state. Whatever interpretation of quantum theory one favors, that inferential distinction is consequential. A proof excluding a specified class of local common-cause explanations is a result about the class with its

assumptions. It does not by itself license the operational conclusion that one observer can encode and transmit a message at will. Classical intuitions about intervening on correlated variables have a domain of application, and physics helps expose its boundaries. (Belousek, local PDF p. 9.)

A distinction among the local methodological papers also bears on the proposed test. In "Philosophy of Science: From Justification to Explanation" (1988), Aharon Kantorovich distinguishes accounting for scientific decisions from providing the norms that justify them. In "On the Inextricability of History and Philosophy of Science" (1977), Richard Burian makes historical considerations part of assessing the support a theory actually earned. Taken together these points give history a precise role. Learning that investigators adopted the chlorine hypothesis because a plot looked striking can explain a change of mind. Reconstructing what a circulation model predicted before the flight can change our judgment of what the plot warranted. Cassandra Pinnick and George Gale's "A Troubling Interaction" likewise identifies the roles of generality and particularity as a point of methodological friction. The investigator should recover what was known when a discriminating test was designed, even if hindsight subsequently supplies additional independent evidence.

Michael Friedman's "Kant, Kuhn, and the Rationality of Science" (2002) adds a difficulty for the notion of a fixed model class. Standards that make a scientific comparison possible can change during scientific revolutions. One cannot always list all live alternatives in advance using a single timeless vocabulary. This objection limits any claim that the working account is a universal decision procedure. It leaves room for a practical comparative judgment within an inquiry and for revision of the inquiry itself. An apparent contradiction among studies may indicate that one needs a new variable, a finer distinction between events, or a different account of what the instrument measures. That possibility is a reason to preserve the assumptions with the conclusion.

VII. A provisional account of causal warrant

A defensible generalization emerges if we ask what would remain uncertain even for a perfectly competent observer. Let the live models consist of the causal arrangements and histories consistent with the design, validated measurements, background knowledge, and available data. Specify a target claim. If every admissible model gives the same answer, the evidence and assumptions identify it. If their answers differ but lie in an interval, the target is partially identified. If several models survive and a particular observation is more expected under some than others, that observation can still alter their relative credibility. For a Bayesian this last point can be made exact through a likelihood ratio. Its force depends on the choice of rivals and on how their auxiliary assumptions distribute probability over what was observed.

This formulation faces an obvious objection. If all the work has gone into selecting the admissible models, saying that they agree may sound empty. The response is to require a public account of what excluded a rival and what observable consequence would let it back in. Snow could cite the two companies' pipes and address lists. Christie could point to the direction of circulation and the chemical measurements. The pump investigator can state the resolution of the logger, the independence of its channels, and the failure modes that could create a false trace. In each instance the bridging premise has a different evidential history from the association it helps interpret. Complete independence is rarely possible. Some capacity for separate correction is essential.

For a population effect the target requires a treatment contrast, an outcome definition and horizon, and a population. Are we asking about offering a device or actually using it? About ever starting or starting before a deadline? About all machines manufactured or the machines still in service? The same data can answer one of those questions while offering only weak clues about the others. What must be established about the world is that the contrast represents a coherent alteration under which the relevant outcome distribution is defined. What must be established about the evidence is that the comparison estimates that alteration under defensible assumptions. An intervention

effect can be well identified even when the many microscopic routes that implement it remain unknown.

For an individual attribution, identify the rival history. Did the effect arise spontaneously, through an alternative producer, or through both? Would it have occurred anyway at a different time? Is the question whether a factor was necessary, whether it contributed, or whether it explains an unusual feature of the result? A population result can give a prior or a bound. An individual measurement adds weight in proportion to its ability to discriminate those histories. A marker found with equal frequency under the two rival routes has little diagnostic value however impressive it looks. A marker confined to one route can be powerful even if the event was improbable beforehand. The marker's specificity, detection quality, and timing need empirical support.

One can state a modest metaphysical conclusion without choosing between all the essays in Cause and Chance. The information required to specify a net response under an intervention and the information required to specify a realized route are logically separable in at least some possible systems. The pump and the two response-type populations demonstrate that separation constructively. A unified theory might nevertheless explain both in terms of a richer common structure. A graph with equations and sufficiently detailed initial conditions is one candidate. A theory of powers manifested in organized mechanisms is another. The examples constrain either theory: it must recover the possibility of unchanged aggregate effects with different individual histories, as well as unchanged actual histories with different available backups. They do not prove that there are two irreducible kinds of causation.

There is a corresponding positive conclusion about knowledge. A correlation can warrant a causal inference even when it does not establish the complete mechanism and even when the effect of changing the correlated sign would be zero. The first possibility is familiar from randomization and natural experiments. The second arises when the sign is a well-understood effect of an earlier cause. The important question in the second case is whether the sign discriminates that antecedent from other possible

sources. An alarm's reliability under deliberate detector activation and its value for diagnosing a fire are empirically distinct. The latter requires information about false alarms and about the prevalence of fires in the setting where the alarm sounded. Causal informativeness extends beyond actionable influence.

A final test is whether a theory can recognize warranted suspension of judgment with some precision. In the numerical experiment, the population gain and an interval for individual necessity were warranted. If the competing circuits left indistinguishable records at the available sampling interval, specific attribution might remain unresolved. If a chemical and a transport process can coexist, a claim that one occurred cannot be promoted without argument to a claim that it fully explains the observed magnitude. Saying exactly which alternatives survive gives useful content to uncertainty. It also identifies which next observation could change the answer and whether the question is empirically decidable under the current description.

The question that remains most worth pursuing is what makes a causal contrast legitimate before we know its answer. Dowe and Hitchcock appeal in different ways to separable paths. Beebee shows how abstraction can manufacture a favorable comparison. Edgington asks what facts are entitled to travel into a hypothetical world. Mill, Snow, and Fisher explain how the world sometimes supplies the comparison for us. These are complementary approaches to a single constraint on inquiry: an investigator must be able to defend the proposed comparison without relying entirely on the verdict it was chosen to produce. That principle is less ambitious than an analysis of causation in noncausal terms. It is strong enough to explain why some correlations, in their actual evidential settings, entitle us to remarkably precise causal conclusions.

VIII. An argument from changing the alternatives

My proposed synthesis begins with a thought experiment about the backup starter. Keep the history of Tuesday's auxiliary circuit fixed, down to the current trace. Now vary how reliable the disconnected primary would have been. If the primary was broken, selecting auxiliary probably improved the chance of a start. If the primary was

nearly infallible, selecting auxiliary lowered it. Tuesday’s current still followed the same route. Next perform the converse operation. Keep both starters’ success frequencies fixed, including the frequencies under assigned switch settings. Vary which circuit is connected on Tuesday by introducing an unnoticed automatic override in a small fraction of machines. We can preserve every aggregate start frequency by compensating elsewhere in that population. Yet a current trace taken downstream of the junction now leaves open which starter produced this start. A measurement on each branch upstream of the junction could distinguish them.

These two operations suggest a criterion with empirical bite. Before using an association to answer a causal question, ask which alterations to the world would leave the evidence intact while reversing the proposed verdict. Some alterations change the unobserved alternative and preserve the actual sequence. Others change the actual sequence and preserve the distribution across alternatives. An investigator earns a stronger causal conclusion by finding evidence that would detect the alterations that matter for that conclusion. Testing the performance of the disconnected primary matters to necessity and policy. Measuring current before the junction matters to production. A single instrument can do both jobs if it independently constrains both alterations. There is no general rule that experiments answer only population questions or that historical traces answer only singular questions.

A small calculation illustrates what “detect” means here. Suppose auxiliary and primary are initially equally plausible producers of a recorded start. For simplicity, assume they are the only possible producers and are mutually exclusive in the machine under examination. A downstream sensor lights with probability 0.99 whichever circuit started it. Seeing the light changes neither relative probability. Now suppose a separately tested branch sensor has a 0.90 probability of lighting if auxiliary produced the start and a 0.10 probability if primary did. Its light multiplies the odds for auxiliary by nine, giving a posterior probability of 0.90 from the stipulated equal prior. Give auxiliary a prior probability of 0.10 and precisely the same evidence instead yields a posterior probability of one-half. An impressive posterior, a discriminating observation,

and a reliable instrument are three distinct evidential virtues. One must ask reliable for detecting what. These values are illustrative assumptions, not new empirical findings.

This account yields a stricter defense against hindsight than merely demanding more evidence. A comparison or route may be discovered after the outcome. That timing alone is harmless. What matters is whether its diagnostic credentials can be assessed separately from its success in the case that suggested it. A previously unnoticed branch sensor can be bench-tested under both routes. A chemical hypothesis developed after inspection of a profile can still predict a tracer distribution that the investigators have yet to measure. Conversely, a distinction written into a protocol before the study can remain uninformative if its two nominal routes always share the same sensor. The relevant independence is the possibility of exposing a mistake, not the calendar date on which a hypothesis was proposed.

One may object that a scientist could imagine unlimited rival histories, including fanciful tampering with every sensor. How much discrimination is enough? The practical answer begins with the claims being made. A logical identification result quantifies over a declared model class and is conditional on it. Comparative evidential support can exist when some rivals remain. A strong historical attribution requires attention to rivals made credible by the system’s known behavior, the selection procedure, and the case record. A newly discovered override that has actually failed in similar machines deserves more weight than a logically possible undetectable saboteur. This judgment is fallible, but it is corrigible by examining independent maintenance records. It neither requires that every possible story be disproved nor permits the investigator simply to discard alternatives they dislike.

The proposal also explains a neglected middle ground. Sometimes a sign discriminates perfectly between two possible source classes while doing nothing to discriminate different routes within either class. In that case one can know the source class and remain wholly uncertain about which individual source acted. Sometimes the reverse holds. A recording can identify the particular switch that closed while telling us little about the family of disturbances that led a person to close it. Causal knowledge can

therefore gain resolution along one dimension without gaining it along another. The appropriate conclusion states the dimension along which the evidence made progress. This is more informative than assigning a single grade such as “suggestive” to an entire study.

Two neighboring proposals help locate what is distinctive in this argument. Ned Hall’s “Two Concepts of Causation” explicitly distinguishes event dependence from production and defends a stronger metaphysical division. I borrow that distinction. My claim is that an investigator can discover which evidence bears on each judgment without first accepting Hall’s division as fundamental. Halpern and Hitchcock’s “Graded Causation and Defaults” explains how judgments about normal alternatives can enter assessments of actual cause. That raises a further question about the backup test: would a disconnected spare count as an alternative if using it was unprecedented or prohibited? For the engineering effect of a switch setting, physical availability and the stated intervention govern the calculation. For explaining why an operator’s omission drew attention, a normal operating procedure can partly determine which contrast was asked about. The change in the explanatory question must be made explicit before a change in causal verdict is treated as a disagreement about the wiring.

Consider an omission. A technician failed to press the reset, and the pump stayed silent. There is no pulse from the missing button press for a branch sensor to find. If repeated controlled resets reliably start the pump in this precise fault state, and records show that the state persisted until someone did press reset, the history can support the omission claim. The task is to discriminate a world in which resetting would have worked from one in which the motor was beyond recovery. The service rule can explain why this particular technician’s omission is singled out among innumerable people who also did nothing. A defensible evidential account must allow that some singular claims earn support through an unrealized but well-tested alternative. Insisting on a transmitted token for every warranted attribution would preclude too much.

Where, then, do I come down on the historical disagreements? Aristotle gives the investigator permission to reason back from a sign. Descartes sets the price of

narrowing that inference to one hidden arrangement: find a consequence on which the arrangements diverge. Locke gives the supervisor permission to act on a demonstrated capacity while the arrangement is unknown. Hume sets the price of extending that decision to another population: establish what remains stable when the circumstances change. Kant and James divide the labor within the individual case. A defensible order makes the contact's release eligible as a cause. The character of the release may distinguish a route within that order. My departure from a purely regularity-based reading is to give such a single episode evidential weight. My limit on that concession is to ask whether a different producer could yield the same experience. These commitments can be tested separately when reasonable investigators disagree.

The most serious philosophical objection is that the proposal offers only an account of evidence conditional on causal models. I accept its limited ambition, with one qualification. The criterion also constrains causal concepts: anyone claiming that the net intervention effect exhausts singular production owes an account of the two alterations above. Anyone claiming that an intrinsic trace exhausts causal necessity must explain why changing the independently available backup changes the necessity judgment. These constraints leave open what fundamentally constitutes a process, whether suitable counterfactuals can capture it, and whether all scientifically useful causal concepts have a common analysis. They close off some identifications while leaving several metaphysical accounts alive. That is a substantive result available before the larger dispute is settled.

The Stanford Encyclopedia of Philosophy provides the baseline for that judgment about what is open. Its entries on probabilistic causation and causal models distinguish distributions from the structures and assumptions used to interpret them. Its account of counterfactual theories treats preemption as a continuing dispute, including replies which argue that more carefully specified counterfactual dependence can survive the objection. Its discussion of causal explanation examines interventionist and mechanistic approaches, while The Metaphysics of Causation keeps questions about the causal relata and the nature of the relation itself in view. I therefore treat the separation of

evidential tasks defended here as my proposed synthesis. I leave the stronger claim that production and dependence are irreducibly different relations undecided. Deference to these entries calibrates the status of the argument; it does not supply its premises.

The question I would carry into a further draft is now fairly precise. For a given body of observations, what independent evidence licenses our choice of the relevant counterfactual alternative, the pathways allowed to vary, and the population across which to average? When we can answer, a correlation may support a population estimate, an individual probability, a specific account of production, or more than one of these. When we cannot, it may still identify exactly where the uncertainty lies. In the pump example, that can be as concrete as knowing which side of an electrical junction needs a second sensor.

Sources and reading notes

Scope of review. The accompanying source ledger covers all 197 files in the History and Philosophy of Science folder, including duplicate copies and nested guidebooks. Searchable pages were screened for relevant passages. Selected texts and case studies received closer reading, including passages from the introduction and every contribution to Cause and Chance. Image-only works received selected-page text recognition. The body of Carnap's legacy e-book could not be extracted. Historical primary passages were checked in local or public editions. SEP entries guide the treatment of disputed questions.

Aristotle. Posterior Analytics I.13. Physics II.3. Local plain-text editions. Posterior Analytics I.13

Descartes, René. Discourse on Method (1637), VI. Principles of Philosophy (1644), IV.203–206. Principles IV

Locke, John. An Essay Concerning Human Understanding (1690), II.xxi.1–4, II.xxvi.1–2 and IV.iii.9–14. Essay II

Hume, David. An Enquiry concerning Human Understanding (1748), IV–VII. Hume Texts Online

Kant, Immanuel. Critique of Pure Reason (1781/1787), Second Analogy, A189/B232 onward. See also the SEP historical entry below.

James, William. The Principles of Psychology (1890), I, ch. XIV, local PDF p. 365. A Pluralistic Universe (1909), appendix B, “The Experience of Activity.” Appendix B

Whewell, William. The Philosophy of the Inductive Sciences (1840). Mill, John Stuart. A System of Logic (1843), III.viii and III.x. Mill, book III

Snow, John. “Cholera and the Water Supply in the South Districts of London in 1854.” Water-supply inquiry

Russell, Bertrand. “On the Notion of Cause” (1912–13). Fisher, R. A. The Design of Experiments (1935), ch. II. Hempel, C. G., and P. Oppenheim. “Studies in the Logic of Explanation” (1948). Russell’s paper

Reichenbach, Hans. The Direction of Time (1956). Suppes, Patrick. A Probabilistic Theory of Causality (1970). Cartwright, Nancy. “Causal Laws and Effective Strategies” (1979). Consulted through the SEP’s comparative discussion. Probabilistic Causation

Dowe, Phil, and Paul Noordhof, eds. Cause and Chance: Causation in an Indeterministic World (Routledge, 2004). Chapter citations in section IV use printed pagination. Volume record

Pearl, Judea. Causality (2000). Woodward, James. Making Things Happen (2003). Hernán, Miguel A., and James M. Robins. Causal Inference: What If (2020), chs. 1–3. What If

Pearl, Judea. “Causes of Effects and Effects of Causes.” Sociological Methods & Research 44 (2015), 149–164. The competing mixtures and circuit examples in this study are original illustrations. Author’s version

Christie, Maureen. The Ozone Layer: A Philosophy of Science Perspective (2001), especially pp. 140–143. Weber, Marcel. Philosophy of Experimental Biology (2005), pp. 281–283. Hacking, Ian. Representing and Intervening (1983), local text on experiment and electron manipulation.

Trout, J. D. “Scientific Explanation and the Sense of Understanding.” Philosophy of Science 69 (2002), 212–233. Tversky, Amos, and Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases” (1974), discussion of regression. Belousek, Darrin W. “Bell’s Theorem, Non-Separability and Space-Time Individuation in Quantum Mechanics,” local PDF p. 9.

Hall, Ned. “Two Concepts of Causation” (2004). Compare the stronger metaphysical thesis with the evidential distinction in §VIII. Author’s text

Halpern, Joseph Y., and Christopher Hitchcock. “Graded Causation and Defaults.” Compare normal alternatives and physically available backups. Author’s version

Cartwright, Nancy. “The Reality of Causes in a World of Instrumental Laws,” reprinted in R. Boyd, P. Gasper, and J. D. Trout, eds., The Philosophy of Science (1991), pp. 379–386. Publisher record

Burian, Richard M. “More than a Marriage of Convenience” (1977). Kantorovich, Aharon. “Philosophy of Science: From Justification to Explanation” (1988). Friedman, Michael. “Kant, Kuhn, and the Rationality of Science” (2002). Pinnick, Cassandra, and George Gale. “Philosophy of Science and History of Science: A Troubling Interaction.” Local papers in the accompanying ledger.

Stanford Encyclopedia of Philosophy. The Metaphysics of Causation

Stanford Encyclopedia of Philosophy. Causal Models

Stanford Encyclopedia of Philosophy. Counterfactual Theories of Causation

Stanford Encyclopedia of Philosophy. Causal Approaches to Scientific Explanation

Stanford Encyclopedia of Philosophy. Causation and Manipulability

Stanford Encyclopedia of Philosophy. Simpson’s Paradox

Stanford Encyclopedia of Philosophy. Kant and Hume on Causality

Stanford Encyclopedia of Philosophy. John Locke

Stanford Encyclopedia of Philosophy. William James

Stanford Encyclopedia of Philosophy. Aristotle on Causality

Stanford Encyclopedia of Philosophy. William Whewell

Stanford Encyclopedia of Philosophy. Carl Hempel