# 43 Causation: Objective or Subjective?

**Wolfgang Spohn** (University of Konstanz)


## Abstract

We, and scientific practice, tend to conceive of causation as an objective relation characterizing the external world. Philosophy has been more ambiguous. This chapter intends to renew the doubts. If causation is only a model-relative notion and if causation is tightly entangled with notions that are best understood in a subject-relative way, then the objectivity of causation is at least undermined. The paper discusses these doubts and concludes that the objectivity of causation must not be presupposed, but must be constructively earned.


## 43.1 Causation: A Bunch of Attitudes

I am glad that philosophy's voice is to be represented in this volume as well—after all, Judea Pearl not only won the Turing Award but also the Lakatos Prize, a highly, if not the most highly renowned award in the philosophy of science—and I am honored that I am invited to contribute as a philosopher. However, philosophy is different; with its more distant view it is prone to have a more critical perspective. Indeed, I feel that this perspective is wanting nowadays.

After the eventual breakdown of positivism, behaviorism, and similar doctrines around 1960, great methodological uncertainty spread, and philosophy, or at least the philosophy of science, seemed much needed. This has thoroughly changed over the past 20 to 30 years. Not that the problems have been solved in a generally accepted way. Philosophy certainly has not solved them; to expect so would be a misunderstanding of the nature of philosophy. Rather, the natural and social sciences have consolidated. They are just no longer irritated. Solving foundational problems has little impact on scientific practice. And methodological problems

have shifted. Data is all-important, but it is also overwhelming, and so data mining, data analysis, machine learning, and computer science in general are the new methodological aids. Philosophical aid seems outmoded.

I can understand this development to some extent, but it is detrimental. I would like to exemplify this with causation, the "cement of the universe."[1] There is hardly any other notion that is of such universal scientific importance than that of causation. Sciences struggle with it every day. It is thus useful to take again the more distant, 2,500-year-old perspective of philosophy. This is not philosophy's private perspective. Rather, almost every cognitive enterprise used to run under the label of philosophy. It's the common heritage of all sciences. Today, though, it's only philosophy that cultivates this heritage. And it is worth doing so. (Of course, if philosophy would do only this, it would be doomed.)

Aristotle, the first and still most embracive universal scientist, distinguished four notions of causes, which would be better called grounds nowadays. A thousand years later, one of them, the notion of efficacious cause—that's our modern notion of causation—took center stage. However, it remained under almost complete theological control for another thousand years. Allah or God is the sole or the ultimate cause of everything. And who would dare question Allah's or God's ways?

The modern discussion starts with David Hume's *Treatise of Human Nature, Vol. I*, in 1739, which he wrote at the age of 28. Well, he did not merely start it; he prepared the entire playing ground on which we still move today. Of course, this is a forbiddingly rough summary.[2] The all-importance of Hume, though, cannot be understated. He confusingly offered two definitions of causation. One, the *regularity view*, is highlighted as an advent of science, although it is recognized as insufficient. The other one is not the counterfactual view, as Lewis [1973b] and his readers, including Pearl [2000, p. 238, 2018, p. 20] state. It is rather what I like to call the *associationist view*. It seems repressed nowadays. According to it, causation is in the eye of the beholder, a habit of thought. This is a gross oddity, it is natural to discard it as a misunderstanding. Hume himself says about it: "I am sensible that of all paradoxes, which I have had, or shall hereafter have occasion to advance in the course of this treatise, the present one is the most violent" (1739, p. 166). In effect, he is so ambiguous about it that interpreters have puzzled over the relation of his two definitions till the present day (see, e.g., Beebee [2006]).

1. This is the phrase of Hume [1740]. More precisely, he says that the "principles of association ... *resemblance* ... *contiguity* ... *causation*... are really to us the cement of the universe." So, actually, and interestingly, he is talking about "epistemic cement."

2. The epilogue of Pearl [2000] gives a much longer, but still brief and very entertaining overview.

"A habit of thought," this sounds so understated. In the more elevated German way, Kant [1781] turned this into a *pure category of thought*. What a label. The idea, though, is basically the same as Hume's. Causation is a relation we impose on the world. It is not a notion we acquire from experience, not an idea of sensation, but rather *an idea of reflection*, in Hume's words—although, of course, experience is required to learn how the relation realizes. I think Kant is right. However, I will not use this chapter to positively defend this claim. The only aim I am pursuing in this chapter is to create some awareness of the fact that contemporary theories of causation are not safe at all from being infected by these old and important ideas about causation.

For, what is the contemporary attitude toward causation? In the positivistic times mentioned above, causation was a shunned notion, bad metaphysics, not imposed by our mind—this would be preposterous—but also not to be found in the world. This changed with Hempel and Oppenheim's [1948] theory of deductive-nomological explanation, which was, *in nuce*, Hume's regularity theory of causation. The irony was: there was no causation in that theory, as became clear about 15 years later. But the ban was broken, and causation is continuously among the hottest topic in the philosophy of science up to the present day. It was not so different in the various sciences, although each has its own speed. Pearl [2018] tells impressive stories about how obstinate the community of statisticians was and still is.

So, the importance of causation is acknowledged almost everywhere now. The natural and social sciences came up with a really surprising variety of ideas and conceptions. If you study them, it's hard to believe that they all talk about the same thing. The goal was to have specific and useful accounts, not just sublime philosophy. However, the matter turned out very difficult, and the ideas were quite idiosyncratic and tentative.

The field is still scattered. However, a certain paradigm emerged around thirty years ago, which by now seems to be the dominating one, sharing wide agreement and applicability. I am referring to the *interventionist theory of causal Bayes nets*, the cornerstones of which are Pearl [1988, 2000], Spirtes et al. [1993], and Woodward [2003]; it was substantially adumbrated, though, in Spohn [1978, 1980]. The three are by no means identical; there is quite a lot of divergence in detail. Still, it is legitimate to subsume them under one broadly conceived heading. And there is no doubt that no one did more than Judea Pearl to familiarize other disciplines with this doctrine and to convince them of its wide applicability—perhaps because as an AI researcher he is closer to the needs of the sciences.

This is tremendous progress and unprecedented success. However, when it comes to the nature of this doctrine, its contenders are surprisingly silent, Judea

Pearl included. Unlike many predecessors, starting with Hume, they don't try to define causation. This may be plausible. There must be some basic concepts, and then causation is likely to be one. Glymour [2004] emphasizes the liberating effect of this move. Similarly, Pearl [2018, p. 27]: His approach, which he attributes to Alan Turing, "is exceptionally fruitful when we are talking about causality because it bypasses long and unproductive discussions of what exactly causality is and focuses instead on the concrete and answerable question 'What can a causal reasoner do?"' However, this strategy does not avoid conceptual issues. If not definable, causation is at least closely related to other basic notions and thus at least infected by their character, as we will see below.

Instead, the main interest was to build causal models, to study their behavior, and to say how they can be tested. This was explored in great constructive detail; only thereby could wide applicability be acquired. What does this procedure leave to be desired? The background ideology certainly is that there are sort of objectively true causal models. This much seems to be tacitly understood, even if one is modest in claiming truth for the models one entertains. And the account of Pearl and others is the best way to get on to the track of the true models. Pearl [2018] does not explicitly speak of true causal models, but he explains the many inferences causal models allow, provided—that's a recurring phrase—"your causal model accurately reflects the real world" (p. 335).

Whenever I talk to scientists, this seems to be their common attitude as well. Of course, causation is an objective feature of the world, and science is there to uncover it. Anything else would undermine the self-conception of science as truth-seeking. And now we finally have a grip on how to do it.

Really? Are Hume and Kant thereby refuted? And the positivists defeated? I would like to cast doubt on this attitude. The objectivity of the notion of causation is not guaranteed at all. A crucial quote from Pearl [2018, p. 21] is: "If I could sum up the message of this book in one pithy phrase, it would be that you are smarter than your data. Data do not understand causes and effects; humans do." But what is it that makes us smarter than the data? The answer Pearl [2018] unfolds is that it is the second and the third rung of his so-called ladder of causation, acting/intervening and imagining the counterfactual. Maybe, though, we are smarter not because of being able to represent more objective truth than the data, but because we are able to add something to the data?

As said, I do not want to defend an answer to this question. But I want to suggest that objectivity must not be simply assumed and is not so easily earned. Subjectivity creeps in from at least two directions, which I want to briefly discuss in this paper. One point is the model relativity of the notion of causation, and the other is the potential subject-relativity of the notions with which causation is at least

intrinsically related. This does not yet confirm Hume or Kant; but it shows that matters are less clear than scientists wish.

## 43.2 The Model Relativity of Causation

Even if there is no general agreement, we have a fairly good conception of what causal models are and how they behave. Thus, we know what causation is, what the causal relations are *within* those models: they are either directly given by the arrows between the nodes or the variables of a causal graph, or they consist in certain probabilistic conditional dependencies among those variables, or they lie in the structural equations relating those variables, and so on. This is our grasp of causation. It is, however, only a model-relative grasp. Is causation hence a *model-relative* notion?

I observe a profound ambiguity concerning this question. On the one hand, I sense an implicit inclination toward model relativity, although it's hard to find it explicitly endorsed. Perhaps I get this sense because people are only dealing with causal models; this is the only frame within which they talk of causation. On the other hand, this attitude clearly won't do. Causation can't be *only* a model-relative notion.

Compare this with the notion of truth, another notion of utter fundamentality. There, Tarski has provided us with the model-theoretic notion of truth, of what it means that a sentence is true in a model.[3] Thereby we have gained a rigorous grip on truth theory, for the first time in history. However, this can't be the full truth about truth. We also have a notion of absolute truth or truth *simpliciter*. "The sun is shining." That's true—full stop (when I am writing this sentence). Relatively speaking, it's only true in one model and false in another (and doesn't get any truth value in a third). So, what's absolute truth? One is tempted to say: absolute truth is truth relative to the true model. But that's blatantly circular.

I won't try to resolve this predicament of truth theory; it's a serious problem. However, the analogy is illuminating. Clearly, acquiescing in the model relativity of causation would introduce an intolerable amount of subjectivity. Causal relations cannot be this way or that way, depending on the causal model we choose; we cannot have it both ways. This would undermine scientific objectivity. Thus, at least implicitly scientists presuppose an *absolute* notion of causation, and this is what their modeling activities try to capture. What is it?

3. Sometimes, people speak only of truth in an interpretation. In any case, this notion of a model differs from, and is much more general than, the notion of a model used in the theory of causation.

The literature, not only in the applied sciences but also in philosophy, hardly comments on this question. To be honest, I find this shocking. Apparently, the question is not really relevant. Somehow, the causal models fit better or worse; so, we know in which direction to improve our causal models; thereby we approach the true causal model; and that's the one that grasps causation in the absolute sense. Thus, there is really no more than the model-relative notion of causation, amended only by the notion of fit or truth of a model. This seems to be the general attitude, and it is certainly the one displayed in Pearl [2018].

I have offered two terms here, "fit" and "truth." "Fit" sounds more cautious, perhaps this is all scientists expect of causal models. However, they cannot waive truth. They may be modest in not claiming to possess the truth. Still, truth must be their guiding aim. Hence, the present discussion is really about the truth of causal models. The general attitude parallels the blatantly circular answer in the case of truth theory. For causation, however, it does not sound circular. But is it any better?

I don't think so. One needs to understand that the truth of a causal model is never a relation between the model as such and reality, as it were. A causal model may fit the data very well, and then there is no reason for suspicion. It may even overfit the data. Often, though, the fit is not so good. One may then adapt the parameters of the model or take similar moves, without essentially changing the model. Usually this won't do, however. Criticism of the model takes the form of an enlarged model accounting for more variables. Almost all discussions are about neglected variables which disturb the picture in one or the other way and because of which all those partial regression and correlation coefficients are misleading. There are common causes, there are confounding variables and selection variables, there are unmeasured and latent variables, Simpson's paradox lurks more often than expected, and so on. By making those neglected variables explicit in the enlarged model one may reach a better fit. This is *always* a possibility, even if the original model fits very well and does not raise suspicion. Surprises can never be excluded.

Of course, practicing scientists rarely aim for perfect models. There is always some slack between the model and the data. Scientists are content when they can be confident to have identified the main causes. They would admit that there always are a lot of further causes blurring the picture. But if they blur it only a little bit, we need not worry. One must always think where to spend one's efforts, and to explore those residual causes may not be worth the efforts. Again, though, surprises can never be excluded. So, this attitude of the practicing scientist only confirms the fact that the truth rather lies in an enlarged model.

Note, however, where this takes us. Isn't this to say that a causal model is true if it is part of an enlarged model, not any enlarged model, of course, but a true enlarged model? And now we are caught again, not in a circularity, but in an infinite

regress. No model is large enough to decide about the truth. We are deferred in the end to what may be called the universal model containing all variables whatsoever, so that no variable is neglected, no further confounding or otherwise disturbing variable can turn up. Surely, though, that's completely ill-defined speech. The universal model is at best a fictitious ideal of which we have no more than the faintest grasp.

For this reason, I am claiming since Spohn [2001] that causation in the intended absolute sense is a model-transcendent notion. Limited causal models do intend it, but whether they grasp it cannot be decided by any other limited model, however enlarged. I want to briefly indicate that this model-transcendence transpires through all of the current theories of causation.

Spirtes et al. [1993, pp. 44f.], for example, make very clear that their basic causal axioms, the causal Markov, minimality, and faithfulness condition apply only to causally sufficient causal models—where, roughly, a model is causally sufficient if it contains all common causes of any two variables in the model. Clearly, if this is taken literally, this means that any causal model must contain the Big Bang, which surely is a remote common cause of any earthly matters. Fairness requires to say that Spirtes et al. have done a lot to weaken this presupposition by exploring how much we can still infer about causal relations in its absence; see, in particular, their second edition.

Woodward [2003] perfectly displays the interventionist agenda on causation. However, if one looks closely, his notion of intervention is model-transcendent, too. If we intervene on the variable $X$ in order to find out whether it is a cause of the variable $Y$, he requires the intervention on $X$ to be statistically independent of any variable that causes $Y$ along some causal path that does not go through $X$; this is condition I4 in Woodward [2003, p. 98]. Here, "any variable" must be taken as quantifying not only about the variables in the causal model but also about all variables outside the model. This is his way of model-transcendence.

The same remarks apply to Pearl's *do*-operator. The model-immanent function of $do(X)$ is to causally separate the variable $X$ from all its causal predecessors in the model. However, this separation is to hold for any enlarged model as well. That is, although $do(X)$ is explained by Pearl as just another variable with a special behavior within causal models, it really has a model-transcendent function. Or in more general words: The truth claim of any causal model always carries the implication—"and there are no further neglected variables, confounding or otherwise, which change the causal picture." This is clearly a model-transcendent implication.[4]

4. This is not to say that by using the relative notion of causation we are bound to make the closed-world assumption (see, e.g., Pearl [2000, pp. 252f].). As such, relative causation is just causation

Where does this leave us? We are not forced to acquiesce in the model relativity of causation. We can get rid of it, but not in the way commonly assumed. The truth of a causal model is not a local affair that could be locally settled. Rather, we are referred to ever larger causal models, but nothing is ever settled due to the model-transcendence of absolute causation. So, in a way, we indeed deal only with model-relative causation; it's always more of the same in ever larger models. But it is important to be clear on what we are up to with causal models, to be clear about what truth could mean for them.

## 43.3 Laws

We may thus have banned the subjectivity entering through the model relativity of causation, though in a somewhat unexpected way. Let me turn, hence, to the other potential source of subjectivity, the nature of the concepts with which causation is closely connected, even if one should have given up defining causation by them. When one surveys theories of causation, the connection always refers to one of two kinds of concepts, either to something like regularities, laws, structural equations etc., or to probability, which is the central notion in all statistical contexts.

Of course, causation is essentially connected to still further notions: action (this relation is perhaps sufficiently reflected in interventionist theories of causation), order (if this is explicated as entropy, one may subsume this under the probabilistic connection), and most importantly, space and time. For physicists this relation is absolutely central. In the social sciences it is often marginal. Surely, if we are to model climate change or the proliferation of a pandemic, space and time are indispensable categories. Often, though, these categories do not even play an implicit role. The reason is clear. There is often no temporal order in the data and hence none in the causal model representing the data.[5] Still, I am wondering how one can ever do causal theorizing while neglecting its first axiom, namely that causes temporally precede their effects.

In the present context, however, we may neglect these other connections because they do not endanger the objectivity of causation. Let me therefore focus on the two connections initially mentioned and first on laws and its ilk. The law connection is the one originally claimed by Hume's regularity of causation. If it would be appropriate, it would bar subjectivity. However, it is not appropriate, for various reasons.

---

relative to the model. Only when we claim that model-relative causation amounts to absolute causation do we claim the closed-world assumption to be true.

5. See also Pearl [2000, section 7.5.1] for a discussion of this point.

First, Hume was not so sophisticated to distinguish between accidental regularities and genuine laws. Certainly, only the latter create causal relations. Although the natural sciences take it to be clear what they are after when they are after laws, I can only warn the reader to enter the philosophical discussion about what laws are; it's a quagmire.[6] One certainly finds various opinions giving up on the objectivity of laws and thus of causation.

However, that's presently only a side issue. There are two more important concerns. One concern is that laws by themselves cannot tell about causal relations. As has often been observed, the counterfactuals describing causal relations really are *counternomologicals*; they refer to which laws still hold when some laws are broken. The laws can never tell this by themselves. Let's consider a simple example and assume that the co-occurrence of falling air pressure, the falling of the barometer, and a thunderstorm were sort of a strict law. Now we break the law between air pressure and barometer by manipulating the barometer. The question determining the causal relations then is: which law still holds, that between barometer and thunderstorm, or that between air pressure and thunderstorm? Not both can still hold. The answer is obvious to us. The point is only that the question and the answer are counternomological ones. The example makes clear that interventions are also invoking counternomologicals; they introduce small miracles, in the terminology of Lewis [1973a, pp. 75ff.].

The next question is: what governs those counternomologicals? The answer is not clear at all. Perhaps a similarity ordering à la Lewis [1973a] does the trick, perhaps some epistemic entrenchment order is working in the background (see, e.g., [Gärdenfors 1988, chapter 4]). Something of this sort is required. However, the objectivity of all these auxiliary notions is at least doubtful. I don't want to say that they are hopelessly subjective. Prima facie, though, they do look subjective. We might reach intersubjective agreement concerning similarity, entrenchment, or whatever, though we would have to study on which grounds we can do so. Possibly we can even claim some kind of objectivity for our agreement in the end; but again the question would be on which grounds we are able to do so.

What I want to emphasize: The issues I am raising here are not issues of the ordinary scientists. They usually proceed from a tacit understanding and agreement. However, if they reach agreement, it's not due to collectively grasping what is objectively there. It's *not* like: "Why does (almost) everybody say that $2 + 3$ is 5? Because $2 + 3$ *is* 5." Rather, agreement comes about in some other way. And if it can claim objectivity, it is not the objectivity of ordinary facts. Our dealings with similarity or entrenchment orders and the like are not a scientific but an epistemological issue

6. If you want to disregard my warning, you may start with van Fraassen [1989].

that requires a different sort of study. The point then is: if causation is essentially entangled not just with laws but with all this additional machinery of doubtful objectivity, then causation is deeply infected by this machinery as well. The objectivity of causation cannot be presupposed, but must be earned and constructed in the way required for this machinery.

The other important concern is that we all have this noble ideal of a natural law, allegedly explored in basic physics. But of course, the laws investigated in the social sciences, economics, geology, biology, even in most parts of physics, and so on, are *not* of that ideal kind at all. The ideal is very misleading. Rather, they somehow are soft, non-strict laws; they are, as we say today, *ceteris paribus* (*cp*) *laws*. A simple physical example is Hooke's law about the proportionality of the extension of a spring and the weight attached to it, which, of course, has countless exceptions. To be sure, all structural equations are of the same kind, wherever they are formulated. This fact is certainly clear to the working scientists, even though they may not have fathomed its epistemological implications. For philosophers of science the insight came quite late; too long were they attached to the ideal. But once they started thinking about them, cp laws turned out to be an utter mystery (see, e.g., Reutlinger et al. [2019]).

Look at "cp, (all) *F*s are *G*" (e.g., "cp, birds fly," or "cp, prices go up, when demand rises"). What does this claim? How must the world look like for this to be true? It's very unclear. Polemically, one might say that it doesn't claim anything at all; it simply says: "all *F*s are *G*, unless they aren't." This is unfair; scientists don't claim platitudes. But it is very hard to avoid this unintended answer. Another reply is that cp laws are statistical laws. Judea Pearl seems to tend to this answer.[7] Most people, however, would reject it. Hooke's law is not a tacitly statistical law about the manufacture and use of springs. Our schematic law doesn't say "most or 99% *F*s are *G*." It rather says "normally or typically, *F*s are *G*." And normality or typicality is not just a matter of proportions. But what is it?

The core problem is that we slip into a similar open-ended situation as we did with causal models in the previous section. We might start with saying: "cp, *F*s are *G*" means "under normal conditions, *F*s are *G*," leaving the task to specify the normal (and the exceptional) conditions. Maybe we can confirm good hypotheses: "whenever normal conditions *N* hold, *F*s are *G*," and "whenever *E* (= not-*N*), *F*s are not *G*." But of course, these hypotheses are not literally true. They are cp laws in turn, and we will find further exceptional conditions $E'$ and $E''$ such that:

7. In Pearl [2000], he explains right on pp. 1f why he turns to probabilities. One reason he gives is that "causal expressions in natural language are subject to exceptions," as are cp laws, and that "probability theory" is "especially equipped to tolerate unexplicated exceptions."

"whenever $N$ and $E'$, $F$s aren't $G$, either" and "whenever $E$ and $E''$, $F$s are $G$, after all." And so on. With a little phantasy, you can easily take three or four rounds of this game with my sample laws. This process is non-monotonic, as logicians say; strengthening the conditions may always reverse the law. And, in analogy to causal models, the process is open-ended; you are never in a position to say: "Now I have exhausted all conditions under which $F$s are, or are not, $G$."[8]

The upshot is that by claiming a cp law we do not make a claim with a plain truth condition to be ascertained or confirmed in familiar ways. The dialectics of normal and exceptional conditions is a different epistemic game. Of course, it is legitimate to play this game; it's the way of science. However, its rules are quite unclear. It's not an ordinary search for truth. How could it be, when the claims made are qualified by cp clauses and thereby lose a plain truth condition? Again, it seems to be the task of the epistemologist to clarify the matter and to find out about the underlying methodology of this cp science.

It is not so clear what the epistemologist will find. To be sure, nothing is objectively normal or exceptional. Normality is, to put it vaguely, an anthropocentric notion. So, subjectivity lurks again. In scientific contexts we can perhaps restrict the notion of normality to its epistemic uses.[9] But even in this case it is basically subjective, and we must again find a different explanation for reaching consensus than that the consensus agrees on objective truth.

Thus, my point is the same as above: If causation is closely entangled with soft cp laws, then it is also entangled with this non-objectivity, with this absence of truth conditions just observed. If so, the objectivity of causation can again not be presupposed. Rather we have to study, by studying the epistemology of cp science, how causation may, perhaps only partially, acquire objectivity.

Where do we stand? If we should have hoped to somehow anchor causation in objective lawhood, this has ended in disappointment; cp laws are not the kind of laws to satisfy our idea of objectivity. And even if they were, laws only would not do; they would have to be amended by some machinery answering counternomological questions. Maybe, though, we can avoid this muddle by taking the turn that most sciences have taken, anyway. Maybe we can avoid all reference to laws and the like and instead look at the connection between causation and probability.[10]

8. As indicated in the previous footnote, this analogy is one motive for Pearl to resort to probabilities. For Woodward [2002] it is a reason to try to analyze cp claims as causal claims. Either way, the problems I am about to display persist.

9. In Spohn [2014] I have tried to explicate this epistemic use in terms of ranking theory, which, I argue, is ideally suited for this job.

10. Van Fraassen [1989] is not about causation. However, it is precisely the probabilistic turn that he propagates there in order to escape the muddle of laws.

This may look promising. However, I would like to indicate in the rest of this paper that we thereby move out of the frying pan into the fire.

## 43.4 Probability

When one looks at contemporary causal theorizing, one is overwhelmed by its probabilistic character. A hundred years ago, this was unthinkable. Causality was firmly tied to deterministic theorizing. Causes were mostly conceived as necessary and sufficient causes. Things changed when physics turned out to be irreducibly probabilistic. At first, it seemed that we had entirely lost causation in the physical realm. But then it became ever clearer that probabilistic causation makes good sense as well. Nowadays we find this attitude also in all of the social sciences. Our data is probabilistic, and when we hope to find causal relations in it, it can only be in the form of probabilistic tendencies. So, it's not surprising that probability now is the key notion with which causality is wedded.[11]

However, what do we mean by probability? In philosophy we discuss several different interpretations—five? or more?[12] It should make a big difference for our understanding of causation with which of these interpretations it is connected. Again, I am surprised how little this is discussed in the relevant philosophical and scientific literature. Is it not important? Is it clear, anyway?

Well, whatever the other interpretations may be, the social sciences (medicine, etc., always included) obviously speak of *statistical probabilities*. This appears to be taken as the only relevant interpretation. However, do we know what statistical probabilities are? Did we check whether they are suitable for connecting up with causation? Again, the literature appears to take this as settled. Let me approach these questions by first briefly explaining how rich and unclear the concept of probability is despite the fact that its mathematical structure is unequivocally fixed.[13]

The clearest interpretation is the *subjective* or *Bayesian* one. According to it, probabilities are rational degrees of belief. There are a lot of arguments why rational degrees of belief must take the form of probabilities. We may leave it open how cogent these arguments are and whether there might be other reasonable

11. What *is* surprising is the far-reaching marginalization of deterministic causation. I find it very unlikely that we searched for a chimera for 2,000 years.

12. Galavotti [2005] and Gillies [2000] are two very commendable presentations of this confusing field.

13. Well, almost. There is some uncertainty concerning σ-additivity and concerning the representation of conditional probabilities via Renyi and Popper measures. This need not worry us here.

conceptions of degrees of belief.[14] There is no doubt, though, that probabilities are by far the most familiar conception of these degrees. Subjectivists, Bruno de Finetti ahead, claim that this is indeed the only intelligible interpretation of probability. However, we need not go so far; it suffices to say that it is at least one good and reasonable interpretation.

I was surprised to read in Pearl [2000], right on p. 2: "We will adhere to the Bayesian interpretation of probability, according to which probabilities encode degrees of belief about events in the world and data are used to strengthen, update, or weaken those degrees of belief." This is a resolute Bayesian avowal extending over the entire book. However, I suspect that he is not consistent in that avowal. The book's later parts are about statistics, and statistics don't refer to single events in the world, as I will point out below.[15]

And the avowal betrays the quest for objectivity. There are no true subjective probabilities. They can and should be well-informed by the data; but then they can change to being even better informed. They might be called true if they conform to objective probabilities. However, this idea is highly problematic. One reason is that objective probabilities themselves are highly problematic, as we shall see. Another reason is that there is more to know about a fact than its objective probability (if it has one), for instance, the fact itself. Thus, perhaps, only a probability of 1 for the fact can be called true?[16] I conclude: we better abstain from calling subjective probabilities true or false or taking them as representing reality.

For causation this entails that there are nothing but causal beliefs, which may be more or less well-informed, which, however, cannot be called true. They do *not* represent any causal reality. This runs counter to the general attitude we meet in the sciences. And it seems to run counter to Pearl's own attitude that I have quoted in Section 43.1. Time and again, he slips into realistic talk, from causal beliefs to belief in causal facts. However, within the Bayesian interpretation this is an illegitimate

14. For decades I have been propagating ranking theory as another model of degrees of belief. Not the least of my reasons is that ranking theory allows to state a theory of deterministic causation (which speaks of causes making their effects possible or necessary) in close parallel to probabilistic theories (which speak of causes making their effects more probable). See Spohn [2012], in particular chapter 14.

15. For instance, Pearl [2000, section 7.5.4] discusses singular versus general causes. But his discussion refers to statistical probabilities concerning populations and not to subjective probabilities about single events.

16. Neither does it help to say that the proper probability is true only before the fact in question realizes, and probability 1 is true only after the fact. This would make truth time-dependent in unwanted ways.

move.[17] So, it seems we should attempt to avoid the Bayesian interpretation in our context.

But beware. Whenever we get into trouble with other interpretations, the subjective interpretation is the only one that always works, that makes sense in every application. We can always resort to making assertions only about our and other people's (causal) beliefs. So, whenever a fallback position is needed, we may well be forced back into the subjective interpretation.

Let me add a few remarks about *Bayesian statistics*. Bayesian statisticians, or Bayesians for short, are not subjectivists; they certainly grant objective probabilities in some sense. They only claim that in doing statistics we need to consider our prior assumptions as well, represented by our subjective probabilities. The traditional Neyman–Pearson school hopes to do without these subjective elements. What sounds like a principled disagreement—it indeed is—apparently turns into a fair cooperation in practice.

However, I think Bayesians have a delicate standing in our present context.[18] It won't do for Bayesians just to use subjective as well as objective probabilities. For, each probability measure must have a uniform interpretation; one cannot mix different interpretations within one measure. So, the Bayesian needs to assume bridge principles translating between objective and subjective probabilities. Such principles are not hard to come by. For instance, if all I know about a given event is that its objective probability is $x$, my subjective probability for that event should obviously also be $x$. By introducing his so-called Principal Principle, Lewis [1980] has initiated a big philosophical discussion about such bridge principles; they may need generalization and modification.

This is, however, not a satisfactory rescue for the Bayesian. One problem is that not all kinds of objective probability are equally suited for such bridge principles. We shall see below that the so-called statistical probabilities are indeed ill-suited. Another problem is that we cannot turn all probabilities the Bayesian refers to into objective ones. The required uniform interpretation of probabilities can only be a

17. More generally, the tendency to slip from conditional belief to belief in conditional propositions is ubiquitous. This step is so easy. However, the move hides all ambiguities between epistemic and realistic world conceptions. It is not innocuous at all. Stalnaker [1984, chapters 6 and 7] is a paradigmatic, but in my view unsuccessful, struggle with what is going on in this move.

18. Pearl [2018, p. 90] complains that "Bayesian subjectivity in mainstream statistics did nothing to help the acceptance of causal subjectivity." The latter means for Pearl that each causal inquiry must start with positing a subjective causal model and must grant the possibility that data may not decide between two different causal models subjectively posited. This sharply differs from the subjectivity I am discussing here.

subjective one. This point then extends to his account of causation, which must be similarly subjective. Thus, we are back at the position above which Pearl avowed, but which we might want to avoid.

Let's turn, hence, to *objective* probabilities. Here, interpretational variance starts. Still, there is a common anchor. Everybody agrees that probabilities somehow ground in relative frequencies; that's their connection to reality. However, this even holds for the Bayesian interpretation; of course, well-informed subjective probabilities listen, and are usually close, to observed frequencies. In particular, though, it holds for all objective probabilities. This grounding is spelled out in the fundamental law of large numbers, proved by Jacob Bernoulli already in 1689 and called the "Golden Theorem" by him. It guarantees that the relative frequencies in infinite independent repetitions converge to the single-case probabilities—though only in a probabilistic sense. This means that some notion of probability is already presupposed by the law of large numbers, and it says then how those probabilities probably manifest in frequencies.

*Frequentism*, which has been very popular among working probability theorists, wants to turn around the relation. It is the doctrine that probability is *defined* as the limit of relative frequency in random sequences, where random sequences are subject to further qualifications, most notably complexity-theoretic ones. Frequentism's crucial problem, with no good answer to date, is that it applies only to infinite sequences, strictly speaking. It cannot be employed for the single case, which is the one we are interested in. We want to know the probabilities governing the next throw of the coin, and this is about the next throw, not about an infinity of throws. Thus, frequentism is *not* supported by the law of large numbers, which already presupposes those single-case probabilities.

Or to address our present concern: Suppose we could isolate a causal system modeled with its probabilities, what does it teach about the causal relations, if we run the system very often and speculate about the limiting frequencies? These relations are in the system, and they are somehow connected with the probabilities in the system and not with the frequencies in the repetitions. With respect to causation, too, we need a notion of objective probability that applies to the single case.

There is such a notion that serves our purpose; philosophers call it the *propensity interpretation*. According to it, the objective probability of, for example, a die showing a 6, is something attributable to the die as such, an intrinsic feature of the die and its set-up, the throwing device, a disposition that can only be described probabilistically and not deterministically, viz. a propensity. The single-case propensity is basic, but it entails, of course, a long-run propensity, which

converges as described by the law of large numbers. A die or a roulette wheel are already good examples, although one may argue about whether they "really" are deterministic devices.

The ultimate examples can be found in quantum and nuclear physics. Radium atoms, for example, have a propensity to decay. We could say that there are many different kinds of radium atoms, each with a different deterministic decay time. Determinism saved. However, this would make no sense at all, since there is no way to tell the kinds apart; it would be a distinction without a difference. Hence, it is much more reasonable to say that all radium atoms have the *same* irreducibly probabilistic propensity to decay governed by an exponential distribution. This is a genuine statistical law: all objects of a certain kind show the same stochastic behavior.

The decay propensity of radium atoms is not immutable. It can change. For instance, we can excite the nucleus by various kinds of radiation and thereby accelerate its decay in various ways. We may set up causal models representing these propensities and their potential changes. Such a model would describe genuine probabilistic causal laws applying to each single case in the same way. Understanding probabilities objectively in this way would thus allow us understanding of causality in the same objective way.

This is what we were looking for. However, the crucial point for the rest of the paper is: Success does not extend; propensity is *not* the kind of probability referred to in most applications of causal models discussed in the literature. These applications belong to the social sciences, medicine, epidemiology, for example, and the statistical probabilities they refer to are not propensities as just described. Let me explain.

Many of those models are a matter of life and death. I certainly have a deterministic propensity to die sometime. But it does not make sense to speak of any propensity of mine to die before 80 or after 80. There are millions of potential causes of my death, most of which are not within my reach. The chance set-up in which my future death is located spreads more or less over the entire surface of the earth and further. The hope that a universal wave function could decide about this propensity would be nothing but a silly reductionist phantasy.

However, aren't there mortality tables? Sure. They don't tell anything, though, about my propensity. They tell how likely men in my age are to reach 80. But I am not an average man, nobody is. It is entirely unclear which specific subgroup would consist of the men relevantly similar to me, and if it were clear, there would definitely not exist any statistics for that subgroup. This is the well-known *problem of*

*the reference class*, which has no good answer.[19] It prevents transferring statistics to the single case; we can't statistically infer single-case propensities.

Certainly, though, we are inclined to reason as follows: If 60% of the men in my age group reach 80, and if you have no information about me that makes me in any way special, then your subjective probability for my reaching 80 should be 60%, too. This reasoning applies a kind of bridge principle relating objective and subjective probability, or perhaps relating only frequency and subjective probability. Presumably, we use this kind of reasoning, at least roughly, whenever we read a statistic.[20] But note that we thereby return to subjective probabilities about the single case, which always make sense. And note that the premise of the argument, the absence of special knowledge, is, strictly speaking, almost never satisfied. Usually, we do have special knowledge about a given single case, which we reasonably conjecture to be statistically relevant, a bit at least, even though we do not have a relevant statistic.

Note how different this is from my physical example. There we could legitimately assume single-case probabilities that entail the statistical behavior of large samples. Here we only have the statistical behavior of large samples without any underpinning by objective single-case probabilities; only shaky subjective inferences about the single case are feasible. This is a *world* of difference.

Of course, I have chosen a graphic example, the probability of death, something potentially caused in more ways than anything else. My point seems obvious in this example. However, the radium atom was an equally clear example for objective propensities. Where on the scale from the one to the other example do the propensities get lost? I do not know; it seems very hard to say. They do get lost through the multitude and the externality of the causes of the object's states to be probabilistically assessed. In view of this multitude what can we still attribute to the object itself? Already a person's propensity to recover from a certain disease after taking a certain drug is a very unclear case, I think. Moreover, the onus is not on me to say where propensities are lost. The onus is on the friend of objective probabilities to show that he is still legitimately speaking of them.

Perhaps, though, he is not speaking of propensities at all. The scientists fitting their causal models to statistical data refer to statistical probabilities; that's what they would say. Let's finally ask, then: what are *statistical probabilities*? Primarily,

19. Probably, the reference class relevant for me consists only of myself—not good for doing any statistics.

20. For an affirmative discussion of this statistical bridge principle, see Schurz [2019, pp. 57–68].

they are just relative frequencies in a given population, which behave like mathematical probabilities. But frequencies are not probabilities. Genuine probabilities enter only through the random mechanism by which individuals are selected from the population. If each individual has an equal chance (= objective probability = propensity) to be selected, then the chance that an individual with a certain feature is selected is the same as the relative frequency of that feature in the population. However, speaking of probabilities in this sense is only a roundabout way of speaking of the frequencies.

Usually, the procedure is the other way around. We cannot register the entire population; we can only observe a representative sample, which is selected by such a random mechanism. Then inferential statistics is needed to probabilistically infer the distribution of the features in the population from that in the sample. But note that these inferred probabilities are not objective probabilities for the shape of the population; they are subjective probabilities expressing our expectations about this shape. Of course, this does not mean that they are arbitrary. They proceed from an objective base in the representative selection mechanism by statistical inference. However, making a random selection from the population does not make the population itself in any way chancy.

Where does this leave us with respect to a causal model? It contains a set of variables that characterize the shape of the population, it contains causal arrows between those variables, and it contains many quantities that look like absolute and conditional probabilities. But these quantities are either observed frequencies in the sample or estimated frequencies in the population. And they confirm, or do not confirm, the causal arrows via the methods of causal inference. However, it must be clear that causes and effects in the model are nothing but relative frequencies in the population. By changing the relative frequencies for the cause variable one can change the relative frequencies of the effect variable. This is most useful information, for sure. But it is *this* kind of information and nothing else. And, I find, it makes the causal model appear quite mysterious because the causal story it delivers is a brute story about the population level without any underpinning from causal stories on the individual level.

This may appear as a very unfair presentation of what is going on in causal models. In particular, my claim about the missing underpinning from the individual level rests on my claim that it rarely makes sense in the applications in the social sciences to speak of individual propensities. I suspect that the general attitude rather is to simply *postulate* those individual propensities. We may not know much about them, and they may have considerable variance. But we do know that they generate the frequencies we observe in the samples or estimates for the population. Hence, what we observe and estimate is an (statistically qualified) average individual propensity.

For instance, if 60% of the men of my age reach 80, then the average propensity of men of my age to reach 80 is 60%. The individual propensities diverge in unknown ways, but they must (roughly) have this average. If a certain drug raises the recovery rate for a certain disease from 20% to 50% in a sample (or in the test group as compared with the control group in a randomized controlled trial) or probably in the population, then the average recovery propensity of those having the disease is raised by the drug from 20% to 50%. Again, the individual propensities will diverge, but they must (roughly) have this average. As stated above, however, any inference to those individual propensities almost inevitably results in subjective probabilities about the individual cases. I have more than 60% confidence to reach 80, and if a person recovers after taking the drug, this is perhaps not just because a 50% propensity has played out well.

This is in no way to question the great value of knowledge about average propensities (= observed or estimated frequencies) and about how to change these average propensities. However, what is the conceptual gain of this move? We now have a hypothetical individual underpinning of the population frequencies. This is indeed a causal underpinning by hypothetical causal stories about hypothetical individual propensities adding up to average propensities. However, we do not know much about that underpinning beyond the frequencies to which it leads. We do not have any statistical laws for the individual cases. And as explained, this underpinning is at best hypothetical and at worst meaningless.

Let me make clear once more what my dialectics on the previous pages was supposed to be. I started out saying that a subjective interpretation of probabilities can be applied everywhere. This would be fine—except that it does not satisfy our objectivistic intuitions concerning causation. This motivated the search for suitable, more objective interpretations that could save the objectivity of causation. This search was perhaps not entirely negative, but the objectivists can hardly be pleased by its weak and problematic results.

My general moral hence is: We must not presuppose the objectivity of causation and of the notions with which it is related. The safe fallback position is always the subjectivistic one; perhaps we should indeed start with Pearl's avowal of Bayesianism. And starting from there, we must work hard to earn and establish objectivity, without guarantee of success. I don't claim that Hume and Kant are thereby confirmed. However, I hope I have succeeded in pleading for more openness toward their doctrines.

This need not undermine the self-conception of scientists as truth seekers. It only suggests a more complicated picture of truth-seeking. Truth seeking is not just somehow adequately representing reality. It has much more to do with subjective belief, with intersubjective agreement, with rational belief and belief change, guided by principles of epistemic rationality, which must be agreed upon in turn.

All of this must be made explicit. When we do so, we may (have to) take recourse to another kind of objectivity, the objectivity of rationality. This is of a normative kind and as such delicate, contested, and not secured at all. It must be earned as well. This picture of science is more complicated, also more difficult to explain to the public than the simple picture of just objectively representing reality. In the end, though, it is a more honest picture.

## Acknowledgments

This research was supported by the DFG Cluster of Excellence "Machine Learning—New Perspectives for Science," EXC 2064/1, project number 390727645. I am most grateful to Judea Pearl for 30 years of partnership. We never did joint work, but we extensively worked at common themes, to my great benefit. Concerning this paper, I am indebted to Eric Raidl for the support and many helpful comments.

## References

H. Beebee. 2006. *Hume on Causation*. Routledge, London.

P. Gärdenfors. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA. DOI: https://dx.doi.org/10.2307/2275379.

M. C. Galavotti. 2005. *Philosophical Introduction to Probability*. CSLI Publications, Stanford.

D. Gillies. 2000. *Philosophical Theories of Probability*. Routledge, London.

C. Glymour. 2004. Critical notice on: James Woodward, making things happen. *Br. J. Phil. Sci.* 55, 779–790. DOI: https://doi.org/10.1093/bjps/55.4.779.

C. G. Hempel and P. Oppenheim. 1948. Studies in the logic of explanation. *Phil. Sci.* 15, 135–175. DOI: https://doi.org/10.1086/286983.

D. Hume. 1739. *A Treatise of Human Nature,* Vol. I: Of the Understanding. (Page numbers refer to the edition of L.A. Selby-Bigge, Oxford: Clarendon Press 1896.)

D. Hume. 1740. *An Abstract of 'A Treatise of Human Nature.'*

I. Kant. 1781. *Kritik der reinen Vernunft.*

D. K. Lewis. 1973a. *Counterfactuals*. Oxford, Blackwell. DOI: https://doi.org/10.2307/2273738.

D. K. Lewis. 1973b. Causation. *J. Phil.* 70, 556–567.

D. K. Lewis. 1980. A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability,* Vol. II. University of California Press, Berkeley, 263–293.

J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

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

J. Pearl and D. MacKenzie. 2018. *The Book of Why*: *The New Science of Cause and Effect*. Basic Books, New York.

A. Reutlinger, G. Schurz, and A. Hüttemann. 2019. *Ceteris paribus* laws. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/ceteris-paribus/.

G. Schurz. 2019. *Hume's Problem Solved: The Optimality of Meta-Induction*. MIT Press, Cambridge, MA.

P. Spirtes, C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search*. (2nd. ed.). Springer, Berlin. 2000, MIT Press, Cambridge, MA.

W. Spohn. 1978. *Grundlagen der Entscheidungstheorie*. Kronberg/Ts.: Scriptor, out of print, pdf-version: *http://www.uni-konstanz.de/FuF/Philo/Philosophie/Mitarbeiter/spohn.shtml*.

W. Spohn. 1980. Stochastic independence, causal independence, and shieldability. *J. Phil. Log.* 9, 73–99. DOI: https://doi.org/10.1007/BF00258078.

W. Spohn. 2001. Bayesian nets are all there is to causal dependence. In M. C. Galavotti, P. Suppes, and D. Costantini (Eds.), *Stochastic Dependence and Causality*. CSLI Publications, Stanford, 157–172.

W. Spohn. 2012. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford University Press, Oxford.

W. Spohn. 2014. The epistemic account of ceteris paribus conditions. *Eur. J. Philos. Sci.* 4, 385–408. DOI: https://doi.org/10.1007/s13194-014-0093-6.

R. C. Stalnaker. 1984. *Inquiry*. MIT Press, Cambridge, MA.

B. C. van Fraassen. 1989. *Laws and Symmetry*. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/0198248601.001.0001.

J. Woodward. 2002. There is no such thing as a *ceteris paribus* law. *Erkenntnis* 57, 303–328. DOI: https://doi.org/10.1023/A:1021578127039.

J. Woodward. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/0195155270.001.0001.