Comments on "Causal Inference without Counterfactuals" by A.P. Dawid

Glenn Shafer*

October 24, 1999

In recent years, a number of statisticians and computer scientists have suggested that casual reasoning requires that questions with hypotheses counter to fact have well-defined answers. Phil Dawid's elegant and insightful article is the first critical examination of this suggestion. As such, it is an essential contribution to the philosophy of probability and causality. It moves the discussion of causality in statistics to a new level of sophistication.

The article should prove an effective exercise in persuasion, because Dawid meets the proponents of counterfactuals on their own ground. He begins with the *counterfactual variables* Y_t and Y_c that appear in the models formulated by Neyman (1923), Rubin (1974, 1978), and Holland (1986), and he makes every effort to understand how much sense and how much use can be made of these variables.

Dawid's central theme is that counterfactuals should be held up to de Finetti's observability criterion. This criterion says that it is legitimate to assess a probability distribution for a quantity Y only if Y is observable at least in principle. On this criterion, it is legitimate to assess probabilities for $Y_t(u)$, because we can apply the treatment t to the unit u and then observe

*These comments benefited from research supported by NSF Grant SES-98199116. I have also benefited from discussions of causality with Phil Dawid over many years. These discussions were most recently pursued in the context of a workshop generously supported by the Fields Institute for Research in Mathematical Sciences, which brought together a number of students of causality, including Vanessa Didelez, Mervi Eerola, Michael Eichler, Paul Holland, Steffen Lauritzen, Wayne Oldford, James Robins, Don Rubin, Richard Scheines, and Ross Shachter, in addition to Dawid and myself. Dawid discussed his article at this workshop, and I am grateful to all the participants for their discussion of the issues it raises.

$Y_t(u)$. It is also legitimate to assess probabilities for $Y_c(u)$, because we can apply the control c to u and then observe $Y_c(u)$. But if we cannot do both, then it is not legitimate to assess probabilities for the pair $(Y_t(u), Y_c(u))$. Dawid vindicates de Finetti's criterion by showing that persuasive examples of causal inference that seem to require a joint distribution for $Y_t(u)$ and $Y_c(u)$ can be reformulated so that they clearly do not involve any such joint distribution.

Dawid's discussion of the instrumentalist use of counterfactual variables, one of the highlights of the article, demonstrates the effectiveness of his conciliatory approach. As Dawid makes clear, reservations about the empricial meaningfulness of counterfactual variables need not prevent us from using them for mathematical convenience.

My main reservation about the article is that it does not take advantage of Dawid's own path-breaking work on predictive probability. (See, for example, Dawid 1985, Dawid 1992, and Dawid and Vovk 1997.) In his effort to find common ground with those who tout counterfactual variables, Dawid emphasizes the case of a finite homogeneous population, where optimal predictions are merely population averages, and he hints that other cases can be reduced to this case by restricting attention to a subpopulation with a specific value of a covariate. This downplays the link between causality and prediction and obscures the potential richness of that link. It makes it difficult, for example, to recognize that the predictions authorized by casual regularities may often fall short of the full panoply of predictions that would be authorized by a probability distribution (Shafer 1996, 1998).

In the end, Dawid concedes too much, especially on the topic of causes of effects. These concessions can be avoided if remember that counterfactual variables do not provide the only framework for discussing causality. Frameworks that make a direct place for probabilistic prediction also have their uses, and they are needed to help us distinguish causal statements that have empirical content from those that are irremediably arbitrary and subjective.

1 Conditional Expected Values Suffice for Deliberation

John and his physician deliberate on whether John should undergo an operation. They decide to go ahead, and John dies on the operating table.

How long would John have lived had the operation not been undertaken? Before the decision is made, it is surely legitimate for the physician to talk about her expectations for how long John will live with and without the operation. If the physician is a mathematician, she may write $Y(\text{John})$ for John's longevity and assess the two expected values $E(Y(\text{John})|\text{operation})$ and $E(Y(\text{John})|\text{no operation})$, where

$$E(Y(\text{John})|\text{operation}) := \\ \text{John's expected longevity if the operation is undertaken} \tag{1}$$

and

$$E(Y(\text{John})|\text{no operation}) := \\ \text{John's expected longevity if the operation is not undertaken.} \tag{2}$$

At this point, before the decision whether to operate, there is nothing counterfactual about either of these quantities. After John's death on the operating table, $E(Y(\text{John})|\text{no operation})$ can be called counterfactual, because it involves a hypothesis that is now contrary to fact. It is a counterfactual expected value. But this name is not particularly enlightening. A better name might be *past conditional*.

There is, I believe, no disagreement about the meaningfulness or usefulness of past conditionals such as $E(Y(\text{John})|\text{no operation})$. Nor is there disagreement about the meaningfulness or usefulness of analogous predictions in cases where we can predict the consequences of treatments on a unit u for certain. In such cases, the conditional expected values $E(Y(u)|\text{treatment})$ and $E(Y(u)|\text{control})$, reduce to the conditional categorical predictions

$$Y_t(u) := \text{the value } Y(u) \text{ will take if } u \text{ is given the treatment } t \tag{3}$$

and

$$Y_c(u) := \text{the value } Y(u) \text{ will take if } u \text{ is given the control } c. \tag{4}$$

Again, there is nothing counterfactual about $Y_t(u)$ and $Y_c(u)$ before the decision is made whether to give u the treatment or the control, although if t is given it is acceptable afterwards to call $Y_c(u)$ a counterfactual prediction.

There is also no disagreement about the importance of quantities like (1)-(4) in causal assertion and hence in causal inference. The essence of causality lies in the fact that different actions will have different consequences or at least different expected consequences, and these differences can still be discussed after the passage of time changes "will have" into "would have had."

Controversy arises only when we ask whether causal questions should always be answered in terms of categorical predictions such as (3) and (4) or whether probabilistic predictions such as (1) and (2) can also be used. The *counterfactual approach* that Dawid criticizes bases causality on quantities of the form (3) and (4) in all cases, even when no such predictions can be made, even in principle. In this approach, even probabilistic predictions are interpreted not as conditional expected values, as we have done in (1) and (2), but as expected values of counterfactual variables. The conditional expected values $E(Y(u)|\text{treatment})$ and $E(Y(u)|\text{control})$ are recast as unconditional expected values $E(Y_t(u))$ and $E(Y_c(u))$.

Dawid concedes too much when he assents to this notational trick. The conditional expected values really are conditional. Yes, $E(Y(u)|\text{treatment})$ becomes $E(Y_t(u))$ when the decision is made to apply t to u. But $E(Y(u)|\text{control})$ remains a conditional expected value, now with respect to past rather than current probabilities. There is no need for us to imagine an alternative universe in which it has been promoted to an unconditional expected value.

2 Why Should the Black Box Contain Determinate Predictions?

As we have just seen, a thorough understanding of causal structure is not needed for deliberation. As Dawid explains, the assessment of the effects of possible actions "can proceed by an essentially 'black box' approach, simply modelling dependence of the response on whatever covariate information happens to be observed for the test unit." But in order to understand the "causes of effects," we need to probe inside the black box.

An autopsy may reveal facts about John that the physician could not have suspected or learned beforehand but which made the operation's failure likely. Any such facts obviously need to be taken into account in a discussion of the causes of John's death. In order for the conditional expected values in (1) and (2) to have causal meaning, they must take all such facts into account. When they do so, will they still be merely probabilities and expected values?

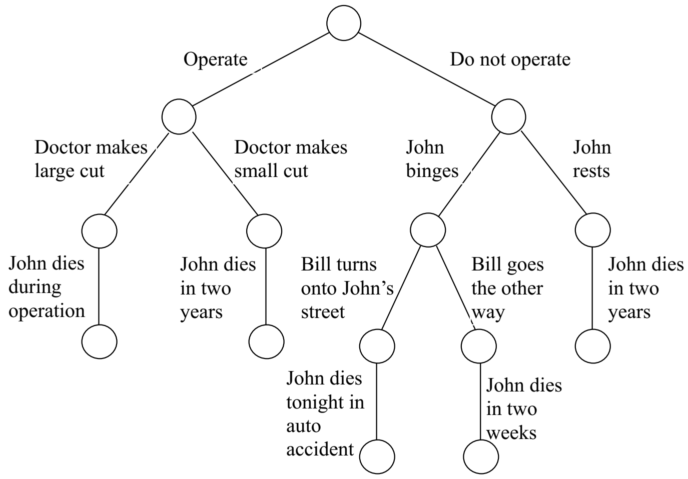

Figure 1: No one can predict exactly the results of treatment and control.

Or will they necessarily become determinate predictions, of the form (3) and (4)?

There are three powerful arguments against expecting determinate predictions.

1. Twentieth-century physics has repeatedly refuted efforts to eliminate probability from the predictions of quantum mechanics.

2. Our mundane experience also provides no support for the proposition that the effects of our actions are always determinate.

3. The very formulation of the question rests on an assumption that John and his physician can choose freely between having the operation and not having it. So how can we coherently deny that there may be later free choices, such as those suggested by Figure 1, that may also effect John's longevity?

The proponents of counterfactuals often respond to the mention of quantum mechanics by suggesting that it is too esoteric to be relevant to everyday concerns. In medicine, business, and law, they argue, we can make do with Newtonian mechanics, where actions have predictable consequences. Unfortunately, as the second argument reminds us, Newtonian laws do not get us very far in understanding the choices we face in medicine, business, and law.

Even Laplace's vision of determinism, in which a superior but human-like intelligence can predict the future states of the world from knowledge of the present state and a small number of laws, demands only the possibility of prediction for states in which the world is actually found. If causal laws predict everything, they predict that the physician will undertake the operation. So the Laplacean vision does not require that the superior intelligence should be able to make a prediction about what would happen if the operation is not undertaken.

The force of the third argument was already acknowledged in the work by Don Rubin that launched the revival of counterfactuals in statistics in the 1980s and 1990s. In Rubin (1978, pp. 39–40), we find three conditions that should be met before one assumes that treatments have determinate results whether or not they are applied:

1. Each treatment should be defined by a series of actions that can be applied to the individual. For example, if we insist on studying the causal effect of being female, we must specify the particular actions to be taken to make the individual female.

2. Any pretreatment manipulations should be included in this series of actions. For example, if different medical treatments are preceded by different physical examinations, the examination should be considered part of the treatment.

3. "...we cannot attribute cause to one particular action in the series of actions that define a treatment. Thus treatments that appear similar because of a common salient action are not the same treatment and may not have similar causal effects."

The third condition can be elaborated by saying that the series of actions defining a treatment must include all human actions that can affect the outcome. Thus we need to include in the definition of John's treatment not only the doctor's actions but also John's and Bill's actions in Figure 1.

3 Causal Structure with Objective Probabilities

We have been led to the conclusion that probabilities with causal meaning—objective probabilities, if you will—are those based on all the information humanly possible to have and use in a given situation. As Dawid might put it, these are the probabilities based on a *sufficient covariate*.

This conception of objective probability hardly new. In the mid-nineteenth century, Antoine Augustin Cournot, adapting Laplace's formulation of determinism, proposed that objective probabilities are those that a superior but human-like intelligence would obtain using the current facts about the world and knowledge of causal regularities. The idea echoed well into the twentieth century, in the work of authors such as Henri Poincaré (1908) and Émile Borel (1924).

What can we say about causality when we have only probabilistic predictions such as (1) and (2) instead of categorical predictions such as (3) and (4)? As I have argued in *The Art of Causal Conjecture* (1996), we can say a great deal. We may assert that a particular action (by a person, an animal, or some inanimate actor, such as a storm or a comet) changes the expected value of some variable or the probability of some particular outcome. The action is, in this sense, a cause. In some cases, we may conjecture broader causal regularities. We may conjecture, for example, that a given type of action always raises the expected value of a given variable. Or we may conjecture that all the causes of one variable are causes of another. For example, we may conjecture that most actions that increase the expected value of a person's smoking decrease the expected length of the person's life.

One of the attractions of counterfactual variables for statisticians over the past several decades has been precisely the fact that they allow us to do without objective probability. At least they allow us to reduce objective probability to the simpler concept of frequency in a finite population. This is attractive to some because of their Bayesian persuasions and to others because of their weariness with long-running debates about the meaning of probability. It now seems clear, however, that nothing has been gained by replacing objective probabilities with categorical counterfactuals. Objective probabilities have an empirical meaning at least in principle—they represent what we might obtain in the limit as we improve our predictions through additional experience and knowledge. Categorical counterfactuals, everyone agrees, are often unknowable even in principle.

4 Asking the Wrong Question

In my view, Dawid concedes too much when he allows that categorical counterfactuals may be needed for inferences about the causes of effects. He concedes too much by agreeing to pose a question that has no meaning.

Something has happened, and we are being asked whether one particular step in the course of events has caused it.

- My headache is gone. Is it because I took aspirin?
- John died on the operating table. Is it because the physician operated?
- Our corn crop failed. Is it because of the variety of seed we planted?

Dawid's poses the general question in these words, "We are interested in whether, for the specific unit u_0, the application of t 'caused' the observed response." He lets us know, with the quotation marks around *caused*, that he is asking a silly question. Unfortunately, the quotation marks do not save him from becoming entangled in silly answers.

Imagine there were a categorical rule about the effect of aspirin on headaches:

- At least two aspirin with at least a cup of water: the headache goes away.
- Less aspirin or less water: the headache persists.

I take the requisite aspirin with the requisite water. My headache goes away. Is it because I took aspirin? We understand the causal structure perfectly, but we cannot answer the question with a simple yes or no.

It is equally silly to isolate a single action and ask whether it is *the* cause when the action's effect depends on something that is settled later—what Dawid calls a "determining concomitant." Figure 2 shows a very simple example. I am required to bet \$1 on the outcome of a toss of a coin. I decide to bet on heads, the coin lands tails, and so I loose my \$1. Did my choice of heads "cause" my handing over \$1 instead of receiving \$1?

Here again, we understand the causal structure perfectly. The coin is fair; the chance of its landing heads is 50% regardless of how I bet. The outcome Y (which will be either either $+1$ or -1) is completely determined by the treatment T (which will be either "bet on heads" or "bet on tails")

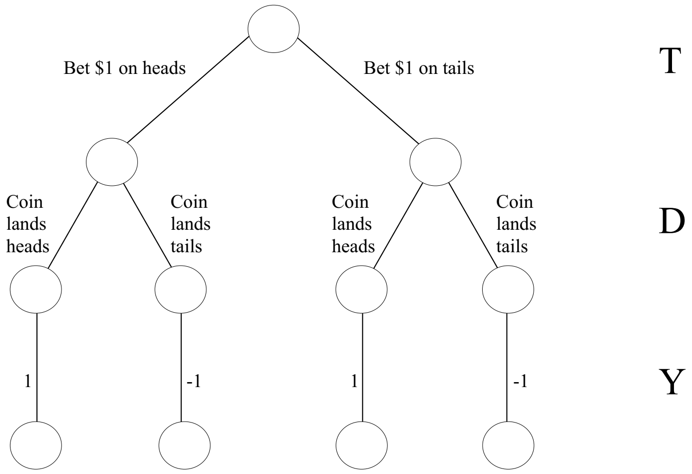

Figure 2: A determining concomitant.

together with the determining concomitant D (which will be either "coin lands heads" or "coin lands tails"). We understand exactly what happened. But this does not enable to us to give a yes or no answer to the question whether T "caused" Y.

Had I instead bet on tails, would the toss come out the same way? What is gained by asking this question or by making up an answer to it? We can make up whatever answer we want. If we assume that this particular determining concomitant D comes out the same in the counterfactual world where I bet on tails, then I win in that world. If choose a different determining concomitant D', whose possible values are "coin lands the way I bet" and "coin lands opposite the way I bet", and we assume that it comes out the same in the counterfactual world, then I lose in that world. What is the point or content of either assumption?

Dawid sees the arbitrariness clearly, and so he arrives at this formulation: "the essence of a specific causal inquiry is captured in the largely conventional specification of what we may term the *context* of the inference, namely, the collection of variables that is is considered appropriate to regard as concomitants..." Specification of the context, he concludes, "is vital to render causal questions and answers meaningful."

In practice, I am willing to trust Phil Dawid to take the air out of silly questions by showing how they depend on the arbitrariness of the choice of a context. But I distrust his formulation, for it seems to say that all singular causal questions partake in this silliness—that all causal answers depend on the arbitrary specification of concomitants.

5 Asking the Right Question

My father quit smoking in the early 1960s, after using cigarettes heavily for more than 20 years. He died in 1997, at the age of 75.

- How long would he have lived had he continued smoking?
- How much did his quitting smoking change his life expectancy?

These are both questions about the causes of an effect. They are both questions about singular causation. They are both questions about my father's quitting smoking as a cause of his observed longevity. The first is a wrong question. It will remain a silly question no matter how many different concomitants Phil Dawid tries out on us. The second is a right question. It is a scientific question, which comes with its own context and requires no arbitrary specification of concomitants by me or you or Phil Dawid.

In the United States, litigation continues between the tobacco industry and the federal government, which seeks compensation for the costs of caring for people whose health was damaged by smoking. The tobacco companies are held to be liable on the grounds that they took actions to increase cigarette consumption in spite of their own knowledge of its ill effects. To measure the effect of these actions, we can ask a scientific question:

> How much was the expected government expense on caring for the ill increased by the actions of the tobacco companies?

Or we can ask a counterfactual question:

> How much less would the government have spent on caring for the ill had the companies had not taken these actions?

The scientific question is very difficult. Its answer is subject to great uncertainty. The counterfactual question adds arbitrariness to the uncertainty. An insistence on the counterfactual question will lead in the end to denial of responsibility. How much blame to place on the tobacco industry becomes not a scientific question but a purely political one: What counterfactual world do we want to imagine?

We all indulge, in anger and regret, in counterfactual talk. "If they had not operated, John would be alive today." "If I had not said that, she would not have left me." "If I had chosen a different publisher, my book on causality without counterfactuals would have sold 10,000 copies." The more fortunate among us have someone to remind us that we are talking nonsense. Calmer heads will remind John's son and widow that his length of life had the physician not operated does not have a determinate value.

The physician's responsibility is to compare (1) and (2) based on the best evidence she can reasonably gather, and to perform the operation, if she does perform it, with expertise and care. We can ask for no more. As Jacob Bernoulli, the inventor of mathematical probability, wrote, *De Actionum humanarum pretio non statuendum ex eventu* (Bernoulli 1713). Do not judge human action by what happens.

References

[1] Jacob Bernoulli. *Ars conjectandi.* Thurnisorum, Basel, 1713.

[2] Émile Borel. À propos d'un certain traité. *Revue philosophique*, 98:321–326, 1924. This note, a review of Keynes (1924), was reprinted on pp. 2169–2184 of Volume 4 of Borel's complete works, *Oeuvres de Émile Borel*, Centre National de la Recherche Scientifique, Paris, 1972. An English translation appeared in Kyburg and Smokler (1964).

[3] Antoine Augustin Cournot. *Essai sur les fondements de nos connaissances et sur les charactères de la critique philosophique.* Hachette, Paris, 1851. Reprinted as Volume II (J. C. Pariente, editor) of the *Oeuvres complètes*, 1975, Paris: J. Vrin.

[4] A. Philip Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13:1251–1285, 1985.

[5] A. Philip Dawid. Prequential data analysis. In M. Ghosh and P. K. Pathak, editors, *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, IMS Lecture Notes–Monographs Series 17, pages 113–126. Institute of Mathematical Statistics, Hayward, California, 1992.

[6] A. Philip Dawid and Vladimir G. Vovk. Prequential probability: Principles and properties. *Bernoulli*, 3:1–38, 1997.

[7] John Maynard Keynes. *A Treatise on Probability*. London, Macmillan, 1921.

[8] Henry E. Kyburg, Jr and Howard E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964.

[9] Henri Poincaré. *Science et méthode*. Flammarion, Paris, 1908.

[10] Glenn Shafer. *The Art of Causal Conjecture*. The MIT Press, Cambridge, Massachusetts, 1996.

[11] Glenn Shafer. Mathematical foundations for probability and causality. In Frederick Hoffman, editor, *Mathematical Aspects of Artificial Intelligence*, volume 55 of *Symposia in Applied Mathematics*, pages 207–270. American Mathematical Society, Providence, Rhode Island, 1998.