Philosophy term paper: Critically evaluating Hume's skepticism of causality

Siddharth Bhat

Monsoon 2019

David Hume (1711-1776) was a contemporary of John Locke and George Berkely. While all of them advocated using empiricism to arrive at knowledge, Hume was the only consistent empiricist among the three, who applied his empirical standards rigorously and unflichingly. This allowed him to arrive at oftentimes counter-intuitive results, such as challenging the very existence of causation.

Since Aristotle, there was thought to be a difference between *belief* and *scientific opinion*. Scientific knowledge consisted of providing explanations in terms of necessary causes for an effect to manifest. As an example, smoke is *caused* by fire, since for smoke to exist, there must *necessarily* exist fire that causes this smoke.

0.0.1 Hume's Guillotine

0.0.2 Convention and Causality

The crux of Hume's argument is that of convention. All that we suppose to be "causal", is in fact simply convention, that we as humans have gotten used to due our experience. He provides a famous example of billiard balls: When we see that the motion of one billiard ball follows another, we are only observing conjunction, not their connection. Indeed, his definition of "cause" is this:

> A **cause** is an object, followed by another, where all the objects similar to the first are followed by objects similar to the second.

1 A resolution by Karl Popper

These problems of causation and their relation to science were eventually resolved (in a sense) by Karl Popper(1902-1984), who gave an account of scientific knowledge in terms of a strictly empirical account driven by falsification, instead of one driven by induction. Roughly, his solution relates to the adage:

> All models are wrong, some models are useful.

Where he views the entirety of science, and as an extension, our model of causality, as precisely that: a model. No knowledge that we possess, not even that of obvious causality, is certain. Knowledge is correct, insofar as it provides a useful model of the world. A model, or a notion of causation can be discarded as soon as it does not align with the reality. In this way, the notion of *falsifiability* takes center stage.

In short, Popper's theory can be summed up as:

- There is no correct way to perform induction, given access to any amount of experimental data. This agrees with Hume's guillotine.
- Our theories (including our notion of causality, which is governed by our theory), is just a conjecture. We must attempt to falsily these conjectures. If we are unable to falsify a conjecture, then we have greater faith in the conjecture. However, we can never truly establish a fact (such as causality). All we can do is to attach a *probability* to causality.

1.1 The Bayesians: Formalizing Popper

The bayesians were a school who took as fundamental Bayes' theorem, a general theorem about the structure of probabilities which describes how to *incorporate* experimental knowledge (obervations) into our current knowledge of the world (the prior) to arrive at an updated knowledge of the world (the posterior). The famous formula is:

$$P(B) = \frac{P(B|A) \times P(A)}{P(A|B)}$$

Far too much ink has been spilled analyzing the above formula, so I wil not repeat that here. The basic point is that this theory provides us with the *best possible way to update our beliefs*, if our beliefs about the nature of the universe are purely probabilistic.

Thus, bringing this back to Humian terms, bayesian theory allows us to decide *which conventions we should choose*, given raw experimental data.

However, this again loses a notion of causation, since we are again making probabilistic statements about the world. In theory, a causation would be repressented as a *perfect correlation.* However, this has its own can of worms. A simple problem with this model is that we lose the **directionality** of cause and effect.

For example, we might have the readings on a thermometer, and the real temperature of that day. Using this data, we would create a formula $r = \alpha t + \beta$, where r is the reading of the thermometer, t is the real temperature, and α, β are constants. This gives us a model for the relationship between the thermometer reading and that of temperature, and indeed these two variables are perfectly correlated. However, crucially, there are *two* possible interpretations of the formula: one saying that the temperature governs the reading of the thermometer

$(r = \alpha t + \beta)$, and the other saying that the reading of the thermometer governs the teperature $(t = (r - \beta)/\alpha)$. The structure of mathematical equations is such that both of these are entirely valid views.

What we really need is a new form of mathematics, that allows us to depict $(r \leftarrow \alpha t + \beta)$ where the $\leftarrow$ indicates that the r is determined by t and not the other way round.

This was believed for roughly a century since the birth of statistics to be impossible, by (fundamentally), an appeal to Popper and Hume: how, indeed, can we establish such a directionality, if all we have is the data at hand?

2 A synthesis of causation: Judea Pearl and the do-calculus

Here enters the fascinating study intiated by Judea Pearl into what are now called *Causal Models*.

The insight is that one can provide a definition of causality, provided *counterfactuals* are available. One can *define* causality as follows. For the sentence:

> The billiard ball A colliding with billiard ball B **caused** billiard ball B to move.

We assign the meaning of the word **caused** as the meaning of this equivalent sentence:

> **In all universes** where the billiard ball A **not** collided with billiard ball B, the billiard ball B **would not move**.

This sentence does not seem to improve our state of affairs by much. It has managed to provide a definition of causality. However, this definition relies on universal quantification of an all universes, which is a tad unrealistic to test.

Note that this is a different notion of causality than that of hume, in that it is verifiable, provided access to an infinite number of universes, and the ability to perform experiments.

Luckily for us, due to the structure of mathematics, *once we fix some causal model*, for example, (A collided with B $\rightarrow$ B moved), we can test how well the *data* agrees with the model.

While as bayesians, we had: $(\text{data} \longrightarrow \text{correlations})$

Note that there is no human-in-the-loop. From the data, we compute, plug and chug to arrive at correlations. To a statistician, this is all that we can say, and indeed, all that we need to say. Unfortunately, this is a disastrous position to hold. After all, as we all know, "correlation does not imply causation". We have yet been unable to escape the long shadow cast by Hume. "correlation" is simply a fancy mathematical encoding of "convention", is it not?

Now as users of do-calculus, we have:

$$
\begin{tikzcd}
\text{agreement (do-calculus)} & \text{causal model} \arrow[l] \\
\text{data} \arrow[u] &
\end{tikzcd}
$$

Notice that we *explicitly* require a model to even begin applying do-calculus. However, the upshot is that we can now discuss how well our proposed model of reality agrees with the data that we have. Indeed, our process of ”doing science” can be now viewed as:

$$
\begin{tikzcd}
 & \text{causal model } n \ (P = 0.1) \arrow[dl] \\
\text{agreement} & \text{causal model 2 } (P = 0.5) \arrow[l] \\
\text{data} \arrow[u] & \text{causal model 1 } (P = 0.2) \arrow[ul]
\end{tikzcd}
$$

The takeaway is that *relative to a model* i, we can judge how well this model captures the causality of the data!

This is *both subjective*, or bound by convention (in terms of the choice of model), but also objective *relative to a model.* Once we choose a model, we can check its ”causal agreement” with the data.

Hence, I argue that do-calculus and our 21st century understanding of causality has settled Hume's arguments to a far greater degree than Hume believed was ever possible: under an intuitive definition of causality (counter-factuals), we can build an edifice of mathematical machinery which allows us to test models of reality against data. While this choice of model is non-canonical, and this is where the human enters the loop, beyond this choice, we have objectivity.

References

De Pierris, Graciela, and Michael Friedman. 2018. “Kant and Hume on Causality.” In *The Stanford Encyclopedia of Philosophy,* Winter 2018, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Hitchcock, Christopher. 2019. “Causal Models.” In *The Stanford Encyclopedia of Philosophy,* Summer 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Hume, David. 2003. *A treatise of human nature.* Courier Corporation.

———. 2016. “An enquiry concerning human understanding.” In *Seven Masterpieces of Philosophy,* 191–284. Routledge.

Jaynes, Edwin T. 2003. *Probability theory: The logic of science.* Cambridge university press.

Lorkowski, CM. 2010. “David Hume: Causation.” *Internet encyclopedia of philosophy.*

Morris, William Edward, and Charlotte R. Brown. 2019. “David Hume.” In *The Stanford Encyclopedia of Philosophy,* Summer 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Pearl, Judea. 1994. “A probabilistic calculus of actions.” In *Uncertainty Proceedings 1994,* 454–462. Elsevier.

Popper, KR, and D Miller. 1953. *The problem of induction.*

Sepetyi, Dmytro. 2013. “Karl Popper’s Solution to the Problem of Induction and Hume’s Problem. The Minimalist Conception of Rationality.”

Thornton, Stephen. 2019. “Karl Popper.” In *The Stanford Encyclopedia of Philosophy,* Winter 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Yudkowsky, Eliezer. 2015. “Rationality: from AI to Zombies.” *Machine Intelligence Research Institute, Berkeley.*