Submission for the FPS conference at TiLPS 2010

Interventions: a case study in formalisation

Jan-Willem Romeijn
University of Groningen

Short abstract

In this paper I discuss probabilistic models of experimental intervention, and I show that such models elucidate the intuition that observations during intervention are more informative than observations per se. Because of this success, it seems attractive to also cast other problems addressed by the philosophy of experimentation in terms of such probabilistic models. However, a critical examination of the models reveals that some of the aspects of experimentation are covered up rather than resolved by probabilistic modelling. I end by drawing a number of general lessons on the use of formal methods in the philosophy of science.

Extended abstract

Ever since the advent of the philosophy of experiment in the 1980ies, interventions have been a topic of explicit philosophical interest; cf. Hacking 1980, Franklin 1986, and Radder 1996. From the very beginning, the consensus has been that observing a system while intervening on it provides more information about, and deeper insight into the system's workings than merely observing it. This intuition is backed by more recent research in cognitive developmental psychology, showing that the possibility of intervention allows for much faster learning than that of mere observation; e.g., Steyvers et al 2007. However, from the philosophy of experiment itself, a further explanation of this consensus view is not immediately apparent.

The philosophical study of intervention was provided with a new perspective when the use of graphs in representing probability functions, an idea originating in computer science, was combined with philosophical ideas on probabilistic causality. The resulting theory of causal Bayesian networks includes a prima facie convincing notion of intervention: the network represents the causal structure of the system under scrutiny, and with some additional suppositions, it also provides a recipe for determining the consequences, for the variables that characterise the system, of an exogenous change to some of those variables. The DO-calculus devised in Pearl 2000 serves as a prime example of this idea but many alternatives have been considered, e.g., in Korb 2007.

One of the attractive features of these probabilistic models of interventions is that they allow us to elucidate the intuition that interventions are more informative than mere observations; see Romeijn and Williamson 200X. An illustration of this idea involves the resolution of underdetermination by means of intervention data in the context of structural equations modeling. If we suppose that the data that was observed initially and the data that was observed after intervention are related to each other according to the recipe as laid down by the network, we can impose further constraints on the parameters characterising the network, and thereby

eliminate unidentified parameters. If, on the other hand, we had treated the additional data simpy as observations, no further constraints could have been derived, leaving the parameters underspecified.

This is a promising result, and it invites us to stretch the probabilistic analysis of interventions to encompass other problems in the philosophy of experiment, towards a complete formalisation of the confirmatory role of experimentation. Quite independently of the success mentioned above, I think there are good reasons for embarking on such a philosophical project. First, despite some efforts by statisticians like Dawid and Rubin, the sciences themselves seem to be in need of tools for dealing with experimental data. In standard statistical techniques, interventions are seen as events that effect randomisation and that create groups with different distributions; with this impoverished view, scientists do not make full use of the information that their interventions might offer. Moreover, the philosophy of science could benefit from a project of this kind. In confirmation theory specifcally, it is standard practice to idealise away from the way in which data are obtained; this causes confirmation theory to loose at least some of its potential relevance.

Rather than sketching the outlines of a formal theory of experiment as part of a future philosophy of science, I will in the remainder of this paper target the philosophical methodology that underlies such a project: formalisation. It is a venerable and perhaps even a key objective in the philosophy of science, and the use of causal Bayesian networks to elucidate experimental interventions provides an interesting case study of it. Can we characterise the problems raised in the philosophy of experiment in formal terms? And if these problems resist a formal characterisation, to what extent is that to do with the subject matter? As it turns out, we can learn a number of things on experimentation by scrutinising the ways in which our formal means for capturing experimentation fall short. I will concentrate on two such shortcomings in particular.

To appreciate the first of these, it may be helpful to stress how daring the idea of experimentation is: to uncover the natural workings of a system of interest we bring it into unnatural circumstances in which it would normally never be found. Of course this is not as crazy as it might seem. Once we assume that there is an inherent structure to the system, bringing it in unusual circumstances can be viewed as providing access to that inherent structure, because in these circumstances it is stripped of all its accidental characteristics. But note that the underlying structure is not just the end product of the experiment; in a general form it is also a presuppostion.

Now, turning to a formal characterisation of this aspect of experimentation, we might say that it is neatly captured by the fact that intervention data can only be framed if we make suppositions on the causal network. The important difference is that in the case of experimentation, we might be completely ignorant on what the inherent structure is, as long as we suppose it is there. We can suppose that we are blind to the workings of our own experimental interventions. They derive their content in part from an external world that may be entirely unknown, much like the meaning of a word may be fixed externally, by what the world happens to be like. In a causal Bayesian graph, any such ignorance will take the form of uncertainty over the graph structure, but in that case we are painting a rather detailed picture of what

it is we do not know. The formal model will have to accommodate the ignorance over the external world in terms of the uncertainty over some set of variables, much like the meaning of a word can be fixed by a list of possible referents. In short, the internalist perspective of the formal model cannot match the externalism of experimentation.

There is a second way in which formal models of experimentation fall short of providing a proper representation of experiment, and which I take to reveal something interesting about experimental interventions. Somewhat speculatively, we might say that experiments are different from mere observation because they allow us a sneak preview of another possible world, or in more technical terms words, because they provide knowledge of a counterfactual nature: we intervene in order to observe how the system diverges from what would have happened if we had left the system to evolve unperturbed. Of course, we presume to know that, if we had not perturbed the system, nothing out of the ordinary would have happened, and hence we can ascribe the observed effects to the causal role of the intervention.

How exactly does this aspect of experimentation find its way into the formal representation? In the causal Bayesian model, the effect of an intervention comes down to a controlled move in a space of probability functions; after the intervention we can formulate a precise probability assignment over all the variables involved, and we can subsequently compare that assignment to the function obtaining before the intervention. This may look like moving from one possible world to another, thus emulating the idea of counterfactual knowledge. But again, I maintain that the formal model is crucially different. In this case, it misses out on the component that gives rise to the illusion of obtaining counterfactual knowledge: the agent that chooses the intervention. Minimally, it is not clear how this aspect of agency can be incorporated into the essentially empiricist formal models of experiment.

Summing up, I argue that a formal representation of experimental interventions in terms of causal Bayesian networks misses out on aspects of experimentation that, at least from the point of view of those resisting formalisation, are relevant to the philosophical discussion of experiments. These aspects have to do with externalism and agency, as spelled out in the foregoing. In my view they mark the limits of formal modelling more generally. Any such modelling will be based on an objective and conceptual reconstruction. The position of the knowing and acting subject as well as her genuine lack of understanding of the environment are in a formal model replaced by a third person perspective and uncertainty over a distinct epistemic domain. Whether we must conclude from this that formalisation has its limits or that the philosophy of experiment needs revision, is another question altogether.