Modelling Experimental Interventions:

Results and Challenges

$\star$

Jan-Willem Romeijn

Faculty of Philosophy
University of Groningen

Contents

1 Overview

The first part of the paper is joint work with Jon Williamson. I deal with. . .

- the common sense idea that we can learn from interventions,

- a formal explication of this by means of Bayesian networks.

These results invite the use Bayesian networks in clarifying experimentation more generally. But I will argue that. . .

- they fail to represent how experiments rely on an external world,

- they do not explicate the modal aspect of causal knowledge.

I end by formulating these criticism as challenges to the idea of modelling interventions by formal means.

2 Interventions and confirmation

The idea that interventions provide preferred epistemic access is commonplace in science, in cognitive psychology, and in the philosophy of experiment.

- History shows the strong ties between the experimental tradition and the birth of modern science (Hooykaas, Galison).

- Philosophers of experiment emphasize that controlled intervention reveals, or even constructs, reality (Hacking, Franklin, Gooding, Radder, Collins).

- Studies in cognitive and developmental psychology show that learning is more efficient in tandem with interventions (Gopnik, Glymour, Steyvers, Tenenbaum).

This paper aims for a fuller understanding of this idea by presenting a formal model of experimentation.

Traditional confirmation theory

In traditional confirmation theory, observation data and intervention data are not treated differently. The focus is on the confirmatory relation between data and model or theory. Some exceptions:

- Bayesian treatments of triangulation, calibration, and the Duhem-Quine problem, in which auxiliary hypotheses on measurement apparatus play a role (e.g., Howson and Urbach).

- Erotetic approaches to scientific experiment, likening investigation to a game of questions and answers (Niiniluoto, Rosenkrantz).

- General philosophy of science on the confirmatory and explanatory role of physically realised models (Morgan, van Fraassen, and many others)

None of these approaches provides clarifies the special confirmatory virtues of intervention data.

Experimentalism and statistics

In the philosophy of statistics we do find particular attention for the confirmatory virtues of experimentation.

- Experimentation as putting hypotheses to severe tests, with a focus on experiments as a way to provoke error (Mayo's school).

- Attention, often critical, for experimental design as a means to generate representative samples (Hacking, Fisher, numerous statisticians).

- The recent use of causal Bayesian networks to model experimental interventions (Spirtes *et al*, Pearl, Korb, and many others).

For revealing the confirmatory virtues, the latter perspective is most promising.

Interventions and causal structure

Bayesian networks can be used to choose between different causal structures.

Example from "The Effect of Country Music on Suicide" in Stack and Gundlach (1992) *Social Forces* 71(2): 211–218.

Interventions and causal structure

By intervening on the control variable *C* and observing the effect variable *S* we may detect the existence of a causal link between the two, or else the existence of a common cause *A*.

The two candidate networks entail distinct and testable implications for the interventions.

3 Unidentifiability in factor analysis

Romeijn and Williamson (20XX) argue that interventions can be also used to resolve problems of underdetermination. They illustrate this by a case study on the rotation problem in factor analysis, a technique that is widely used in the social sciences.
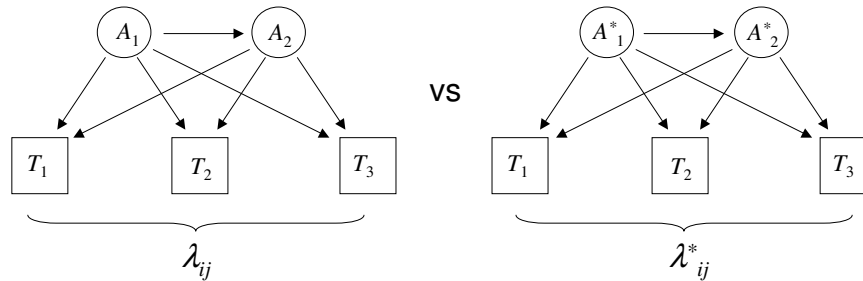
The rotation problem is that there are infinitely many ways to regress a set of correlated manifest variables, e.g. test scores, on a set of latent variables, e.g. psychological attributes.

Fear and loathing in Bayesian networks

Say that fear F and loathing L are both binary manifest variables, and consider a single latent cause, depression D. Observations are of individuals being fearful and loathsome or not, so there are four categories.
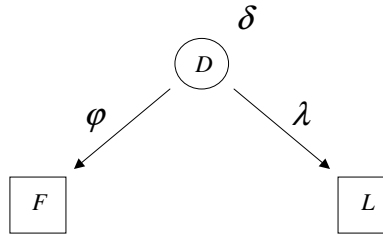
Every probability assignment consistent with the graph is a hypothesis with statistical parameters δ, ϕ_d, and λ_d. In other words, the hypotheses are Bayesian networks on the given graph.

Unidentifiability in Bayesian networks

We may count the degrees of freedom in the statistical model, after applying a fit criterion. We have a total of five parameters in the statistical model:

- the chance of an individual for being depressed,

- two separate chances for being loathsome, depending on whether the subject suffers from depression, and

- two such chances for being fearful.

But we have only 4 observed relative frequencies, namely for the observations of the four categories concerning fear and loathing, with the further restriction that they add to 1. There is therefore a 2-dimensional continuum of statistical hypotheses that fit the data perfectly.

4 Using intervention data

The basic idea of using interventions for the purpose of solving this kind of underdetermination is as follows.

- An intervention changes the distribution over the latent variable of the subjects. In the example, the probability for depression is altered.

- It does not change the probabilistic relations between the latent and the manifest variables. In this case, we keep the probability of fear and loathing conditional on depression fixed.

- After the intervention, we obtain a new estimation problem for the parameters in the statistical model.

The key point is that to accommodate the intervention data, we have a smaller space of parameters available.

Drugs to the rescue

Say that we intervene on the depression by administering a drug E to the subjects. We model this by an additional node, setting the probability for depression to a new but unknown value ε.
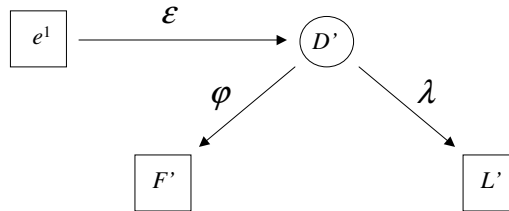
Note that in order to frame the intervention, we assume that the latent variable model is correct and that we intervene only on the depression node.

Unidentifiability resolved

Because we observe the fear and loathing of the individuals after the intervention, we have 3 new and additional observed relative frequencies.
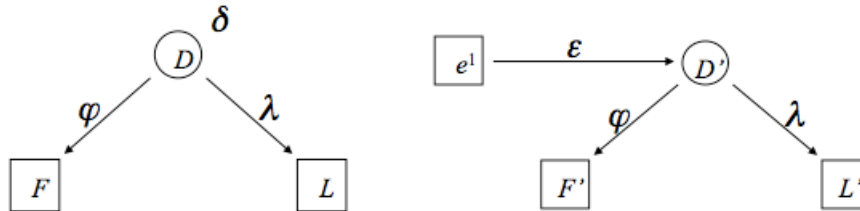
This brings the total to 6 observed relative frequencies. But we have only one additional parameter in our problem: the chance of a subject suffering depression after taking the drug. This brings the total number of parameters on 6 as well, so there is a unique best fit!

Drugs as a guidance to theory change

It may also happen that the fit with the collected observational and interventional data is poor, according to some model selection or fit criterion.
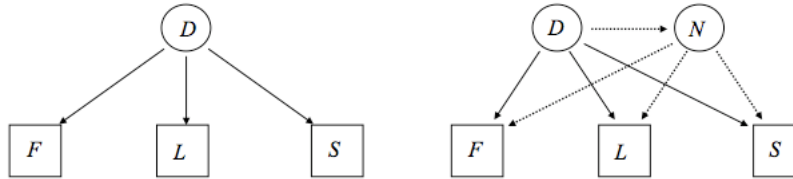
In that case we can add latent variables to the model, enlarging the statistical model. The intervention data suggest specific ways of doing this: different extensions of the model lead to different fit. They guide the conceptual change.

Interventions replace theoretical criteria
The resolution of underdetermination and the generation of theory are typically achieved by theoretical considerations.

- Choosing among a set of empirically adequate hypotheses involves their theoretical virtues, e.g., explanatory power, coherence with other theories, and so on.

- Despite research into the process of discovery, the inclusion of new mechanisms or entities in a model is often associated with the imagination of the scientists.

In the case of factor analysis, we see that such theoretical considerations are partly replaced by empirical ones.

5 A formal philosophy of experiment?

The above results on the formalisation of experiments invites us to approach other conceptual and methodological puzzles in the philosophy of experimentation in similar fashion.

- The experimenter's regress (Collins, Latour): what are we measuring in experiment if, in the act of measuring, we investigate what we measure?

- The isolation of theoretical concepts (van Dyck, Chang): how do experiments help to fix the reference of the theoretical terms in our theories?

- Methods for experimatation (Rubins, Dawid): what is the correct statistical tool for incorporating evidence from experiment into our theories?

Instead of embarking on these bigger projects, I want to indicate some fundamental problems with the formal framework just laid down.

6 Externalism and types of ignorance

As illustrated by the example, the use of experimental data relies on the presupposition of a causal structure.

- Only if we assume that some causal structure is responsible for the data we obtained, we can maintain that the experimental intervention leads us to the more specific causal structure.

- The exact confirmatory role of the intervention data, i.e., the way the data bear on the hypotheses, is prescribed by the assumed causal structure.

I argue that this aspect of experimentation interferes with the aim of explicitly modelling experimental interventions.

Meaning externalism
The meaning of terms in our language may well be fixed by states of affairs in the world rather than facts of the matter about the world.
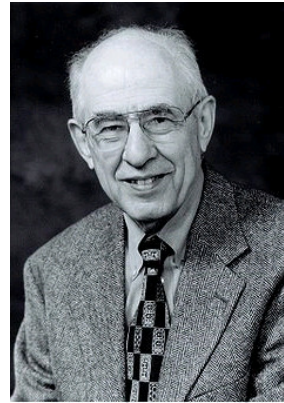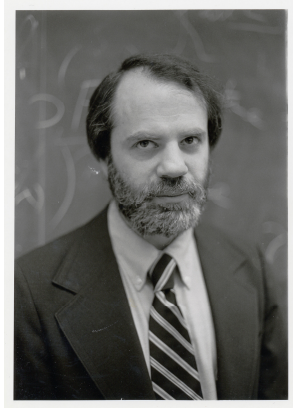
So in our use of language, we rely on the world having a particular structure.

Confirmational externalism

In the use of experiments in science, we encounter the same kind of reliance on external structures, of which the experimenter may be perfectly ignorant. It need not be clear to the experimenter what it is she is manipulating.

- In a formal model such ignorance will have to be captured in terms of the uncertainty over some definite domain of possibilities. But what if we have not yet chosen such a domain?

- Often the experimenter first establishes stable patterns in the data, only to attach variables thereafter. So even if we choose a definite domain, what if we have not yet delineated the variables?

Here we encounter the problem on internal and external languages that Carnap addresses: first we need to settle on a language, and then we can do philosophy. Or can we?

7 Modality vs empiricism

Experiments seem to allow us a "peek into another possible world". The idea is that they provide knowledge of a counterfactual nature.

- We intervene in order to observe how the system under scrutiny diverges from what it would have been like, if we had left it unperturbed.

- If we had not perturbed the system, nothing out of the ordinary would have happened, hence we can ascribe the observed effects to the causal role of the intervention.

I argue that this aspect of experimentation is at variance with the Humean empiricism that seems inherent to the formal model of experimentation.

The problem of induction
To the disgrace of philosophers, there is still no solution for the problem that causal facts do not follow from observations alone.
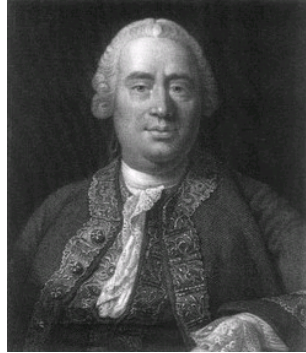
To learn something new from interventions, we need to assume that the data obtained before and after the intervention are related in a way that is regulated by a causal model.

Intervention by stipulation
The problem is not so much that we rely on this supposition of causal structure, but rather that we consider the decisions of the experimenter as the ultimate exogenous variable in that structure.

At the end of the day, all the data are gathered in the same actual world. It may well be in virtue of our sense of free agency, or utter randomness, that we frame the data as pertaining to multiple worlds.

8 Formal methods in philosophy

I fear that these problems are not idiosyncratic for the programme of using causal Bayesian network models of experimentation, but endemic to much of formal philosophy of science.

- The priority of language in formal modelling flies in the face of the externalist aspects of much of scientific activity.

- The traditional empiricist backdrop of many programmes of formal modelling is at variance with the modal character of almost all scientific knowledge.

Of course, this is a rather negative reading of the foregoing. Instead, the audience is welcome to consider the criticisms as challenges!

Thank you

The slides for this talk will be available at http://www.philos.rug.nl/~romeyn and the full paper will also be posted there. For comments and questions, email j.w.romeijn@rug.nl.