

The only other alternative is to understand Alf's beliefs about reasons on the right-hand side of the previous bi-conditional as desire-like attitudes. The resulting view would say that Alf has a sufficient reason R if and only if after a respectable amount of good reasoning he would want to act because of R. This would make Alf's reasons a direct function of what he wants and thus Gaus's view would be a simple form of subjectivism about reasons, which most metaethicists reject for good reasons. Of course there are many metaethical expressivists who claim that utterances about reasons express our desire-like attitudes and as we saw above Gaus has a lot of sympathy for this view. However, even expressivists are not committed to the idea that Alf's reasons depend on what he wants. The expressivists can perfectly well express their own positive attitudes towards Alf's actions that do not depend on Alf's desires.

All of this matters. If *The Reasons One Has* fails, Gaus's motivation for thinking that reasons to accept moral principles must be accessible to each agent from her own deliberative perspective is undermined. If this is right, liberal members of the public who rank suggestions about more extensive states low wouldn't be able to veto merely on the basis of what reasons they recognize from their actual deliberative perspectives. This would enable us to justify more extensive states to everyone on the basis of the reasons they have but cannot recognize.

Jussi Suikkanen
University of Birmingham, UK

REFERENCES

- Enoch, D. 2013. The disorder of public reason *Ethics* 124: 141–176.
Hare, R.M. 1952. *The Language of Morals*. Oxford: Oxford University Press.
Schroeder, M. 2010. *Noncognitivism in Ethics*. New York, NY: Routledge.

doi:10.1017/S0266267114000091

Evidence-Based Policy: A Practical Guide to Doing it Better, Nancy Cartwright and Jeremy Hardie. Oxford University Press, 2013, ix + 196 pages

Policymakers are increasingly privileging randomized control trials (RCTs) as the best evidence for causal claims. In an RCT one randomly assigns subjects into treatment and control groups. If the randomization

that determine what we have sufficient reasons according to the *Reasons One Has*. The problem with this is that this view assumes that there are pro tanto reasons. It is then natural to think that our pro tanto reasons determine what we have sufficient reasons to do rather than our beliefs about them.

is successful, then the difference in expected outcomes between the two groups provides an estimate of the average causal effect of the treatment on the outcome for the population in the study. RCTs can only show that a causal relation obtains in the study's population. To extrapolate a causal relationship from the study population to a different population (the *target population*) further assumptions are required. Specifically, one must assume that within the target population the causal factor is capable of playing a similar role to the one it plays in the study population and that this factor is accompanied by the background conditions required for it to bring about its effect. In *Evidence-Based Policy: A Practical Guide to Doing it Better*, Nancy Cartwright and Jeremy Hardie provide an extremely accessible guide for how policymakers can use their background knowledge to evaluate whether these assumptions are met in a particular case.

The authors model extrapolative inferences as having the form of a deductive argument, which they call *the effectiveness argument* (45).¹ The conclusion of the argument is that a particular factor that had a positive causal effect in the study population will have a positive causal effect in at least some members of the target population. This is a weak conclusion that is compatible with the policy having a net negative effect. Although it is not sufficient to justify implementing a particular policy, it is necessary. Policymakers must establish *at least* this conclusion before implementing a policy.

The effectiveness argument contains three premises. Premise 1 is that a factor, *X*, has a positive effect on an outcome in one population. This is what an ideal RCT establishes.² It is a mistake to infer from the first premise that *X* will have a similar effect in other populations; to make this inference two additional premises are required. Premise 2 is that *X can* play a similar causal role in the intended population. Premise 3 states that the *support factors* necessary for *X* playing this role are present in the target

¹ The effectiveness argument contains a set of assumptions that, if true, would justify an extrapolation. In addition to presenting these assumptions, the authors also give an account of evidence for when an extrapolation is justified. According to this account (19), any evidence *e* for a premise in the effectiveness argument is also evidence for the conclusion of the argument. This account is untenable, since evidential relevance is not, in general, transitive; just because *e* is evidence for a premise that is evidence (relative to an argument) for a conclusion does not entail that *e* is evidence for that conclusion (Hesse 1970). That a card is red is evidence that it is the queen of hearts, which entails that it is a queen. But that a card is red is not evidence that it is a queen. Fortunately, none of their claims about extrapolation depends on this theory of evidence.

² Some have criticized RCTs on the grounds that we have no assurance that the populations will be even approximately balanced in studies with small samples (Worrall 2007); see Reiss (2013, chapter 11) for discussion. Cartwright and Hardie purposely put this issue to the side. They assume that RCT are valid for the test population and ask whether their results can be generalized to other populations.

population. Support factors for X are other factors required for X to have its effect.

Premises 2 and 3 block two ways that a causal claim can fail to generalize from one population to another. To illustrate, consider the following example. A study in Tamil Nadu established that educating mothers promoted healthier infants. Unfortunately, a similar intervention in Bangladesh failed to improve infant health. Why? The authors suggest that what explains the difference is that in Bangladesh mothers-in-law (rather than mothers) are in charge of distributing the food in the family. Premise 2 does not obtain, since educating mothers does not play the same causal role in Bangladesh as it does in Tamil Nadu. Educating mothers-in-law, in contrast, could play a similar causal role.

In saying that educating mothers-in-law *could* play a similar causal role, the authors leave open the possibility that it might fail to do so were certain support factors absent. Educating the mothers-in-law might have no impact on infant health if the family lacks an adequate food supply. Causes do not typically work in a vacuum, but rather require other factors to bring about an effect. Cartwright and Hardie borrow J.L Mackie's terminology on which causes are *INUS conditions*. An INUS condition is an *Insufficient but Necessary* part of an *Unnecessary but Sufficient* condition for an effect. In other words, when X is an INUS condition for Y , Y obtains if and only if $BX \vee Z$ is true, where BX is a minimal sufficient condition for Y and Z is a disjunction of other minimal sufficient conditions for Y . Within this framework, one can easily see that B is a support factor for X , since only in conjunction with B does X bring about Y . The authors, like those concerned to identify causes, pick out one factor (X) as *the* cause, but there is no non-pragmatic distinction between causes and support factors. When X 's support factors are not present, premise 3 does not obtain and the policy will not have its intended effect.

Although premises 2 and 3 are intuitively distinct, one must refer to what the authors call *causal principles* to make this distinction precise. Here is the causal principle for Tamil Nadu³:

$$(TN) I = a_1 + a_2 I_0 + a_3 B_m E_m + a_4 Z$$

The lowercase ' a 's are coefficients and the uppercase letters are random variables – I refers to infant health, I_0 is infant health at an earlier time, E_m is education of the mother, B_m are the support factors for E_m , and Z represents all other causes of I that do not interact with $B_m E_m$. The equation represents how infant health would change if one were to intervene on one of the right-hand-side variables while holding the others

³ I have altered the notation of the causal principles in several ways to improve clarity. All of the coefficients are adjustable parameters, so a_1 in one principle need not have the same value as a_1 in another.

constant. The difference between a failure of premise 3 and a failure of premise 2 is as follows. Premise 3 is false if the value of B_m differs in the two populations. Premise 2 is false if the variable E_m does not appear in the causal principle for one of the populations. According to the authors, the educational intervention failed in Bangladesh because premise 2 was false. Bangladesh has the following causal principle:

$$(BD) I = a_1 + a_2 I_0 + a_3 B_{ml} E_{ml} + a_4 Z$$

E_{ml} refers to the education of the mother-in-law. Since (BD) does not contain a variable for E_m , premise 2 does not obtain.

But what determines whether E_m appears in Bangladesh's causal principle? Suppose that instead of treating (BD) as Bangladesh's causal principle, we used the following causal principle, which applies to both populations:

$$(C) I = a_1 + a_2 I_0 + a_3 B_m E_m + a_4 B_{ml} E_{ml} + a_5 Z$$

(C) contains both E_m and E_{ml} , so premise 2 is satisfied. Since the values of the support factors can differ between the populations, the effects of E_m and E_{ml} can differ as well (as, in fact, they do). If one represents Bangladesh using (BD), premise 2 does not obtain, but if one represents it as (C), it does. Absent some reason for choosing (BD) over (C), whether premise 2 obtains will be objectionably language dependent.

One reason to prefer (BD) to (C) is that if one models the difference between the populations with (C), one misses the fact that the policy's success depends not on which particular member of a family one educates, but rather on whether one educates the person with power over the family's food distribution. At one point the authors suggest that for each population, the relevant causal principle should look as follows:

$$(P) I = a_1 + a_2 I_0 + a_3 B_{pw} E_{pw} + a_4 Z$$

The subscript pw means "person with the power". I'd like to suggest that instead of considering (P) as an alternative to the distinct principles for each population ((TN) and (BP)), we should rather think of it as an alternative to (C). Like (C), (P) applies to both populations, but only (P) captures the common causal role played by the variables E_m and E_{ml} in (C).

Cartwright and Hardie talk as if one can determine whether premise 2 obtains by considering whether a factor appears in a population's causal principle, but populations do not wear causal principles on their sleeves. A population can have one causal principle relative to one set of measured variables, and a different principle relative to another set. If the treatment variable is 'mother's education' one causal principle applies, if it is 'education of the person in power' another does. The insight behind premise 2 is that choosing one variable set over another

can aid extrapolation. This insight has been neglected in the literature on causation. In order to make this point, however, one needs to separate the cases in which one compares two populations using a single model from those in which one compares two ways of modelling the same population. Premise 3 concerns the way that two populations could differ relative to a single way of specifying the variables. Premise 2 concerns the question of whether the factor under consideration would be a variable in the optimal model.⁴

Cartwright and Hardie suggest that a policymaker should perform two searches – a *horizontal search* and a *vertical search* – prior to implementing a policy. These searches correspond to premises 3 and 2, respectively.⁵ In a horizontal search, one considers whether the support factors in the study population obtain in the target population as well. In a vertical search, one thinks about whether one has described the cause at the right level of description.

How useful are these searches for determining whether a policy will succeed? Cartwright and Hardie describe an intervention to improve reading scores by means of reducing class size that was successful in Tennessee, but failed in California. A horizontal search would have revealed that California was missing support factors that were present in Tennessee. Specifically, unlike Tennessee, California had a shortage of both teachers and classroom space. In cases like this, where one knows some of the necessary conditions for a policy to work, horizontal searches are clearly useful. In situations where both populations have the conditions necessary for bringing about the effect, horizontal searches are less useful. Would it have been worth performing the intervention had California had enough teachers to implement it, but fewer teachers-per-student than in Tennessee? All else being equal, this would reduce the efficacy of the intervention, but all else is never equal. Perhaps the teachers in California are better on average and this compensates for the negative effects of the higher student-to-teacher ratio. Alternatively, maybe good teachers can only do so much if the classes are too big. Knowing what the support factors are is insufficient for determining how varying these factors changes the effect. For this reason, horizontal searches are better suited for ruling out policies in which support factors are absent than for justifying policies when they are present.

We've already seen an example of a vertical search in the Tamil Nadu case. The principle 'educate the person in power' extrapolates to

⁴ The question of which model is optimal is related to that of why models with fewer adjustable parameters are preferable to those with more, cf. Forster (2007), Forster and Sober (1994), Whewell (1840).

⁵ The authors do not explicitly note the correspondence between premise 3 and a horizontal search and between premise 2 and a vertical search.

Bangladesh; 'educate the mother' does not. The level of abstraction at which we describe a causal factor is important. How can we translate this insight into practical advice? By abstracting away from the properties of a population we end up with claims that apply to a wider range of populations, but not all ways of abstracting work equally well. In the Tamil Nadu case, switching from 'educate the mother' to the more general 'educate the person in power' worked, but why should we abstract to *this* general principle. Why not 'educate the person who supervises the child' (supposing that mothers play this role in Tamil Nadu)? This principle is as abstract as the one they suggest and it yields different advice for applying the lessons from Tamil Nadu to Bangladesh. How can one know which principle to adopt by looking just at Tamil Nadu? Without some guidance regarding which ways of abstracting are preferable, vertical searches do not yield a verdict on whether a causal relation extrapolates to the target population. Cartwright and Hardie identify this need, but they do not provide much guidance concerning how to satisfy it.

In horizontal and vertical searches a policymaker relies on her background knowledge in considering whether a policy will work. The authors say little about how to determine if one has reliable background knowledge in the first place. Consider the case Cartwright and Hardie discuss of a nurse who is able to quickly detect whether an infant has a certain disease (131–2). Since this disease is treatable only if it is detected early, the hospital would like to teach the nurse's skill to other nurses. Through careful deliberation, the nurse discovered that she detects the disease through monitoring whether the infant changes colour, shows heightened activity, and has reduced appetite. Assuming that the nurse is correct about how she makes her diagnoses, it will be possible to teach the other nurses how to make similar diagnoses by looking for these changes. In this case, the nurse was in fact correct, and the hospital was able to teach other nurses to make better predictions. Yet, even though the nurse's judgement was reliable, there is little reason to think that people's causal judgements are generally reliable, especially when one is implementing a complicated policy. This is why we need RCTs in the first place. It would therefore be unsatisfactory if extrapolation relied entirely on causal intuitions.

Fortunately, the nurse's hypothesis about how she makes correct predictions is testable. Consider the following model for the case (Figure 1):

This model represents the possible causal relations between the variables. It includes three measured variables on the path from the disease to the diagnosis. These measured variables are called *mediators*. The arrow going directly from the disease to the diagnosis represents all the causal paths between the treatment and the outcome that do not go through the measured mediators. Using *causal mediation* techniques,

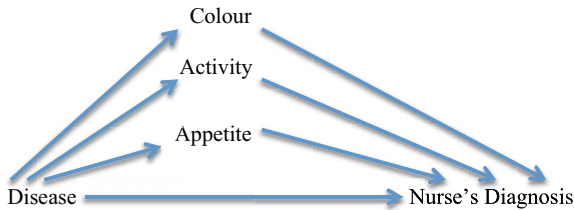


FIGURE 1.

one can determine how much each path contributes to the total effect. Doing so requires more complicated experimental designs than standard RCTs (Imai *et al.* 2011). Initially, one might think that one could measure the influence on a path going through a mediator by randomizing the mediator. The reason this does not work is that when one randomizes the mediator, one severs the causal connection from the treatment to the mediator. Randomizing the mediator enables one to estimate the effect of the mediator on the outcome, but this is not the quantity one wants to estimate in causal mediation. The desired quantity is the causal contribution of the path going from the treatment to the mediator to the outcome, but randomization disrupts this path. Despite this complication that arises in measuring the relative contributions of the different paths, they are in principle measurable (Pearl 2012) and social scientists have developed preliminary experimental designs for measuring them (Imai *et al.*, 2013). The nurse's hypothesis about how she makes her predictions can be verified by measuring the contributions of the paths going through the mediators.

Causal mediation techniques aid in extrapolation, since testing a hypothesis about the way a cause operates in the study population often enables one to predict whether it will work in other populations. If the nurse's predictions were largely based on infant colour, then other people capable of detecting these colour changes would probably make similarly good predictions. A central thesis of *Evidence-Based Policy* is that knowing *how* a cause works (which requires more than knowing the support factors and the causally relevant description) is essential to knowing whether it will generalize. But the authors say little about how we can learn what we need to know. Causal mediation techniques help answer this question.

Cartwright and Hardie intend their book as a practical guide for doing evidence-based policy better and succeed in their intention. They encourage policymakers to ask a broader set of questions than merely whether the policy has been shown to work somewhere. Without considering these additional questions, policymakers have little basis for thinking that a policy that worked elsewhere will also work in

their particular situation. Through horizontal and vertical searches, policymakers can use their background knowledge to avoid investing resources in projects that are unlikely to succeed. After doing these searches, one still needs to determine which projects are likely to succeed. I have here suggested that causal mediation techniques are one way to enhance extrapolation. Whether causal mediation is the most fruitful path remains to be seen.⁶

Naftali Weinberger
University of Wisconsin – Madison, USA

REFERENCES

- Forster, M. R. 2007. A philosopher's guide to empirical success. *Philosophy of Science* 74: 588–600.
- Forster, M. & Sober, E. 1994. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science* 45: 1–35.
- Hesse, M. 1970. Theories and the transitivity of confirmation. *Philosophy of Science* 37: 50–63.
- Imai, K., L. Keele, D. Tingley and T. Yamamoto. 2011. Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105: 765–789.
- Imai, K., D. Tingley and T. Yamamoto. 2013. Experimental designs for identifying causal mechanisms. (With discussions.) *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 176: 5–51.
- Pearl, J. 2012. The mediation formula: a guide to the assessment of causal pathways in nonlinear models. In *Causality: Statistical Perspectives and Applications*, ed. C. Berzuini, P. Dawid and L. Bernardinelli. Chichester: John Wiley & Sons.
- Reiss, J. 2013. *The Philosophy of Economics*. New York, NY: Routledge.
- Whewell, W. 1840. *The Philosophy of the Inductive Sciences: Founded upon their History* (Vol. 1). London: John W. Parker.
- Worrall, J. 2007. Evidence in medicine and evidence-based medicine. *Philosophy Compass* 2: 981–1022.

⁶ I would like to thank Nancy Cartwright, Dan Hausman, Fabienne Peter and Reuben Stern for helpful feedback.