New tools for thinking about causation

Arnaud Pocheville, Brett Calcott, Karola Stotz, Paul Griffiths

{paul.griffiths, arnaud.pocheville}@sydney.eud.au, {karola.stotz, brett.calcott}@gmail.com

http://griffithslab.org

Introduction

Three qualitative notions play an important role in current debates about causal explanation:

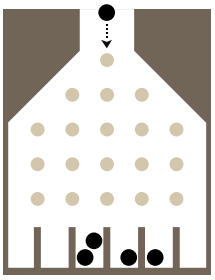
Specificity "the notion we are trying to capture is that the state of C exerts a fine-grained kind of control over which state of E is realized" (Woodward 2010, 305)

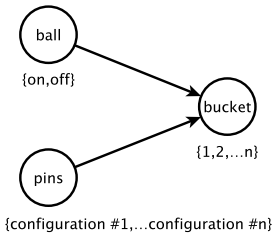
Proportionality "proportional in the sense that they should be just 'enough' for their effects, neither omitting too much relevant detail nor containing too much irrelevant detail" (Woodward 2010, 297)

Stability "the stability of this relationship of counterfactual dependence has to do with whether it would continue to hold in a range of other background circumstances" (Woodward 2010, 292)

Causal information theory applies Shannon information measures to the results of interventions on a causal graph. It allows us to clarify these three concepts and define measures of them.

Specificity



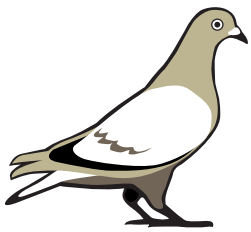


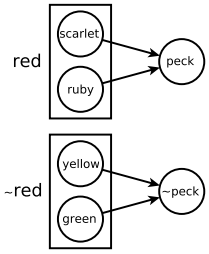
Spec: $I(E;\widehat{C}) = H(E) - H(E|\widehat{C})$.

Causal specificity is measured by the reduction in uncertainty about the value of the effect variable that results from intervening to set the value of the cause variable. (Variables with hats are set by intervention.)

C is a cause of E iff Spec > 0.

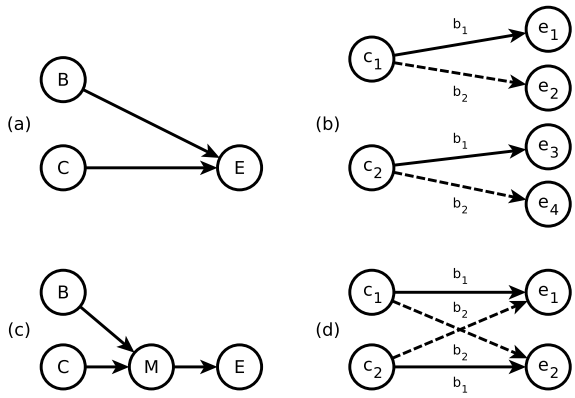
Proportionality





Proportionality constraint: given an effect variable E that is a target of intervention or causal explanation, a causal variable C should be discretised so as to minimise the entropy of C whilst maximising specificity for E.

Stability



(a) C and B are causes of E. B can be thought of as a set of background factors.
(b) The diagram represents the mapping from values of C and B to values of E. Here B provides additional causal information but does not interact with C.
(c) M is a variable representing the mapping from C to E. If B interacts with C, B will have some specificity for M.
(d) Here interaction is maximal, since B is necessary to determine to any degree the mapping from C to E.

So, the instability of C $\rightarrow$ E is the degree to which B interacts with C. This is measured by interaction information: $I(\widehat{C};E;\widehat{B}) = I(\widehat{C};E|\widehat{B}) - I(\widehat{C};E)$.

References

1. Ay, N. and Polani, D. Information flows in causal networks. *Adv. Complex Syst.* 11, 17–41 (2008).
2. Korb, K. B., Hope, L. R. and Nyberg, E. P. in Information Theory and Statistical Learning (eds. Emmert-Streib, F. and Dehmer, M.) 231–265 (Springer US, 2009).
3. Lizier, J. T. and Prokopenko, M. Differentiating information transfer and causal effect. *Eur. Phys. J. B* 73, 605–615 (2010).
4. Woodward, J. Causation in biology: stability, specificity, and the choice of levels of explanation. *Biol. Phil.* 25, 287–318 (2010)
5. Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H. and Knight, R. Measuring Causal Specificity. *Philos. Sci.* 82, 529–555 (2015).
6. Pocheville, A., Griffiths, P.E. & Stotz, K. Comparing causes: an information-theoretic approach to specificity, proportionality and stability. In Proc. of the 15th CLMPS. (in press)

This publication was made possible through the support of a grant from the Templeton World Charity Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Templeton World Charity Foundation.

Poster by Paul Griffiths, Arnaud Pocheville, and Stefan Gawronski

1 A primer on information theory

Information theory captures how information about the value of one variable reduces uncertainty about the value of a related variable[1].

When a discrete variable has only two values, its value can be known by answering a single question (yes or no). The answer conveys one unit of information (1 bit).

If the variable has 2^n equally likely possible elements, n dichotomous questions (n bits) are needed to determine the actual value. The quantity of information in the actual value is thus $n = \log_2(2^n)$. If each possible value has equal probability $p = 1/2^n$, knowing any actual value of the variable brings $-\log_2 p$ bits of information.

In general, the information gained by knowing the actual value of a variable is measured as an average over the probabilities of the different values. This quantity is the *entropy* of the probability distribution of the variable:

$$H(X) = -\sum_{i=1}^{N} p(x_i) \log_2 p(x_i)$$

(x_i represent values of the variable X and N is the number of different values). Uncertainty is maximised (*maximum entropy*) when each value is equiprobable.

If X and Y are two random variables, we can define the entropy of the couple, H(X, Y). This enables us to define the *conditional entropy*, representing the amount of uncertainty remaining on Y when we already know X:

$$H(Y|X) = H(X,Y) - H(X)$$

In a similar way, the *mutual information*, that is, the amount of redundant information present in X and Y is obtained by:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Conditional entropy is null, and mutual information is maximal, when Y is completely determined by X. Note that conditional entropy is generally asymmetric while mutual information is always symmetric. By contrast our measure of causal specificity is always asymmetric: $I(\widehat{X};Y) \neq I(\widehat{Y};X)$.

The relationships between these three different measures are represented in Figure 1.

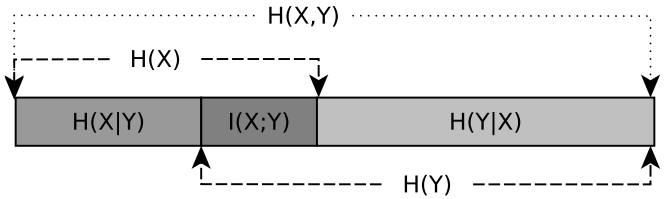


Figure 1: Diagram of the relationships between the different informational measures, entropy H(X), conditional entropy H(X|Y) and mutual information I(X;Y).

[1] Excerpt from [1]. For details see [2].

2 Causal power and actual causation

Measuring Spec with different probability distributions over $\widehat{C}$ corresponds to different views of causal specificity in the existing, qualitative literature.

Woodward's [5] fine-grained influence (INF) can be understood as intervening in an unbiased way, to make every value of C equiprobable:

INF: $I(\widehat{C};E)$, where the distribution of $\widehat{C}$ has maximum entropy.

However, another option is to construct a distribution which maximizes specificity, which need not maximize the entropy of C.

MaxSpec: $I(\widehat{C};E)$, where the distribution of $\widehat{C}$ maximises Spec.

Whereas INF measures how much influence C exerts on E in an unbiased set of intervention experiments, MaxSpec measures how much influence C exerts on E under ideal conditions. This is the 'causal power' [6] of C with respect to E and can be thought of as a measure of C's *potential* influence on E. We suggest MaxSpec best explicates the intuitive idea of the intrinsic causal structure of a system [7].

We can also measure how much difference a cause *actually* makes to an effect – 'specific actual difference making' [4] or (SAD). Our interventions mimic the observed distribution of C:

SAD: $I(\widehat{C};E)$ where the distribution of $\widehat{C}$ is identical to the actual distribution of C in some population.

This measure has been termed 'information flow' [3].

1. Griffiths, P. E. *et al.* Measuring Causal Specificity. *Philosophy of Science* **82,** 529–555 (2015).
2. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 2006).
3. Ay, N. & Polani, D. Information flows in causal networks. *Advances in complex systems* **11,** 17–41 (2008).
4. Waters, C. K. Causes that make a difference. *The Journal of Philosophy,* 551–579 (2007).
5. Woodward, J. Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy* **25,** 287–318 (2010).
6. Korb, K. B., Hope, L. R. & Nyberg, E. P. en. in *Information Theory and Statistical Learning* (eds Emmert-Streib, F. & Dehmer, M.) 231–265 (Springer US, Boston, MA, 2009).
7. Pocheville, A., Griffiths, P. E. & Stotz, K. in *Proceedings of the 15th Congress of Logic, Methodology and Philosophy of Science* (College Publications, London).