When local causes are more explanatorily useful

Pierrick Bourrat
Department of Philosophy
Macquarie University
2109 North Ryde, NSW, Australia
p.bourrat@gmail.com
pierrickbourrat.com

&

The University of Sydney
Department of Philosophy & Charles Perkins Centre
2006 Camperdown, NSW, Australia

Abstract
Madole and Harden plead for better integration of causal knowledge of different depths to understand complex human traits. Classically, local causes—a particular type of shallow causes—are considered less useful than more generalisable causes, giving a false impression that the latter causes are more useful and desirable. Using a simple example, I show that sometimes the contrary is true.

Main

Madole and Harden provide an insightful analysis showing that different types of causes can play different roles in helping us understand the etiology of complex traits. They make a point often underappreciated—namely, that average treatment effects (ATEs), such as those obtained from random control trials (RCTs), have some of the same limitations as those often attributed to heritability estimates and single-nucleotide polymorphism associations. Similarly, heritability estimates have often been charged with being only local parameters—when such a charge is rarely made against RCTs. Further, the charge of locality gives the false impression that a less local causal relationship—one that could be observed in a broader range of conditions—is always more useful than a local one. I show here that this conclusion does not follow; in some cases, which I illustrate with a theoretical example, local causal knowledge can be more useful for explanation and intervention than more generalisable knowledge.

Since Lewontin (1974), it is commonly accepted that heritability estimates originating from an analysis of variance suffer from the problem of locality: one estimate obtained in one population, even if unbiased, cannot and should not be extrapolated to other populations. This position, particularly its extreme form, is questionable (see Sesardic, 2005, pp. 75–80); however, generally, locality is considered a detrimental feature for establishing causal relationships. Being able to generalise a result is an important aspect of science, and locality stands in its way.

The problem of locality ties in with the analysis of causation provided by Woodward (2010) in the context of biological science (see also Bourrat, 2020, 2021, for discussions in the specific context

of heritability). Intervening on a variable (X) permits establishing whether X is a cause of another variable (Y) but not comparing it to different causes. To that effect, several dimensions of causal relationships have been proposed in the literature, among which is stability. Some causal relationships are less stable than others—that is, they break down more easily when the background (i.e., the variables of the system that are not X or Y) changes. Thus, locality and stability are inversely related. The more a cause is local—the less it would generalise beyond the population where it was established—the less stable it is.

An ATE measures whether X makes a difference on Y, while all other variables in the background are randomised. As such, if a difference in Y is observed, one can be confident that the relationship tested is as stable as there is variation in the background. More importantly, however, it tells us nothing about the specificities of this relationship, only that X makes, on average, a difference on Y (often with a certain magnitude). To uncover the specificity of the relationship, which might be desirable both for explanation and intervention, assuming we are dealing with two binary variables, one would need to fix some variable(s) in the background at one value and assess whether a difference in Y is still observed when X is intervened upon. The operation would have to be repeated for different background values.

The lesson from this simple case is that local or unstable causal relationships can have more value than more stable ones when the causal relationship is characterised by averages. This flies in the face of the commonly accepted view that more is better when it comes to causal stability and uncovers a well-known trade-off in the philosophy of modelling literature between generality and precision (Levins, 1966).

To make the case slightly more concrete (see Figure 1), suppose a global population of individuals with two possible genotypes (G_1 and G_2) in equal proportions. Each genotype is associated with either two phenotypes with the same probability, T_1 and T_2 (e.g., two levels of anxiety, 'low' for T_1 and 'high' for T_2), depending on the background with two randomised states in equal proportions, Z_1 and Z_2, that could represent the environment. Intervening on G in the global population would lead to the conclusion that the genotype is not a cause of T. However, suppose that (unknown to the experimenter) in a local population '1', containing the same proportion of the two genotypes but where only the background Z_1 exists, intervening on G would lead to a deterministic change in T (with $G_1 \rightarrow T_2$ and $G_2 \rightarrow T_1$). Further, in another local population '2', identical to the first except that only Z_2 exists, the opposite deterministic causal relationship would be established (with $G_1 \rightarrow T_1$ and $G_2 \rightarrow T_2$). In each local population, the conclusion would be that an individual's genotype causes the trait, but that a different genotype causes a different trait's value in the two populations.

This last conclusion would be more adequate than the conclusion reached in the global population that G does not cause anxiety. This is so because intervening on the background of some individuals in the population experiencing the 'wrong' environment could affect their phenotype and have the benefit of reducing their anxiety.

A similar demonstration using more complex variables than binary variables—although more tedious—could be devised to show that a small global average effect can be the result of two (or more) large effects established in more local backgrounds but going in different directions.

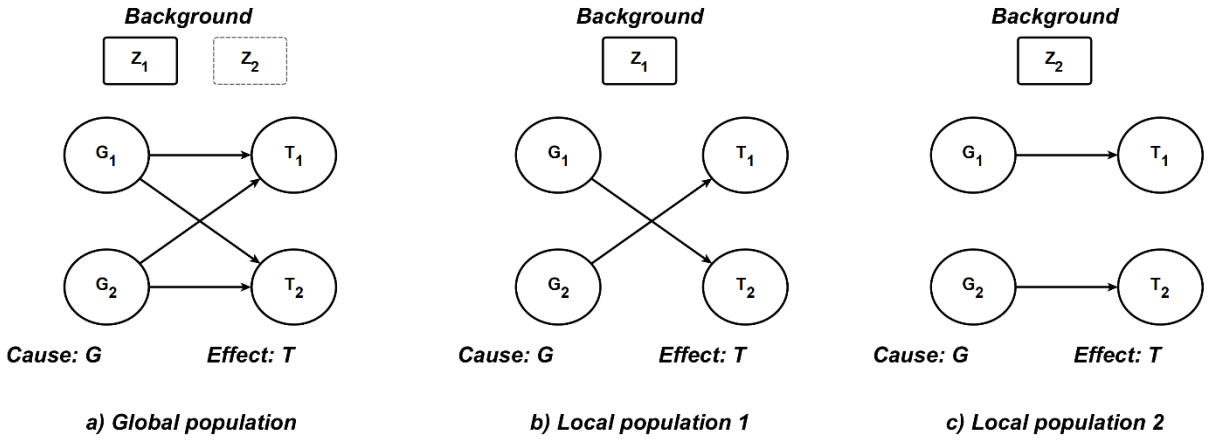

Figure 1: Causal relationship between G and T in three populations. a) In a global population with two randomised backgrounds (Z_1 and Z_2), G does not appear to be causing T: intervening on G, on average, does not affect the probability of expressing one of the two values of T. b) In the local population '1', with a constant background Z_1, intervening on G leads to a change in T, and it is established that this relationship is $G_1 \rightarrow T_2$ and $G_2 \rightarrow T_1$. c) In the local population '2', with a constant background Z_2, the same is observed as in the local population '1', except that the relationship is reversed so that $G_1 \rightarrow T_1$ and $G_2 \rightarrow T_2$.

The point sketched here speaks directly to Madole and Harden's urge not to dismiss shallow causes once integrated into a more thorough causal analysis.

Acknowledgement

I thank James Madole and Paige Harden for discussions on this topic.

Funding statement

This work was supported by an Australian Research Council Discovery Early Career Research Award (Grant ID: DE210100303).

Conflicts of Interest

None.

References

Bourrat, P. (2020). Causation and Single Nucleotide Polymorphism Heritability. *Philosophy of Science*, *87*, 1073–1083.
Bourrat, P. (2021). Heritability, causal influence and locality. *Synthese*, *198*, 6689–6715.
Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*, 421–431.
Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*, 400.
Sesardic, N. (2005). *Making Sense of Heritability*. Cambridge: Cambridge University Press.
Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*, 287–318.