Barbara Osimani
(University of Camerino)

# The challenges of context-specific causality and the virtues of evidence amalgamation:
## Grading causality rather then evidence

Much theoretical debate on causality may be subsumed under two sub-headings: 1) necessary and sufficient conditions for defining causality (metaphysical/semantic projects); 2) perfect indicators of causality to distinguish authentic from spurious causes (epistemological/ methodological) project. I would like to advance a third program, methodological in scope and purpose but grounded on metaphysics, which aims to provide imperfect indicators of causality and on this basis aid the amalgamation of heterogenous evidence for the purpose of causal diagnosis. This program stems from the specific challenges posed by the complex behaviour of causes in contexts and the difficulties they pose for the evaluation of interventions.

Indeed, the main problems regarding the evaluation of policy/health interventions may be summarized under the following headings:

1. External validity: whether the expected effect will occur in target population;
2. Extrapolation: whether the same kind of functional relationship holds for values of the variables which have not been observed;
3. Change of underlying causal structure by intervention itself;
4. "Side effects": unexpected, possibly unwanted harmful consequences of interventions.

I will concentrate on the fourth point by interconnecting with the other three when relevant. Current methodologies have been developed with a focus on identifying direct causes by isolating them, and by generally assuming modularity and linearity. However causes come seldom alone, and they act within a web of contributing/interfering/inhibiting factors. This causal web is generally characterized by different kinds of relationships (positive and negative feedback, attrition or threshold effects, back-up mechanisms – net-effect: no result –,

overcompensation mechanisms – net-effect:: opposite result; multiple realizability, moderating and mediating factors, low/high integration among subsystems; see also Joffe, 2011). Assessing the net (average) effect of the putative cause in isolation makes very little sense in this context. Cartwright (2009) elaborates on this by focusing on the nature of structural/path coefficients in causal models; e.g. β in $Y(u) = a(u) + \beta(u)X(u) + W(u)$, and emphasizes that β does not represent a single factor but a complex function of further factors that together fix whether and how much X contributes to Y: $\beta = f_1 (z_{11}, \ldots, z_{1n}) + \ldots + f_m (z_{m1}, \ldots, z_{mp})$. One corollary is that you will get the same result in the target population only if it shares the same mean value of β (i.e. the same distribution of different values of β, which in turn represent different combinations of values of z factors). Another corollary of this state of affairs is that you might have that together with $\beta = f_1 (z_{11}, \ldots, z_{1n}) + f_2 (z_{21}, \ldots, z_{2p})$ the intervention produces an intended outcome Y, whereas with $\delta = f_1 (z_{11}, \ldots, z_{1n}) + f_3 (z_{31}, \ldots, z_{3p})$ it produces (also) an undesired outcome (e.g. adverse drug reaction) Q:

$Y(u)\ c = a(u) + \beta(u)X(u) + W(u).$
$Q(u)\ c = k(u) + \delta(u)X(u) + Z(u).$

Factors which make up the causal coefficient and determine the success of intervention may be for instance ancillary conditions such as the right kind of information to the relevant target category; adequate communication style; or involvement and motivation, for educational campaigns; genetic make-up and clinical history, for pharmaceutical treatments and so on. Furthermore, interventions should be evaluated in the *system of policies* in which they are implemented: we can talk in this sense of "connectedness" of interventions, and this brings an additional element of complexity to the picture.

Hence, no intervention (whether in the health care, economic policy, or education) can be seen as immune to so called "externalities", or more generally, unintended and possibly unexpected and harmful consequences of an intervention aimed at some main effect. The case of adverse drug reactions is paradigmatic in this sense, but also ordinary policy interventions such as fiscal laws or educational campaigns provide stock examples of such phenomena. Thus, interventions are supposed to be evaluated on the basis of their net benefit: i.e. intended and desired minus unintended, undesired and possibly unexpected effects. However, methodological emphasis has been mainly placed on evidence for efficacy,

with evidence for possibly occurring negative outcomes being evaluated with the same methods developed for evaluating intended effects.

In fact, both philosophical causal theories as well as common methods for causal inference have been developed with a focus on "effects of causes" rather than on "causes of effects" (Holland, 1986). In this framework, causation has come to be identified as the net (average) effect of the cause independently of other (interfering) causes (*ceteris paribus* clause). Accordingly, methods for assessing causation have been generally developed with the goal of isolating the causal link under investigation from other influences, either physically or by holding the interference factors fixed (experimentally or through statistical adjustment). The focus on confounders and bias originates in this horizon. Criteria underlying evidence standards are focused on the quality of the causal signal precisely in the sense of eliminating noise, and abstracting cause from context. This is the reason why RCTs and meta-analyses of RCTs are ranked at the top of the hierarchy. Although better equipped with respect to the evaluation of context-specific causality, also alternative methodologies, such as (recursive) Bayesian nets, or agent-based modelling are designed with a focus on effects of causes rather than causes of effects.

I will elaborate on this point and insist on the distinction between grading evidence (see e.g. GRADE system or the Oxford CEBM levels of evidence, but also Clarke's et al. (2013) proposal to use both evidence of mechanisms and of probabilistic dependence for the purpose of grading evidence) vs. evidence amalgamation, with respect to the purpose of evaluating side effects of interventions. In particular, I will contend that rather than grading evidence, we should grade causal claims on the basis of the amount, scope and reciprocal support of evidence with respect to the target context.