Supplement to Theories of Change: Making Value Explicit

Additional notes on the definition of "Theory" as presented on page 39 in Powell, S. (2019). Theories of change: Making value explicit. *Journal of MultiDisciplinary Evaluation*, 15(32), 37-52.

- "Theory" in this sense refers only to causal theories about what affects what and not, for example, theories expressing general principles, classification or categorisation. "Theories" in this sense do not have to express any particularly profound, general, or complete piece of knowledge. They should simply express, in a Bayesian sense, our best guess at how things work.
- Also, although this is an explicitly causal, not correlational, approach, I have not defined "causal", "causality", etc; these are probably best left as primitive, undefined, concepts, see (Pearl, 2000, p. 96).
- With this definition, a Theory consists of variables, things which can take (or could have taken) different values or levels or scores. (It is more usual to talk about the "value" of a variable, e.g., the value of the variable "current year" is "2018", but here I will use the word "level" or "state" as I am already using the word "value" in a different sense). We can also include variables (actually, sets of variables) which stretch or repeat across time or bind together sets of cases such as student scores within a school class. So we can talk about a "variable" like "total number of hours of sunshine today" which is actually a set of many different variables, one for each day – or like "student score in the maths test", which can have one score for each of, say, 30 students, and so is really a set of 30 variables. This contrasts with the way "variable" is defined in statistics as a *set* of data points, but is consistent with the broader use of "variable" in other areas of mathematics, and consistent with Pearl's (2000) usage. Variables, so conceived, might have numerical levels, or ordered but non-numerical levels, or just the levels "true" and "false"; or they might be more vaguely-defined sets of more vaguely-defined possibilities. Perhaps we can even include variables which are a fuzzily-defined bunch of narrative possibilities: for example, we can think of the sentence "the impressive and insightful workshops led to the teachers' enthusiastic promotion of the new ideas in their teaching" as describing a causal relationship between two very rich, narrative-type variables. For example, instead of being charismatic and insightful, the workshops could have been confusing, or divisive, or simply pleasing, and so on, and anything in between. The teachers' response could have been reluctant, or inspiring, or any of a host of other things. Neither of these variables can be reduced to a simple set of numeric or ordered levels. Yet we can understand the causal story because we possess enough psychology, or at least enough folk psychology, to see how the state of the first variable leads to the state of the second variable.
- A Theory is usually presented in the form of a kind of network of arrows or "directed graph" linking the variables, so the set of all the variables pointing to some variable are those which influence it causally. As Pearl (2000, p. 13) emphasises, the main advantage of presenting a Theory in the form of a network is that it restricts the direct causal influences of a given variable only to its parents – we can predict the state of that variable only by referring to its parents, without worrying about all the other variables in the Theory. There are two important assumptions: 1) all the variables in a Theory are connected into one network, i.e. there are no separate fragments; 2) we only allow graphs without loops, i.e. we restrict the definition of Theory to so-called Directed Acyclic Graphs (Pearl, 2000, p. 13). This does not mean that we forbid "feedback loops", because a such a loop as usually understood actually comes back to a *different variable* – identical to the first in every way but at a different point in time, which makes it a different variable. Dealing with this is beyond the scope of the present article.
- In most practical cases, at least some parts of the composite Theory will be derived from different

sources in different contexts with different kinds, and quality, of evidence. More details about each causal connection, whether vague or precise, should ideally be provided as a constituent part of the Theory in order to answer questions like: "if this parent variable is set to this level …, what level will the child variable take?". These influences can be noted directly on the child variable or the incoming arrows, or elsewhere, as appropriate. Such influences can be incomplete, probabilistic, non-linear, fuzzy, or, as in Figure 1, involve other unspecified and unknown factors. Scriven (2012) argues persuasively against the idea that "approximations [have] no place in laws of nature". All we need for a Theory about the influence of one thing on another, is reason to believe that manipulating the influence Variable(s) might, sometimes, make some difference to the consequence Variable.

- The *direction* of the connections (most frequently, positive or negative) is often not explicitly stated in theories of change, and nor are they in Figure 1; often, as here, it is safe to interpret them as, broadly speaking, positive: "*The more* teachers are exposed to the new teaching method, *the more* creative teachers are, and in turn *the more* creative teachers are, *the more* student academic achievement will improve".