Where there is no single Theory of Change: The usefulness of Decision Tree models

Rick Davies, Cambridge, April 2013

Theories of Change (ToC) are in the limelight. In 2012 reviews of the use of ToCs were commissioned by DFID, CARE, and Comic Relief, and a number of training courses specifically about the use of ToC were also put on offer. An explicit ToC can be a great aid to evaluation. At best, it clarifies expectations of outcomes and how they will be achieved, in a way that is evaluable. But ToC have at least three limitations. Firstly, because of our limited capacity to understand complex models, few ToCs adequately represent the complexity of reality. Secondly, large projects with many stakeholders mean multiple perspectives on the design of a project, which are often not readily reconciled in a single ToC. Thirdly, even in the case of apparently simple interventions, like cash transfers, every additional design parameter that needs to be correctly specified doubles the size of the combinatorial space in which a best design might be found. With as few as 10 possibly relevant attributes (e.g. relating to context, partners and procedures) there are 2^{10} or 1,024 possible best designs. A project's official Theory of Change might represent a small sub-set of these combinations. In parallel to the limits of a Theory of Change approach, the continuing expansion of computer memory capacity is providing an ever expanding volume of data available to be analysed, an issue touched upon by Robert Kilpatrick's discussion of "big data" in the March 2013 issue of Connections.

Large companies and big science enterprises have responded to the scale and complexity of their work by developing an array of data mining tools designed to search for patterns (which can described in terms of clustering and association rules). These include algorithms designed to learn and to generalise from one data set to another. One type of algorithm that is widely used for such "predictive modelling" is called a Decision Tree. A Decision Tree model can summarise how different combinations of conditions are associated with different kinds of outcomes. Applied to an existing set of data it provides a summary classification. When this derived classification is applied to a new but comparable set of data it provides testable predictions about what outcomes will be associated with what combinations of conditions.

It is my view that Decision Tree models have considerable potential as an analytic tool to be used by evaluators. There are five main arguments for this proposition, which I have provided in brief summary form here.

Firstly, Decision Tree models can represent **sophisticated forms of causality**. They can describe causal packages involving multiple conditions and a single Decision Tree model can contain multiple packages of this kind. This is what Ragin (1989) has described as "multiple and conjectural causation". This capacity enables what systems theorists describe as equifinality - multiple routes to the same end. Causal asymmetry can also be represented, recognising that the absence of an outcome is not always simply because of the absence of conditions required for its presence. There may be other inhibiting factors at work. A simple example is shown in Box 2 below.

Within this capacity to describe "multiple and conjectural causation" is the ability to visibly differentiate sufficient and/or necessary causes and INUS causes (insufficient but necessary parts of condition that is unnecessary but sufficient). This ability then provides the means to differentiate attribution from contribution, in the sense used by Mayne (2012), and to go one step further and actually enumerate the extent to which a given condition is an influential contributory cause, relative to others in the model.

Somewhat surprisingly perhaps, the second argument emphasises the **user friendliness** of Decision Tree models as representations. This advantage is relative to the results of regression models, which are expressed in mathematical notation and the results of Qualitative Comparative Analysis (QCA), which are expressed in Boolean logic statements. In a Decision Tree the outcomes of interest are represented by the “leaves” and the conditions that are associated with a given outcome are represented by the branches and sub-branches leading to that leaf. A simple example is shown in Box 1 below. Ease of understanding is important if results of analyses need to communicated to non-expert audiences

The third argument emphases **testability**. It is standard practice that when a Decision Tree model is developed, it is developed using only part of the data set available, which is commonly around 60% (randomly selected). At this stage the focus of assessment is on the adequacy of the Decision Tree model as a descriptive model. There are a number of different measures that can be applied, regardless of the specific contents of the model. One is the percentage of cases correctly classified within each leaf. A good model correctly classifies a high percentage. Another is the ratio of cases correctly classified compared to their overall incidence, called “uplift”. Another is the simplicity of the tree. Although accuracy can always be improved by having more and more sub-branches this more complex structure risks “over-fitting” the model and weakening its predictive validity when later tested against the remaining “test” data set. It is the latter measure that really matters with many commercial applications of Decision Tree models.

The fourth argument is about **inter-operability** (the ability of diverse systems to work together). Decision Tree models are based on simply structured data sets, where cases are listed row by row and their attributes and associated outcomes of interest are listed column by column. The same kind of data set is used by QCA and is also usable by Social Network Analysis tools, which is another approach to analysing complex contexts and interventions. The results of QCA analyses, expressed in Boolean notation, can also be expressed as Decision Tree diagrams. Because QCA and Decision Tree algorithms use different methods of analysis but can be applied to the same data set, their combined use provides an opportunity for triangulation. Finally, it is also possible to integrate data from control and intervention groups and compare differences in outcomes within and between different configurations of conditions.

Related to the above argument, Decision Tree models are **scalable**. They can be used with very small numbers of cases and very large numbers of cases. With small numbers of cases Decision Tree models can be developed using ethnographic methods, as done by Gladwin (1989), and potentially with other more participatory methods. For large numbers of cases there are now a range of software packages available, including open source packages that include other data mining algorithms in addition to Decision Trees.

Caveats and counter-caveats: Decision Tree models are pre-eminently *predictive* models. Explanatory models can be seen as a sub-set of predictive models. They should be able to deliver testable and hopefully accurate predictions. But not all predictive models need to be explanatory models. It is useful to be able to identify poor households based on a predictive model, but that model does not have to also provide a good causal explanation of why people are poor. But where there is reasonable care with the selection of case attributes a predictive model can also provide a plausible causal model. For example, the re-analysis of the Krook data set below, where attributes were chosen with plausible causal links in mind. The validation of the casual content of a model then requires attention to the *mechanisms* that might link the associated attributes. These will come from close attention to the workings of individual cases in the data set, including the presence of “smoking guns” and “hoop tests”, as described by Mahoney (2012) in his recent paper on “The logic of process tracing tests in the social sciences”.

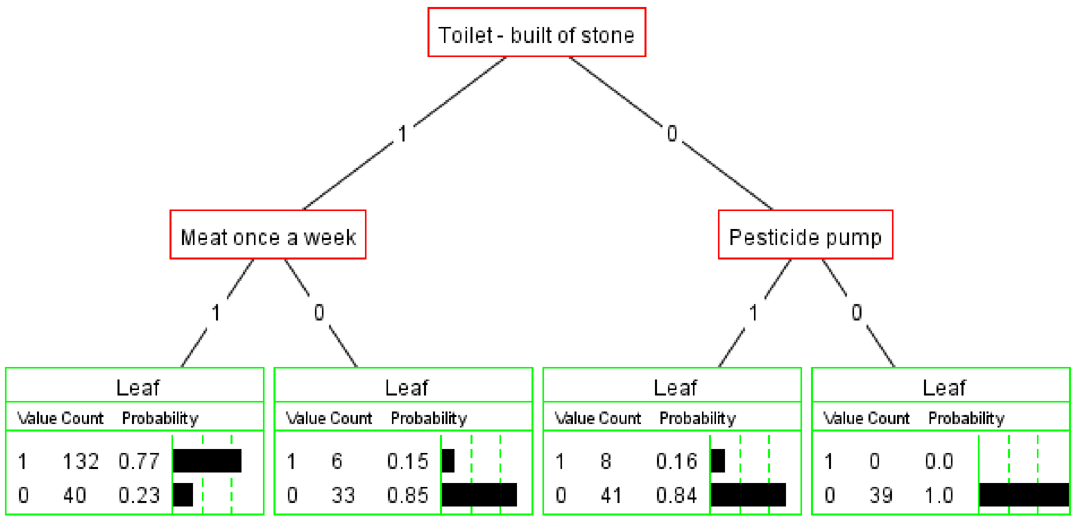
In conclusion, for readers who are now interested, I could recommend two different texts. One is Siegel’s (2013) popular science account of predictive analytics and the other is Goertz and Mahoney’s (2012) A Tale of Two Cultures, a comprehensive exposition of the value of a set-theoretic approach to qualitative analyses of causal process, which is consistent with the use of both Decision Tree models and QCA.

Box 1: Decision Tree model based on household poverty data from Ha Tinh province of Vietnam in 2006

The simple Decision Tree shown below was generated by an analysis of a randomly selected 50% of 596 responses to a Basic Necessities Survey that asked about 23 aspects of people’s households. Reading the tree from the top, we see that if a household has “a toilet built of stone” and they “eat meat once a week” then there is a 77% probability they will be non-poor. On the other hand, if a household has neither there is a 100% probability they will be poor.

When this simple model was tested against the second half of the data set its overall accuracy was 82%. Further small improvements were made by developing more branches, using additional attributes to split the three existing “leaves” into purer sub-sets.

Given that the survey asked about 23 household attributes, there were 2^{23} possible combinations that could have been the best predictors of a household’s poverty status i.e. 8,388,608 possibilities



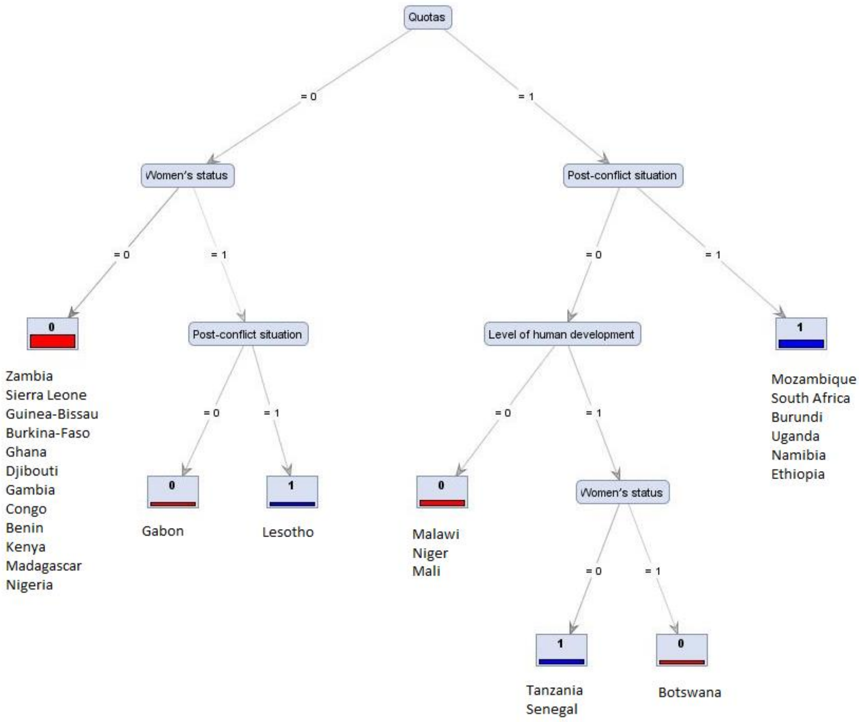
Link = 0 = Households did not have the above attribute Link = 1 = Households did have the above attribute	Leaf value = 0 = Household was poor (low BNS score) Leaf value = 1 = Household was not poor (high BNS score)

Box 2: Decision Tree model of conditions favourable to high level of women's participation in parliament in Africa

This model was the results of a re-analysis of Krook's (2010) QCA data set and analysis. The results were in agreement with Krook. Bear in mind however that there was no questioning of the relevance or accuracy of the original set of attributes chosen for analysis.

The presence of a "post-conflict situation" was neither necessary nor sufficient condition for high levels of women's representation in parliament. But it was a contributory factor in seven out of the eight cases. "Women's status" was a contributing factor in two of the eight cases.

The conditions associated with 12 cases of low levels of women's representation were not a mirror image of those associated with conditions associated with the six cases high levels of women's representation. They were asymmetric.



Key: 0 = attribute not present, 1 = attribute present.
Red = countries with low levels of women in parliament,
Blue = countries with high levels.

References cited

The arguments in this article have been developed in more detail in a longer paper of the same name, available online at http://mande.co.uk/2012/uncategorized/where-there-is-no-single-theory-of-change-the-uses-of-decision-tree-models/

Gladwin, C. H. 1989. Ethnographic Decision Tree Modelling. Vol. 19. Sage Publications, Incorporated. http://books.google.co.uk/books

Goertz, Gary, and James Mahoney. 2012. A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences. Princeton University Press.

Kirkpatrick, R. (2013) Evaluation for the 21st Century, Evaluation Connections, March: 3-4

Krook, M. L. 2010. “Women’s Representation in Parliament: A Qualitative Comparative Analysis.” Political Studies 58 (5): 886–908.

Mahoney, James. 2012. “Mahoney, J. (2012). The Logic of Process Tracing Tests in the Social Sciences. 1-28.” Sociological Methods & Research XX(X) (March 2): 1–28. doi:10.1177/0049124112437709.

Mayne, John. 2012. “Contribution Analysis: Coming of Age?” Evaluation 18 (3) (July 1): 270–280. doi:10.1177/1356389012451663.

Ragin, Charles C. 1989. The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. University of California Press.

Siegel, Eric. 2013. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. John Wiley & Sons.