Influences and Connections Between System Dynamics and Decision Analysis

Ken R. McNaught

Cranfield University

Applied Mathematics & Operational Research Group, Engineering Systems Department, Cranfield University, RMCS Shrivenham, Swindon SN6 8LA, UK.

Tel: +44(0)1793 785288
Fax: +44(0)1793 784196
E-mail: K.R.McNaught@rmcs.cranfield.ac.uk

Abstract: System Dynamics and Decision Analysis have much in common, but usually address different problem situations. Nonetheless, there appears to be scope for some cross-fertilization of ideas. After briefly considering what the two fields have in common, we then address their differences. Firstly, the two fields emphasise different aspects of a problem situation, leading to models with a different focus. Secondly, the representation of time is quite different. The meaning and representation of causality and dependence are key to both fields, suggesting that there should be areas of mutual interest which could be explored further in a cooperative vein. Uncertainty, on the other hand, has traditionally been of more concern in Decision Analysis than in System Dynamics. It is suggested that there are occasions when a more thorough treatment of parameter uncertainty within SD modelling would be beneficial, and a simple example is provided for discussion.

Keywords: graphical model, influence diagram, uncertainty, conditional independence

WHAT DO SYSTEM DYNAMICS AND DECISION ANALYSIS HAVE IN COMMON?

Both fields are concerned with providing practical approaches to support decision makers across a wide spectrum of application areas and at all levels, but particularly at higher managerial levels.

Both employ quantitative and qualitative models to address a wide variety of problem situations from the well-structured (hard) to the messy (soft).

Both maintain a unique identity but are embraced by the common umbrella of Operational Research / Management Science.

WHERE DO THEY DIFFER?

1. Emphasis

The emphasis of SD is on understanding the behaviour of systems in terms of their structure. A model of the system is developed and used to identify potential system improvements, often in a fairly ad hoc fashion, reminiscent of the stereotypical engineer.

The emphasis of Decision Analysis (DA) is on identifying and evaluating potential alternatives available to a decision maker. Often, a model describing the problem is developed which is then used to identify solutions which are in some sense optimal. By its nature, DA encompasses a wider range of techniques and models than SD. Decision trees and multi-criteria approaches such as multi-attribute value theory (MAVT), and the Analytic Hierarchy Process (AHP), are the kind of models which are likely to be most familiar to the SD community, since these have long been the most widely taught DA subjects in OR/MS courses and in other quantitative methods courses. In recent years, however, approaches such as Markov decision processes, Bayesian belief networks and influence diagrams have emerged to stake a claim for inclusion and, of these, the latter has been the most successful in this regard. The decision theoretic influence diagram, popularised by the Stanford school of decision analysis, particularly Howard and Matheson (1984) and Shachter (1986), addresses the same kind of sequential decision problems as decision trees, but offers an alternative representation which avoids the 'bushy mess' of medium to large scale decision trees.

By virtue of having so much in common, it is perhaps not surprising that both the System Dynamics and the Decision Analysis communities separately coined the same term, 'influence diagram', to describe one of their graphical models. The common name has undoubtedly caused a degree of confusion, particularly in the minds of some students! While both use nodes to represent variables of interest and directed arcs to represent the main influences between variables, the two constructs are quite different. The most obvious visual difference is that the decision theoretic influence diagrams (from now on abbreviated to DTIDs) are acyclic, i.e. they do not permit loops, unlike the influence diagrams of SD (from now on referred to as Causal Loop Diagrams (CLDs). The relationships represented by the influences or arcs are also of a different form, to be described later.

2. The representation of time

Sometimes, decision analytic models are criticised for not handling time adequately. However, while time may not be represented as explicitly as in SD models, it is certainly not always ignored. Decision trees, for example, display a very clear chronology from left to right across the diagram. Time is not measured in clock time, but by the state of a decision maker's knowledge. As time passes, potentially more information becomes available to the decision maker. Important quantities which were previously uncertain (as far as the decision maker is concerned) become certain as events take place and questions are resolved.

DTIDs do not immediately display such a clear chronology as they emphasise dependencies between variables, but a DTID does take account of time. Furthermore, a DTID can be converted to an equivalent decision tree - the conversion is achieved by changing the focus to time dependence. Arcs going into a decision node always denote time precedence, although other arcs denote probabilistic dependencies as well as deterministic relationships (particularly payoffs). Those that enter decision nodes display which decisions were taken before that one, and which uncertainties, represented by chance variables, have been resolved at the time the decision is taken.

Although DTIDs are closely related to Bayesian belief networks (BBNs), one of the chief distinctions is that BBNs are essentially static models. The greater flexibility of DTIDs follows from the fact that they permit additional types of nodes (decision nodes and value nodes) and, as alluded to above, the meaning of an arc depends on the node types it connects. Nevertheless, an arc between two chance nodes has the same meaning in both - probabilistic dependence.

A BBN provides an efficient representation of a joint probability distribution across some domain of interest. Evidence and observations on the variables in the network can be entered and their effect on the probability distributions of other variables in the network are then calculated. Although the type of systems represented by BBNs are usually stationary, system evolution and time-lagged cycles can be represented, and lead to dynamic Bayesian networks.

The kind of situation where time plays no part is where a single decision is to be made and there is no uncertainty to be resolved - then the model will be static. Multi-criteria approaches such as MAVT, MAUT and AHP usually fit this description. Clearly, so do many of the classic OR approaches such as linear programming, but these fall outside the conventional remit of decision analysis, so will not be considered here.

3. The representation of causality and dependence

Both SD and Decision Analysis employ graphical representations of causality and dependence. While SD thrives on feedback loops, such cycles are not permitted in DTIDs or in Bayesian belief networks - logically, they are not required.

As stated previously, Bayesian belief networks are closely related to DTIDs. An arc between two chance nodes has the same meaning in both. It denotes probabilistic dependence, or relevance. It can represent causality, but the relationships do not have to be causal. (See Pearl 1988 for a discussion of causality, relevance and dependence.) These dependencies are usually displayed in the direction of causality, if that direction is clear, or otherwise in the direction that leads to the most natural conditional probability assessment. In the solution of the model, however, the structure of the DTID can be manipulated, providing some basic rules are observed (see Shachter 1986). Some arcs can be removed or reversed. An arc from variable X to variable Y means that Y is conditioned on X, so reversing such an arc leads to X being conditioned on Y, and the required conditional probability distribution is calculated using Bayes' theorem.

By contrast, in SD, the focus is more on how the system under consideration changes through time. The causal relationships between the variables and parameters are captured through a set of equations that define the model. The equation for a particular variable, X, will be a function of all the variables and parameters which have arcs that enter X.

4. The representation of uncertainty

Within the standard tools of decision analysis, such as decision trees, DTIDs and Bayesian belief networks, uncertainty is represented by chance nodes. Each chance node is an uncertain variable with an associated probability distribution.

Uncertainty is usually unrepresented in SD models, which are traditionally deterministic. When it is discussed, it is usually in the context of random sampling within an equation (e.g in a production and inventory control model, daily demand might be sampled from a probability distribution) in which case the number of sampled values is the number of times the variable is updated during a run. Actually, the vector of sampled values equates to only a single realization, however, as the system is in a different state every time a new value is sampled. This kind of sampling represents random variability within a run.

A second type of uncertainty relates to parameter uncertainty and is often handled by 'sensitivity analysis' (facilitated by the most popular SD software packages). This is epistemic uncertainty and is potentially more serious for the conclusions drawn from a study. A single SD model run requires the assumption that every parameter takes a particular known value at time zero. For some parameters, however, there may be a considerable degree of uncertainty about what their true value is. Furthermore, some of these parameters might be unconnected in our model but actually be dependent in a probabilistic sense (i.e. their true values are correlated).

There appears to be little in the SD literature regarding the handling of this uncertainty, but there does not appear to be any problem in principle with directly applying the methods of uncertainty and risk analysis to the parameters of SD models.

DOES DECISION ANALYSIS HAVE ANYTHING TO OFFER SYSTEM DYNAMICS?

1. Integration of multi-criteria approaches with SD model output

There does appear to be considerable scope for the use of multi-criteria approaches in SD. While the formulation of an objective function is sometimes frowned upon in SD circles, and for good reason, it nonetheless provides a mechanism to summarize what is often complex, multivariate output from a model. Such a mechanism can only be enhanced by taking on-board best-practice guidelines from the field of MCDA. For example, the use of nonlinear value functions and swing weights in the formulation of objective functions would seem to be a step in the right direction. As yet, there are few examples to inform SD practice, a notable exception being provided by Santos, Belton and Howick (2001). They consider the integration of SD and MAVT in the context of performance measurement systems.

A possible explanation for the relative paucity of successful examples of the integration of these methods is provided by Andersen and Rohrbaugh. Among other things, they highlight the difficulty of longitudinal trade-offs. While the nature of some systems might make it sensible to trade off only the final values which occur at some meaningful point in time, more often than not, there will be no such endpoint of interest. Then, some kind of averaging or integrating or discounting equation will be needed for each measure of interest. Even then, some people will prefer a steady improvement in performance to mild decline followed by steeper improvement or steeper improvement followed by mild decline. They may also prefer lower frequency and lower amplitude oscillations in performance, regardless of any calculated numerical advantage of a more volatile pattern. The issue of long-term versus short-term improvement is another interesting trade-off which they highlight, having obvious implications for political decision-makers who are often more concerned with short-term objectives.

Coyle (2002) integrates SD with an MCDA aproach in an entirely different way. He suggests the use of the Analytic Hierarchy Process (AHP) to help the modeller decide whether a qualitative model should be further developed ino a quantitative model.

2. The notion of conditional independence

Conditional Independence is a notion which is usefully employed in probabilistic graphical models, such as Bayesian belief networks and DTIDs. Put simply, if A, B and C are uncertain events, then if $P(A \mid B, C) = P(A \mid C)$, we say that A and B are conditionally independent given C. In other words, provided we are certain about C, additional information about B does not affect our belief about A. Such a relationship would be represented graphically in a BBN as shown in Figure 1.

Nadkarni and Shenoy (2001) recently suggested that this concept could be useful in causal mapping. It would appear that it could also prove to be a useful concept when discussing and formulating CLDs. Discussions of causality and dependence are often fraught with misunderstandings. So too are the notions of direct and indirect influences. We have probably all struggled with clients or students who are tempted to put too many direct influences onto a CLD, leading to clutter, confusion and misrepresentation of the underlying system. For example, if while considering part of a system, a client produces a graph such as that in Figure 1, but with an additional arc from B to A, they could be asked whether the value of A would change when C was fixed, but B could take different values. If the reply is negative, then the extra arc is superfluous and B's influence on A is indirect, mediated by C.

Thinking about the function or equation you would use to represent the variable in a quantitative SD model helps to clarify whether an influence is direct or indirect. So too does the notion of conditional independence, but in a way that does not require equation formulation. Hence, it may prove to be a useful tool during model elicitation with a domain expert who is not familiar with quantitative SD or comfortable with equation building.

3. Accounting for parametric dependencies and uncertainty

As Ford and Sterman (1998) have noted, 'The literature is comparatively silent, however, regarding methods to elicit the information required to estimate the parameters, initial conditions, and behaviour relationships that must be specified precisely in formal modeling.' This is particularly true when it comes to considering possible correlations / dependencies between parameter values. It is almost as if parameters / constants are not considered worthy of further thought since their values cannot change during a model run, and different values can always be tried out later, perhaps as part of a sensitivity analysis. However, that ignores their obvious importance in defining the scenario. A sensitivity analysis in which several parameters are each allowed to take a range of different values is effectively assuming that each parameter value is independent of all the others. Unless this is very carefully thought out, it may result in a large number of inappropriate scenario configurations, which are all valued equally. Consider the following simple example of a recruitment model.

In this simplified model of the recruitment process, we only consider a single component level, workforce size, as shown in Figure 2. This is sufficient to demonstrate the essence of the approach being advocated. The SD model in Figure 2, then, contains this single level, together with an inflow - recruitment rate, and an outflow - leaving rate. The leaving rate is the product of staff turnover (defined as the proportion of the staff leaving each month) and workforce size. Recruitment rate is governed by proportional control - the discrepancy between the actual workforce size and a target workforce size is divided by average recruitment time. Consequently, there are three parameters in the model - staff turnover, average recruitment time and target workforce size.

If we now consider possible dependencies between these parameters, in particular between staff turnover and average recruitment time, further evocative, background parameters become relevant - in particular, the attractiveness of the company to employees and the level of unemployment. For example, when unemployment is high and company attractiveness is high, you would not expect high staff turnover or long average recruitment times. It may be, however, that the model only includes turnover rate and recruitment time as explicit parameters. The background effects of unemployment and company attractiveness might be considered implicitly and so play an evocative role in the assignment of explicit parameter values (see Howard 1989 for a discussion of evocative variables in the context of decision theoretic influence diagrams). On the other hand, they might not be considered at all, possibly resulting in highly implausible combinations of parameter values depending on what the actual background conditions are.

The joint probability distribution for the two explicit model parameters and the two background parameters can be elicited most naturally by concentrating on the causal structure of the domain. This is displayed in the form of a BBN in Figure 3. In order to fully specify the BBN, prior marginal distributions are required for the two parentless nodes, company attractiveness and unemployment, while for each node with parents, a conditional distribution is required for each possible combination of parent states. In this example, it has been assumed that each parameter can be in one of three states, corresponding to high, medium and low values. The prior marginal distributions and the two conditional distributions used for this example are given in Tables 1 to 4. This leads to the joint probability distribution shown in Figure 4 (the values on the belief bars are percentages).

Table 1. Prior distribution for Unemployment.

Unemployment:	low	medium	high
Probability:	0.25	0.5	0.25

Table 2. Prior distribution for Company Attractiveness.

Company Attractiveness	low	medium	high
Probability	0.2	0.6	0.2

Table 3. Conditional distribution for Recruitment Time (R.T.) given Company Attractiveness and Unemployment.

Company Attractiveness	Unemployment	P(R.T. = low)	P(R.T. = med)	P(R.T. = high)
high	high	0.9	0.09	0.01
high	medium	0.75	0.23	0.02
high	low	0.5	0.4	0.1
medium	high	0.7	0.28	0.02
medium	medium	0.25	0.5	0.25
medium	low	0.2	0.4	0.4
low	high	0.2	0.6	0.2
low	medium	0.1	0.4	0.5
low	low	0.01	0.19	0.8

Table 4. Conditional distribution for Staff Turnover (S.T.) given Company Attractiveness and Unemployment

Company Attractiveness	Unemployment	P(S.T. = low)	P(S.T. = med)	P(S.T. = high)
high	high	0.9	0.09	0.01
high	medium	0.8	0.18	0.02
high	low	0.6	0.3	0.1
medium	high	0.7	0.25	0.05
medium	medium	0.25	0.5	0.25
medium	low	0.2	0.5	0.3
low	high	0.2	0.6	0.2
low	medium	0.02	0.28	0.7
low	low	0.01	0.09	0.9

If we are interested in a scenario where unemployment is high and company attractiveness is high, then these parameters can be assumed to be at these levels with certainty. The resulting joint probability distribution is shown in Figure 5. Alternatively, if unemployment is low and company attractiveness is low, the joint distribution shown in Figure 6 is obtained. The joint distributions of the included model parameters can then be used to identify the most likely combination of parameter states given the assumed background conditions. They could also be used to attach probabilistic weights to the model outputs resulting from the various possible parameter combinations. For example, staff turnover and average recruitment time are conditionally independent given unemployment and company attractiveness. Hence, once these background parameters have been fixed, the joint probability distribution of staff turnover and average recruitment time is obtained simply by taking the product of the two marginal distributions. This is displayed in Table 5 for the situation corresponding to Figure 6.

Table 5. Joint distribution of Staff Turnover and Recruitment Time given Unemployment = low and Company Attractiveness = low.

	Rec.Time = low	Rec.Time = medium	Rec.Time = high
S.T. = low	0.0001	0.0019	0.008
S.T. = medium	0.0009	0.0171	0.072
S.T. = high	0.009	0.171	0.72

Another possibility is that a background variable, such as company attractiveness, is considered important, but is also unknown and difficult to observe or measure, i.e. it can be considered a latent variable, unlike unemployment, for example. In such a case, the quantitative modeller might be tempted to leave it out for fear of having to concoct dubious equations or tables to describe its effects. An alternative approach in this case, however, would be to directly model the dependence relationships between the unobservable and observable variables, by eliciting the appropriate conditional probability distributions from the system experts. Then, fixing the observable parameters at levels which reflect the scenario to be modelled allows the appropriate joint distribution of explicit model parameters to be calculated and subsequently used in attaching likelihoods to the various scenarios. This is in contrast to allowing unobserved parameters to take virtually any combination of values from their permitted ranges, and then treating the resulting assortment of model runs as though they are all equally likely when in fact many are highly unlikely, given the observable parameters. For example, if in the situation we are interested in, unemployment is known to be high and staff turnover is known to be medium, but we have no information on average recruitment time or company attractiveness, we enter the information that we have got, as in Figure 7, and then obtain the probability distribution of the unobserved parameters given the observed ones.

The approach outlined here is illustrated in Figure 8. The graph is an amalgam of a CLD and a BBN, with the arcs of the BBN shown as dashed. This approach requires the modeller and the system expert(s) to think deeply about possible dependencies not just between included model parameters, but also between evocative background variables and model parameters, and adds another dimension to the elicitation process. As Eden (1992) has previously pointed out, however, 'The elicitation process is designed to be a cathartic experience which provides "added value" because it changes thinking...'

CONCLUSION

System Dynamics and Decision Analysis both have significant track records of success in providing support to high-level decision makers. As ever-more complex problem situations are exposed to analysis, ways of extending and enhancing these successful tools must be sought. In this paper, some such extensions have been suggested. Further work is required to test these ideas and explore their consequences.

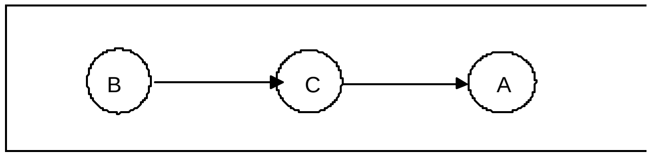

Figure 1. A and B are conditionally independent given C.

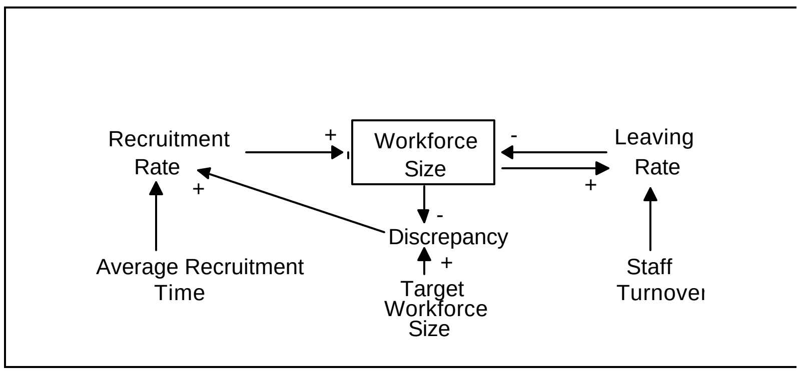

Figure 2. CLD for the simple recruitment process.

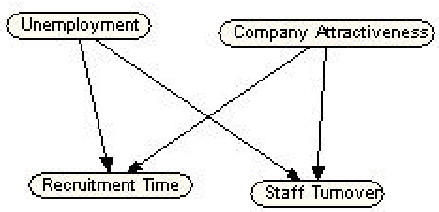

Figure 3. BBN displaying the dependencies between the background and explicit model parameters.

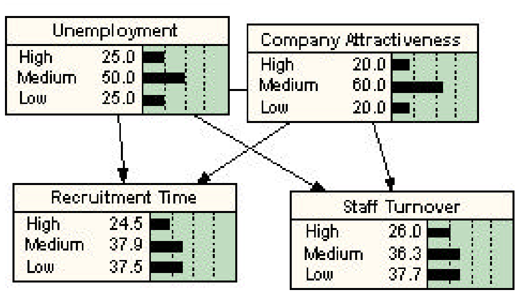

Figure 4. Initial BBN with no parameters observed.

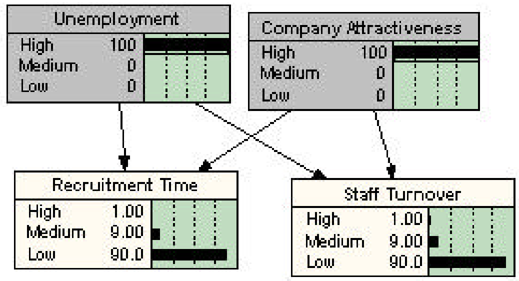

Figure 5. Resultant distribution with Unemployment and Company Attractiveness both high.

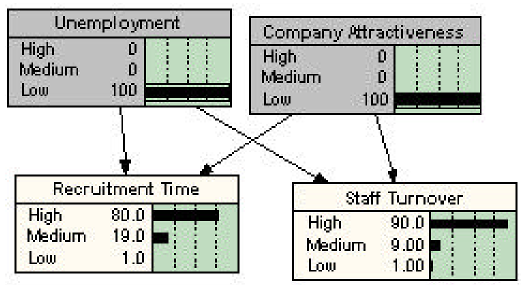

Figure 6. Resultant distribution with Unemployment and Company Attractiveness both low.

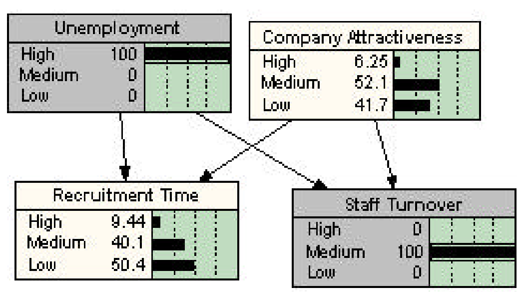

Figure 7. Resultant distribution with Unemployment high and Staff Turnover medium.

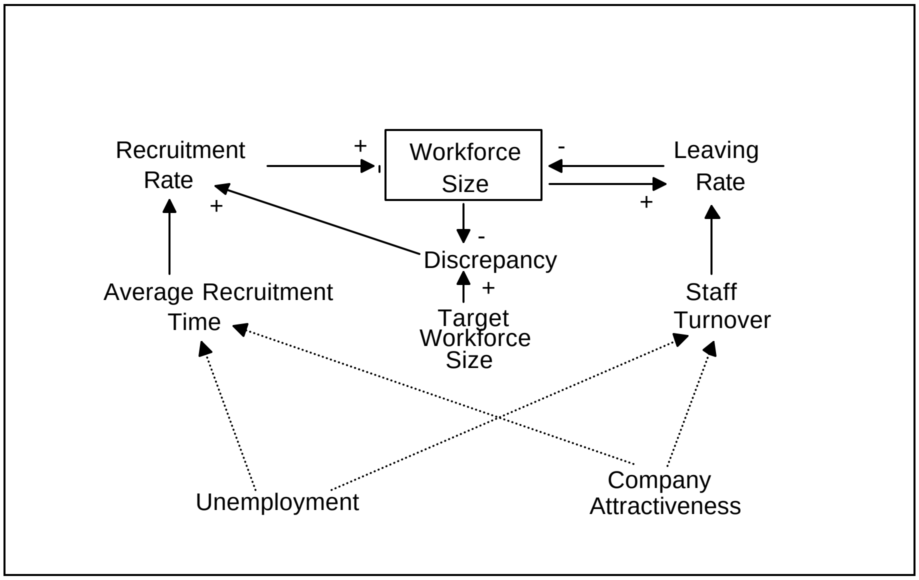

Figure 8. Combined CLD and BBN for the recruitment process.

References

Andersen, D. F. and J. Rohrbaugh. 1992. Some conceptual and technical problems in integrating models of judgment with simulation models. *IEEE Transactions on Systems, Man, and Cybernetics* 22: 21-34.

Coyle, R. G. 2002. A possible method for assessing the relative values of alternative system dynamics models. *Proceedings of the International System Dynamics Conference*, Palermo.

Eden, C. 1992. On the nature of cognitive maps. *Journal of Management Studies* 29: 261-265.

Eden, C., F. Ackermann and S. Cropper. 1992. The analysis of cause maps. *Journal of Management Studies* 29: 309-324.

Ford, D. N. and J. D. Sterman. 1998. Expert knowledge elicitation to improve formal and mental models. *System Dynamics Review* 14 (4): 309-340.

Howard, R. 1989. Knowledge maps. *Management Science* 35: 903-922.

Howard, R. and J. E. Matheson. 1984. Influence diagrams. In *Principles and Applications of Decision Analysis, Vol. 2.* Strategic Decisions Group, Menlo Park, California.

Nadkarni, S. and P. P. Shenoy. 2001. A Bayesian network approach to making inferences in causal maps. *European Journal of Operational Research* 128: 479-498.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco.

Shachter, R. D. 1986. Evaluating influence diagrams. *Operations Research* 34: 871-882.

Santos, S., V. Belton and S. Howick. 2001. Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations and Production Management* 22: 1246-1272.