What's in a mental model of a dynamic system? Conceptual structure and model comparison

Martin Schaffernicht
Facultad de Ciencias Empresariales
Universidad de Talca
Avenida Lircay s/n
3460000 Talca, CHILE
martin@utalca.cl

Stefan N. Groesser
University of St. Gallen
Institute of Management
System Dynamics Group
Dufourstrasse 40a
CH - 9000 St. Gallen, Switzerland
Tel: +41 71 224 23 82 / Fax: +41 71 224 23 55
stefan.groesser@unisg.ch

Contribution to the International System Dynamics Conference 2009
Albuquerque, NM, U.S.A.

Abstract

This paper deals with the representation of mental models of dynamic systems (MMDS). Improving 'mental models' has always been fundamental in the field of system dynamics. Even though a specific definition exists, no conceptual model of the structure of a MMDS has been offered so far. Previous research about the learning effects of system dynamics interventions have used two methods to represent and analyze mental models. To what extend is the result of these methods comparable? Can they be used to account for a MMDS which is suitable for the system dynamics methodology? Two exemplary MMDSs are compared with both methods. We have found that the procedures and results differ significantly . In addition, neither of the methods can account for the concept of feedback loops. Based on this finding, we propose a conceptual model for the structure of a MMDS, a method to compare them, and a revised definition of MMDS. The paper concludes with a call for more substantive research.

Keywords: Mental models, dynamic systems, mental model comparison, graph theory, mental model measurement

Introduction

Mental models have been a key concept in system dynamics from the beginning of the field (Forrester, 1961). One can say that the purpose of system dynamics modeling as well as

simulators based on system dynamics models is to develop or improve mental models: "improving human judgment and decision making" (Forrester, 1985:134). During many years, the term 'mental models' has been used without referring to a precise definition. About ten years ago, the concept of a 'Mental Model of Dynamic Systems' (MMDS) has been defined and discussed (Doyle and Ford, 1998, 1999; Lane, 1999). According to them, "a mental model of a dynamic system is a relatively enduring and accessible, but limited, internal *conceptual representation* of an external system (historical, existing or projected) whose structure is analogous to the perceived structure of that system" (Doyle and Ford, 1999: 141; emphasis added). In an earlier contribution, Doyle (1997) has positioned mental models as one kind of mental representation along with others like 'schemas' and 'scripts'. Interestingly, the content of MMDS is not described beyond stating that it is conceptual and a representation.

In the years after Doyle, no attempt to enhance the definition has been published. In addition, it seems noteworthy that inquiries into changes of mental models which result from system dynamics interventions refer to Doyle, but do not use it in operational terms (Capelo and Fereira, 2008). This is understandable, because the definition of MMDS does not account for a structure of mental models. Measurement and comparison of mental models require a conceptual structure. Since system dynamics claims to be able to improve mental models, it is important to go beyond belief (Maier and Grössler, 2002) and provide evidence for mental model development. One possibility is to measure MMDS before and after an intervention and compare the results.

In other fields, e.g., psychology and organizational research, several studies have been undertaken which compare changes in mental models. In this, they had to assume that the mental models have a conceptual structure which can be captured by the selected approaches. In addition, researchers seem to have pursued different ways to compare mental models. Some have applied the so-called distance ratio method (Doyle et al., 1996 and 1998), others have used the so-called closeness approach (Ritchie-Dunham, 2002; Capelo and Fereira, 2008). The fact that the literature on the 'distance ratio' does not discuss the 'closeness' approach, and vice versa, calls our attention.

This paper addresses three issues related to this state of affairs. First stands the question if the 'distance-ratio method' and the 'closeness method' result in similar representations, and second, if they are able to capture the effects of system dynamics interventions on MMDS. We use a textbook example to compare both methods and answer the question. As it turns out, both methods differ significantly in the context can capture and their results. Third, even though both approaches consider the elements 'links' and 'variables', they completely ignore the concept of feedback loops. For system dynamics, a feedback loop is a central element. A method accounting for feedback loops would yield significant insights for system dynamics modeling. We propose a preliminary idea about what a conceptual representation of a MMDS contains. Moreover, we propose a way for an adequate application.

The following section introduces the distance ratio and the closeness method. It then provides an exemplary case study and the results of the application of both methods. Thereafter, we critically discuss the methods' strengths and weaknesses. The subsequent section presents our argument in favor of a method capable of accounting for feedback loops and a tentative revision of the definition of a MMDS. Our conclusions mainly call for further contributions to a method for comparing system dynamics models.

Comparison of two Methods to Represent and Analyze Mental Models

In the following, we introduce two methods for the comparison of existing causal maps. Here, we consider causal maps as explicit and formalized representations of mental models. In the third section, we develop a conceptual structure of a mental model – a representation of the fundamental assumptions the methods of comparison. To keep the paper brief, we exclude aspects related to the elicitation of these mental models (see Ford & Sterman, 1997 for elicitation techniques).

The Distance Ratio Method (DR-Method)

This method was initially developed by Langfield-Smith et al. (1992) and improved by Markóvski and Goldberg (1995), whom we follow in the description below. It allows measuring the correspondence between different mental models, A and B in our case. An extension by Langan-Fox et al. (2000; 2001) allows to compare group mental models. The method has been applied in an early system dynamics study (Doyle et al. 1996 and 1998).

The distance ratio is a non-negative number that expresses the degree of difference of A and B. The comparison is based upon the two sets of variables and causal links of A and B, respectively. The ratio ranges from 0 (identical models) to 1 (no commonalities). In order to compare the two mental models, each of them is represented as 'association' or 'adjacency' matrix, *A* and *B* respectively, where each of the model's variables constitute a row and a column. Rows are numbered from 1 to *p* using an index *i;* columns from 1 to *p* using index *j*. Each variable is assigned a row and column with a specific number and *i=j*. If variables *x* and *y* are located at row *r* and column *c* respectively, possible links between them will appear in cells a_{rc} and b_{rc}. Links from a variable *x* to a variable *y* are denoted as "1" for positive polarity and "-1" for negative polarity; "0" indicates no causal connection. For instance, a_{rc} = 1 and b_{rc}=-1 signify that model *A* has is a positive link from *x* to y, and model *B* contains a link with a negative polarity from *x* to *y*. We use *p* to denote the total number of possible nodes; P_c is the set of common nodes in *A* and *B*, and p_c is the number of common nodes. P_{uA} is the number of nodes unique to *A* and P_{uB} the number of nodes unique in *B*. N_A and N_B are the total sets of nodes in the models.

Parameter α expresses the possibility to include 'self-loops' in the representation (0 = possible, 1 = not possible); in our case we select α = 1. Parameter β represents the highest possible link strength, which is 1 in our case. For approaches which take link strength and polarity into account, it may be justified to give different significances to polarity differences: in case a link is positive in one model and negative in the other, this difference is more relevant if the link has a higher strength. Parameter δ indicates the importance to polarity change according to the strength of links involved; in our case, δ = 0 to not increase the difference. The modeling approaches do not often take into account the possible states of polarity of causal links. Parameter ε represents the possible number of possible polarities; in our case ε = 2. Parameter γ accounts for the circumstance that links are absent from a model because (1) a causal link between two variables is believed not to exist, or (2) that one or both of the involved variables are not part of the model. If the (1) is accounted for differently ten

(2), γ represents this as in our case: $\gamma = 2$. The formula for the distance ratio is provided in Equation 1.

$$DR(A,B) = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} diff(i,j)}{(\varepsilon\beta+\delta)p_c^2 + \gamma'(2p_c(p_{uA}+p_{uB}) + p_{uA}^2 + p_{uB}^2) - \alpha((\varepsilon\beta+\delta)p_c + \gamma'(p_{uA}+p_{uB}))}$$

Equation 1: The formula for the distance ratio

In this equation, one has to sum up the possible differences between the elements of A and B using one of the following possibilities:

diff(i,j) = 0 if i = j (main diagonal) and $\alpha = 1$ (no self-loops);

$\Gamma(a_{ij}, b_{ij})$ if either i or j $\notin P_c$ and i,j $\in N_A$ or i,j $\in N_B$ (a variable neither in A nor B);

$|a_{ij} - b_{ij}| + \delta$ if $a_{ij} * b_{ij} < 0$ (if there is a difference between the polarities);

$|a_{ij} - b_{ij}|$ otherwise.

For the case that one variable is only part of one model, additional considerations are necessary:

$\Gamma(a_{ij}, b_{ij})$ = 0 if $\gamma = 0$;

0 if $\gamma = 1$ and $a_{ij} = b_{ij} = 0$;

1 otherwise.

The denominator constitutes the largest possible difference that can exist between *A* and *B*. Recall that we have set $\alpha = 1$, $\beta = 1$, $\delta = 0$, $\varepsilon = 2$, $\gamma = 2$. In addition to the previous, $\gamma' = 0$; if $\gamma = 0$; and 1 if otherwise; in our case: $\gamma' = 1$. These specifications simplify Equation 1 as following:

$$DR(A,B) = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} diff(i,j)}{2p_c^2 + (2p_c(p_{uA}+p_{uB}) + p_{uA}^2 + p_{uB}^2) - (2p_c + 2(p_{uA}+p_{uB}))}$$

Equation 2: Formulation of the distance ratio specified for the use of system dynamics

As can be seen, the number of variables that exist only in one of the compared models is important to calculate differences. The DR-method accounts for variables and causal links with positive or negative polarity. By this, a causal loop diagram which has been reduced to a causal diagram without feedback loops. In principle, a causal diagram can be used to compare mental models that are articulated in the context of system dynamics modeling.

The Closeness Method

As the distance method, the closeness method also uses elements and algorithms from graph theory to calculate the relationship between two models (Schvaneveldt, 1990; Goldsmith and Davenport, 1990). Ritchie-Dunham (2002) and Capelo and Fereira (2008) have used it to compare models created with system dynamics. The method develops a network of nodes, representing variables, and their corresponding links.

For our comparison, we take again models A and B. First, we need a reference network containing all variables and possible links against which each of the models are compared. This reference model is R = A∪B. Departing from the adjacency matrices of A and B, one can construct the extended adjacency matrix and then define each cell $r_{i,j}$ of R as the Boolean sum of $a_{i,j} + b_{i,j}$. For practical reasons, we take the maximum function $r_{i,j} = \max(a_{i,j}, b_{i,j})$. Then the degree of similarity (DS) is calculated as (Equation 3):

$$DS(A,R) = \frac{numberoflinks(A)}{numberoflinks(R)}$$

Equation 3: The closeness method computes the degree of similarity to two models

The resulting number indicated the degree of similarity of both models where "1" signifies that both models are 'identical'; "0" indicate two 'completely different' models. Again, the loops of a causal loop diagram have to be reduced; the remaining diagram can be used to apply this method.

Conceptual structure underlying the methods of comparison

Both methods to compare mental models share some commonalities. These we elaborate in this section and term it the "conceptual structure". Both methods consider mental models to consist of variables and links with their respective polarity.

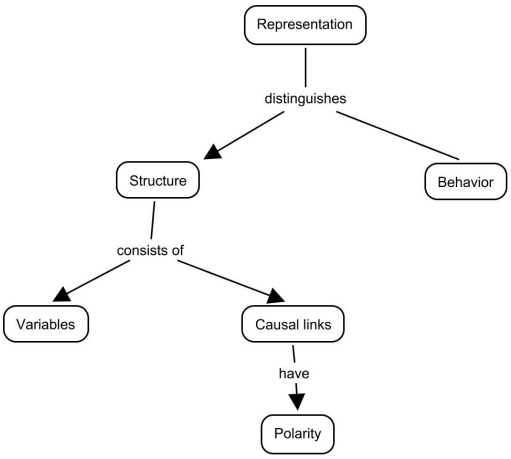

Figure 1: Conceptual structure of a mental model.

Any such structure is used to think about the behavior of the system or the problem. Both methods represent variables as nodes and causal links as connectors (lines) constituting a digraph, which allows to use graph-theoretic tools (Warfield, 1989).

Application

Preparations

Both methods have been used in system dynamics; hence, we can infer that both methods are equally useful for the representation of mental models? Do they result in the same level of similarity of two mental models? We apply both methods to two simple causal loop models (taken from Morecroft, 2007, Chapter 7). The causal loop diagrams are assumed to capture essential parts of the subjects' mental models. This is a common approach in the field (Capelo and Fereira, 2008).

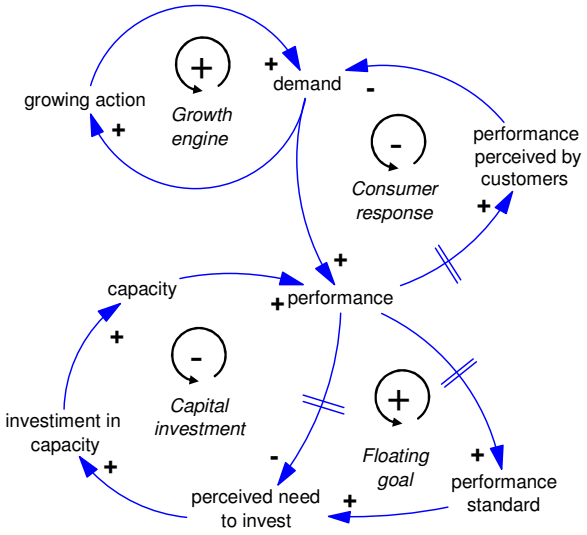

Figure 2: Model A presents the structure of the "growth and underinvestment" archetype (Morecroft, 2007: 194)

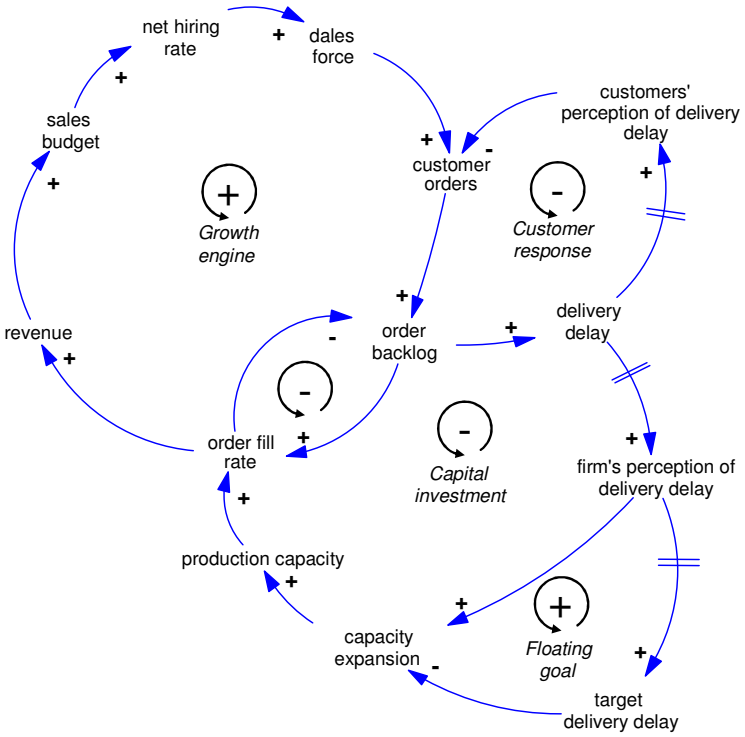

Figure 3: Model B depicts the “market growth” case (Morecroft, 2007: 198)

Model A presents the structure of the “growth and underinvestment” case; Model B depicts Forrester’s “market growth” model (Morecroft, 2007: 194). Both models consist, in principle, of same loops and variables similar in their meaning. For instance, ‘customer orders’ (model B) match ‘demand’ (model A). In the following, we choose abbreviations for the variables which ease the following comparison and analysis. By this step, we have also synthesized equivalent variables. Note that this has only been done where a direct correspondence was possible; one could argue that the variable “growing action” is sufficiently similar to the causal chain of “revenue-sales budget-net hiring rate-sales force”. However, we did not establish this kind of aggregates to not bias the results of the comparisons. Table 1 presents the abbreviations.

Model A's variables	*Model B's variables*	Abreviations
Growing Action		GA
	Revenue	R
	Sales Budget	SB
	Net Hiring Rate	NHR
	Sales Force	SF
Demand	Customer Orders	D
	Order Backlog	OB
Performance	Order Fill Rate	P
	Delivery Delay	DD
Performance Perceived by Customers	Customers' perception of delivery delay	PPC
	Firm's Perception of Delivery Delay	FPDD
Performance Standard	Target delivery delay	PS
Perceived Need to Invest		PNI
Investment in Capacity	Capacity expansion	IC
Capacity	Production capacity	C

Table 1: Variables of the two models[1]

Figures 4 and 5 show the causal diagrams in which the variables are substituted by the abbreviations.

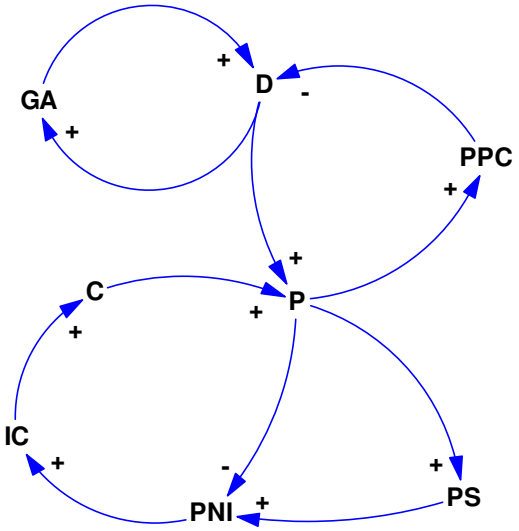

Figure 4 : Model A's causal diagram

[1] Abbreviations are built from the capitalized initial letters of the variables. Where there are two 'candidates' in a row, the chosen one appears in italics.

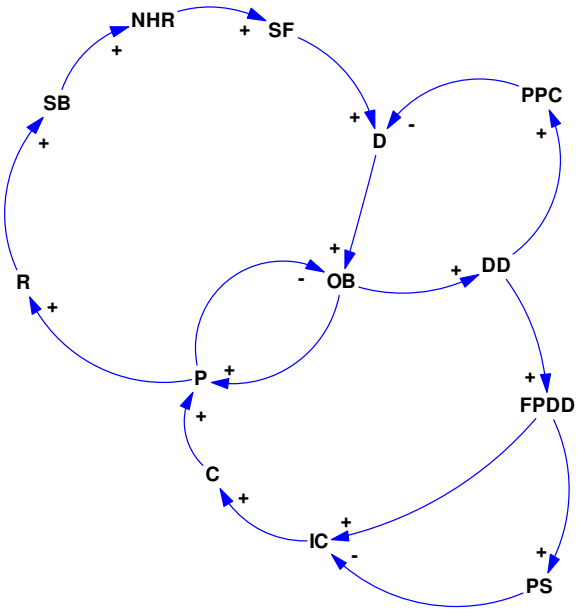

Figure 5: Model B's causal diagram

Consider, two individuals, one having read the “growth and underinvestment” case and the other the “market growth” paper. They both create a diagram about what they internalized (Figure 4 and 5, respectively). Our intention is to explicate, if the two representations of mental models are different to the “reality”; i.e., compare (1) the explicated mental model of the subjects and the original and (2) both explicated mental models with each other.

Applying the Distance Ratio Method

Both models are first converted into an adjacency matrix with the causal connections represented by the factor (+/-) 1 (Tables 2 and 3).

	GA	*D*	*P*	*PPC*	*PS*	*PNI*	*IC*	*C*
GA		1						
D	1		1					
P				1	1	-1		
PPC		-1						
PS						1		
PNI							1	
IC								1
C			1					

Table 2: Model A's adjacency matrix

	R	SB	NHR	SF	D	OB	P	DD	PPC	FPDD	PS	IC	C
R		1											
SB			1										
NHR				1									
SF					1								
D						1							
OB							1	1					
P	1					-1							
DD									1	1			
PPC					-1								
FPDD											1	1	
PS												-1	
IC													1
C							1						

Table 3: Model B's adjacency matrix

Since each of the models uses different variables, the matrices are different in some rows and columns. In order to make them comparable, both are converted into an extended adjacency matrix, which includes all variables from both models (Table 4).

A	R	SB	NHR	SF	D	OB	P	DD	PPC	FPDD	PS	IC	C	GA	PNI
R															
SB															
NHR															
SF															
D							1							1	
OB															
P									1		1				-1
DD															
PPC					-1										
FPDD															
PS															1
IC													1		
C							1								
GA					1										
PNI												1			

B	R	SB	NHR	SF	D	OB	P	DD	PPC	FPDD	PS	IC	C	GA	PNI
R		1													
SB			1												
NHR				1											
SF					1										
D						1									
OB							1	1							
P	1					-1									
DD									1	1					
PPC					-1										
FPDD											1	1			
PS												-1			
IC													1		
C							1								
GA															
PNI															

Table 4: The models' extended adjacency matrices

Model A has 8 and model B has 13 nodes. Using the definitions and specifications of Equation 2, we obtain $p = 15$, $p_c = 6$, $P_{uA} = 2$, and $P_{uB} = 7$. The denominator of Equation 2 results in 203.

In the following, we calculate the difference ratio between the extended adjacency matrices. The cells in Table 5 have the value 1, if there is a difference between the corresponding cells in matrices A and B ($A_{D,P}$ <> $B_{D,P}$ => $DIFF_{D,P} = 1$); otherwise the value is 0 ($A_{IC,C} = B_{IC,C}$ => $DIFF_{IC,I} = 0$).

DIFF	*R*	*SB*	*NHR*	*SF*	*D*	*OB*	*P*	*DD*	*PPC*	*FPDD*	*PS*	*IC*	*C*	*GA*	*PNI*	
R	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
SB	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
NHR	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
SF	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
D	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	3
OB	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
P	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	5
DD	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
PPC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FPDD	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2
PS	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	2
IC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GA	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
PNI	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
															Diff	**22**

Table 5: Difference matrix of adjacent matrixes A and B

The additional column to the right sums row elements; in total, there are 22 differences between A and B. The denominator of Equation 2 is 203. This results in the following Distance Ratio: *DR = 10.84%*. The distance ratio indicates that both models are similar to a high degree. This supports our initial intuition.

Application of the Closeness Method

The first step in the application of the closeness method is to construct the reference model or network R:

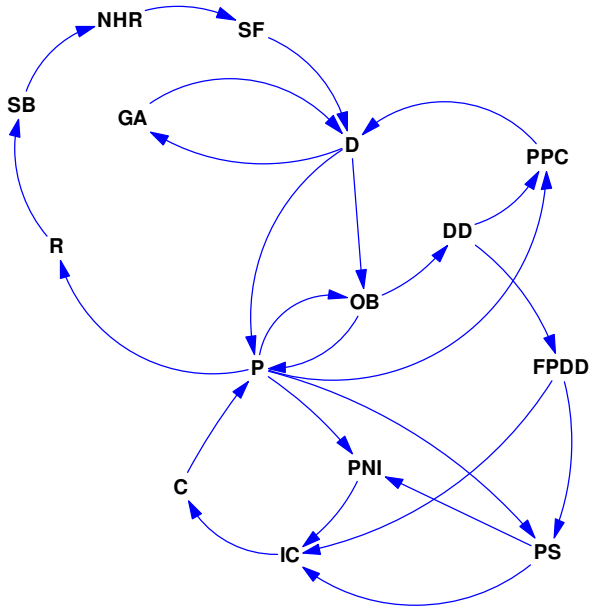

Figure 6: Reference model R synthesizes both the abstracted model A and B

The reference model synthesizes both the abstracted versions of models A and B. It contains the variables and links of both models. The polarity of the causal links is not accounted for by the closeness method. Table 6 shows the matrix for the reference model.

NET	R	SB	NHR	SF	D	OB	P	DD	PPC	FPDD	PS	IC	C	GA	PNI
R		1													
SB			1												
NHR				1											
SF					1										
D						1	1							1	
OB							1	1							
P	1					1					1				1
DD									1	1					
PPC					1										
FPDD											1	1			
PS												1			1
IC													1		
C							1								
GA					1										
PNI												1			

Table 6: Matrix form of the reference model R

The total number of links is 24. We can then use the extended adjacency matrices (Table 3 and 4) to determine the closeness measure. With 11 links for A and 17 for B, the closeness ratios are: CR(A,R) = 45.83%; CR(B,R) = 70.83%. The method indicates that model B has a high level of similarity; model A seems to be rather dissimilar.

One may feel uncomfortable with the fact that this method constructs a reference model R, where the previous method directly compared between A and B. An imaginable alternative would be to compare A to B and B to A, each time using the other model as reference model (and its links as 'all possible links'). For this, we have to determine the number of common links between A and B:

A and B	R	SB	NHR	SF	D	OB	P	DD	PPC	FPDD	PS	IC	C	GA	PNI
R															
SB															
NHR															
SF															
D															
OB															
P															
DD															
PPC					1										
FPDD															
PS															
IC													1		
C							1								
GA															
PNI															

Table 7: Common links between A and B

Each of these three common links has to be compared to the maximum number of links of the respective reference model. It follows that: DS(A,B) = 3/11 = 27.27% and DS(B,A) = 3/17 = 17.65%. Thus the closeness method suggests that models A and B are not similar to one another.

Critique of the Compared Methods

The distance ratio between models A and B is 10%. The degree of closeness of models A and B with respect to the reference model R is 46% and 70% respectively (or 27% and 17% respectively). These are surprisingly different results. Are models A and B close to each other, as suggested by the distance ratio, or distant, as indicated by the closeness measure?

The methodological assumptions of the distance ratio and the closeness method account for the different results. The distance ratio method uses the variables, the links, and the links' polarity, whereas the closeness method considers only links, not the link polarity and also not the variables. As can be seen above, models A and B have 8 and 13 variables respectively; summing up to 15 in total. Six of these are common to both models. We can compare each to the reference set (shown in the extended adjacency matrix, Table 6) resulting in the fraction of shared variables:

A/R = 6/15 = 35.33%

B/R = 13/15 = 86.67%

Alternatively, we can determine the share of variables that A has common with B (and vice versa):

Common variables of A in B = 6/13 = 46.15%

Common variables of B in A = 6/8 = 75.0%

Models A and B are more similar to each other when variables are taken into account, and not only links. This clearly indicates that commonality or difference at the level of variables is relevant for model comparison. If the set of variables is predetermined by the shape of the inquiry process – which seems to be usually the case in the literature concerning the similarity method – then it may seem that there cannot be differences at this level. Still, if for a given subject, some variables of the reference model are not given weight in any link, then this indicates that the variable is not important in the respondent's mental model. Not taking this into account seems debatable from the perspective of system dynamics.

In this sense, it is not too surprising that the distance ratio suggests more similarity than the closeness measure. Besides this difference, not all links have the same polarity. By not prompting subjects to articulate if their causal belief says 'inverse' or 'proportional' – corresponding to a naive notion of polarity as discussed in Schaffernicht (2007) – the closeness method ignores an important concept in system dynamics.

Considering these aspects, the closeness method may have limited usefulness for the analysis of system dynamics models, as compared to the distance ratio method: when differences between variables and the polarity are relevant, the distance method is more appropriate.

However, there is something more to say. What does a distance of 10% or a closeness of 45% or 70% between models A and B indicate for a discipline that uses feedback loops as fundamental components of social systems (Forrester, 1968)? When systems thinking capabilities are assessed (Booth-Sweeny and Sterman, 2007), much attention is paid to feedback loops – however, the concept does not in any of the two methods. This has undesirable consequences.

Both methods assume that all links and nodes are equally important. This is debatable as can be seen from considering the feedback loops of models A and B:

Loop	**Models**	
	A	**B**
Growth engine	D->GA	D->OB->P->R->SB->NHR->SF
Consumer response	D->P->PPC	D->OB->DD->PPC
Capital investment	P->PNI->IC->C	P->OB->DD->FPDD->IC->C
Floating goal	P->PS->PNI->IC->C	P->OB->DD->FPDD->PS->IC->C

Table 8: Comparing feedback loops between models A and B.

Both models are made up by the same feedback loops (the loop between P and OB in model B is a consequence of the 'physics' – P being the outflow of OB; so it is not of major interest and does not even receive a name), and despite the differences at the level of variables, the loops have the same polarities in both cases.

As argued above, model B is more disaggregated than model A, and the distance seems to stem from this fact. Still, at the descriptive level of feedback loops, one would rather say that both models are the same: as suggested by Table 8, despite the different amount of detail each of the four loops can keep its name, since they signify the same.

Then again, if one of the loops disappeared as consequence of a missing link, the distance ratio would react insignificantly . But when looking at the loop level, we would conclude that both models are significantly different in this case. Does this not indicate that links and variables that belong to one or several feedback loops are more important than those that do not? How can a method for comparing models assess what is important at the level of feedback loops if it does not take them into account?

Eliminado: ¶

Towards the representation and comparison of MMDS

Defining the structure of "conceptual representation"

The points outlined above suggest that system dynamics would benefit from a clear definition of what is meant by "conceptual representation" in the definition of a MMDS (Doyle and Ford 1998). We elaborate a tentative definition and propose a way of comparing such mental models.

When someone studies a situation applying the system dynamics methodology, we must assume a situation in which the approach can be appropriately applied. Then we should expect the articulated maps – external representations of cognition - to be developed according to the language's vocabulary and grammar. The following figure decomposes "conceptual representation" into elements used in system dynamics:

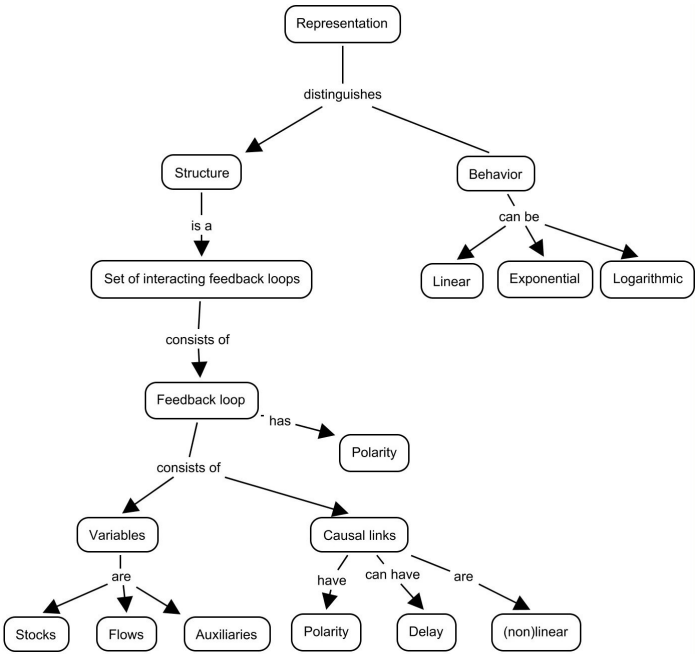

Figure 7: The conceptual representation of system dynamics. It shows how dynamic systems are conceptually represented.

The first thing to note is that structure is separated from behavior. Structure refers a hierarchy of elements, where three different levels of description are distinguished. In this point, both methods coincide. Next, the system is thought to be a set of interacting feedback loops (Forrester, 1968). Then appears the feedback loop, which has a polarity and consists of various types of variables bound together by causal links.

In this sense, we propose to modify the definition of MMDS to "a mental model of a dynamic system is a relatively enduring and accessible, but limited, internal *conceptual representation* of an external system (historical, existing, or projected) *in terms of reinforcing and balancing feedback loops emerging from state and flow variables that interact in non-linear and delayed ways*, whose structure is analogous to the perceived structure of that system".

Outline for comparing MMDS

While the distance ratio and the closeness methods are well suited for modeling paradigms with one level of representation, they are incomplete from the point of view of system dynamics. The methods are not capable to account for the emergent level of individual feedback loops and the level of the set of interacting feedback loops. We might overcome this limitation and develop an extension capable of taking into account feedback loops (both identity and polarity) and their interaction over a subset of variables, based upon preliminary work by Schaffernicht (2006). We have also argued that the current methods are not able to explain all relevant differences in understanding because it treats all variables equally; therefore, we complement it with qualitative analysis of model differences based upon the method proposed therein.

The starting point is to consider the 'model' to be a sequence of 'versions' that rise over time. In between these versions, differences can arise in terms of the boundary (variables included), the time horizon and the respective sets of model components: variables (V), causal links (CL), and loops (L). If we wish to analyze the difference between the models A and B, the following intermediate indicators are calculated (adapted from Schaffernicht, 2006):

Number of	*in A*	*in B*	*In A and not in B*	*In B and not in A*	*In B and in A*	*In A and in B but modified*
Variables	$NumV_A$	$NumV_B$	$OnlyV_A$	$OnlyV_B$	$bothV_{A\_B}$	$modV_{A\_B}$
Causal links	$NumCL_A$	$NumCL_B$	$OnlyCL_A$	$OnlyCL_B$	$bothCL_{A\_B}$	$modCL_{A\_B}$
Loops	$NumL_A$	$NumL_B$	$OnlyL_A$	$OnlyL_B$	$bothL_{A\_B}$	$modL_{A\_B}$

Table 9: Indicators for model comparison

Interestingly, if we replace the 'number of' elements' by 'identify', then the different cells of the previous table contain the identifiers of the respective components, which gives a more detailed view about the different sets. The following table shows the results of comparing A and B in this way:

Number of	in A	in B	In A and not in B	In B and not in A	In B and in A	In A and in B but modified
Variables	GA, D, P, PPC, PS, PNI, IC, C	R, SB, NHR, SF, D, OB, P, DD, PPC, FPDD, PS, IC, C	GA, PNI	R, SB, NHR, SF, OB, DD, FPDD	D, P, PPC, PS, IC, C	n/a
Causal links	GA->D D->GA D->P P->PPC PPC->D P->PS P->PNI PS->PNI PNI->IC IC->C C->P	P->R R->SB SB->NHR NHR->SF SF->D D->OB OB->P P->OB OB->DD DD->PPC PPC->D DD->FPDD FPDD->PS FPDD->IC PS->IC IC->C C->P	GA->D D->GA D->P P->PPC P->PS P->PNI PS->PNI PNI->IC	P->R R->SB SB->NHR NHR->SF SF->D D->OB OB->P P->OB OB->DD DD->PPC DD->FPDD FPDD->PS FPDD->IC PS->IC	PPC->D IC->C C->P	-
Loops	Growth engine; Customer response; Capital investment; Floating goal.	Growth engine; Customer response; Capital investment; Floating goal.	-	-	-	Growth engine; Customer response; Capital investment; Floating goal.

Table 10: Qualitative comparison

The rows concerning variables and causal links contain elements we have already seen above. However, the fact that both models have the same loops but all of them have differences as for some of their variables now becomes vivid.

Based on this information, the next Table 11 displays the corresponding quantities:

Number of	in A	in B	In A and not in B	In B and not in A	In B and in A	In A and in B but modified
Variables	$NumV_A$=8	$NumV_B$=13	$OnlyV_A$=2	$OnlyV_B$=7	$bothV_{A\_B}$=6	n/a
Causal links	$NumCL_A$=11	$NumCL_B$=17	$OnlyCL_A$=8	$OnlyCL_B$=14	$bothCL_{A\_B}$=3	$modCL_{A\_B}$=0
Loops	$NumL_A$=4	$NumL_B$=4	$OnlyL_A$=0	$OnlyL_B$=0	$bothL_{A\_B}$=0	$modL_{A\_B}$=4

Table 11: Quantitative comparison

Now, we can compute insightful ratios. For each of the concepts, we can determine how similar and how different each model is with respect to the other. The common elements of a model, divided by the number of elements in the other model indicate the degree of similarity. The number of unique elements of a model divided by the number of elements in this model shows how the level of difference.

Aspect	Model A (with respect to B)	Model B (with respect to A)
Variable Similarity	$bothV_{A\_B}$ / $NumV_B$ = 6/13 = 0.46	$bothV_{A\_B}$ / $NumV_A$ = 6/8 = 0.75
Variable Difference	$OnlyV_A$ / $NumV_A$ = 2/8 = 0.25	$OnlyV_B$ / $NumV_B$ = 7/13 = 0.54
Links Similarity	$bothCL_{A\_B}$ / $NumCL_B$ = 3/17 = 0.18	$bothCL_{A\_B}$ / $NumCL_A$ = 3/11 = 0.27
Links Difference	$OnlyCL_A$ / $NumCL_A$ = 8/11 = 0.72	$OnlyCL_B$ / $NumCL_B$ = 14/17 = 0.83
Loops Similarity	$modL_{A\_B}$ / $NumL_A$ = 4/4 = 1	$modL_{A\_B}$ / $NumL_B$ = 4/4 = 1
Loops Difference	$OnlyL_A$ / $modL_{A\_B}$ = 0/4 = 0	$OnlyL_B$ / $modL_{A\_B}$ = 0/4 = 0

Table 12: results from quantitative comparison

The *loop* comparison poses a challenge: if loops are 'the same' only when they have exactly the same variables, then the four loops in models A and B are different. However - as we have seen before - they refer to the same meaning, have the same polarity and for the most part, the more detailed model B can be interpreted as a disaggregated version of model A. From this viewpoint, the loops appear to be identical, though modified. For this reason, the column "In A and in B but modified" has been used in this case. Clearly, more discussion is needed concerning loops comparison.

The *variable similarity* rejoins the results of the *closeness* method, which had found

DS(A,R) = 45.83%

DS(B,R) = 70.83%

The *links difference* produces approximately the same results as the alternative way in which the closeness method was used:

DS(A,B) = 3/11 = 27.27%

DS(B,A) = 3/17 = 17.65%

These indicators suggest that models A and B are rather similar in their structure regarding variables, but rather different in their causal link structure.

The *loop* indicators clearly point to that at this level, both models are "almost the same"; the "almost" stems from the fact that the loops contain different variables.

Even though originally proposed to monitor the traces of learning along longer sequences of versions, this method clearly shows structural aspects of the differences between two models that the DR and the DS method cannot show. In combination with the "identify" version of the previous table, we produce a precise image of the changes that exist between both models: one can go beyond the indicator and consider what the differences and similarities are.

Conclusions

The currently accepted definition of a "mental model of a dynamic system" (MMDS) does not specify the structural content of "internal conceptual representation"; consequently researchers who desire to represent and analyze MMDS receive no support for how to represent MMDS. We use two methods – the distance ratio and the closeness ratio – which have already been applied in previous system dynamics studies, to represent and analyze. This paper inquires if (1) the methods are comparable in their results and (2) if their underlying structure of the "conceptual representation" is satisfying for system dynamics purposes.

The first aim was then to inquire if these methods – the distance ratio and the closeness ratio – lead to similar results when applied to a system dynamics model. It was found that the approaches differ widely in what they take into account and what they produce. The distance method processes information about the variables and links (with polarity), while the closeness method only uses the information about links (without polarity). Accordingly, the results of both methods when comparing two similar, but different models do not indicate the same distance or closeness. This is not necessarily a negative assessment concerning any one of the methods, but it alerts us to choose the method to represent and analyze with care.

Then it was argued that system dynamics tries to help improving mental models using a specific language with specific symbols and meanings that go beyond the expressive power of the usual conceptual structure assumed to represent mental models. If such learning effects exist, the representation and analysis of mental models should be able to detect it. Therefore, the conceptual structure of MMDS should contain the elements to represent feedback loops . We have developed a preliminary method which can account for the characteristics of system dynamics. The consecutive application to exemplary cases has shown that the relationship between two models can be assessed more accurately when similarities, differences, and feedback loops are considered. We have found, in addition, that the distance and closeness ratios are highly condensed indicators which need a qualitative component. With our method, we try to provide this improvement.

References

Booth-Sweeny, L. and Sterman, J. 2007. Thinking about systems: student and teacher conceptions of natural and social systems, *System Dynamics Review* **23**(2/3):285-312

Capelo, C. and Ferreira, J. 2008. A system dynamics-based simulation experiment for testing mental model and performance effects of using the balanced scorecard, *System Dynamics Review* **25**(1): 1-34

Doyle, J. K. 1997. The cognitive psychology of systems thinking. *Systems Dynamics Review* **13**(3), 253-265.

Doyle, J. K. & Ford, D. N. 1998. Mental Models Concepts for system dynamics Research. *System Dynamics Review*, **14**(1): 3-29.

Doyle, J. K. & Ford, D. N. 1999. Mental Models Conepts Revisited : Some Clarifications and a Reply to Lane. *System Dynamics Review*, **15**(4): 411-415.

Doyle, J. K., Radzicki, M. J., & Trees, W. S. 1996. Measuring the Effect of System Thinking Interventions on Mental Models. Paper presented at the 1996 International System Dynamics Conference, Cambridge, Massachusetts.

Doyle, J. K., M. J. Radzicki and W. S. Trees 1998. *Measuring change in mental models of dynamic systems: An exploratory study*. Unpublished manuscript, Department of Social Science and Policy Studies, Worcester Polytechnic Institute, Worcester, MA.

Ford, D. N. and J. D. Sterman (1997). "Expert Knowledge Elicitation to Improve Formal and Mental Models." *System Dynamics Review* 14(3): 309-340.

Forrester, J. W. 1961. *Industrial Dynamics*. Cambridge MA: Productivity Press.

Forrester, J. W. 1968. *Principles of Systems*. Cambridge: The MIT Press.

Forrester, J., 1985. The “model versus a modeling process”, System Dynamics Review 1: 133-134.

Forrester, J. 2007. System dynamics – the next 50 years. System Dynamics Review 23(2/3): 359–370

Lane, D. C. 1999. Friendly Amendment : A Commentary on Doyle and Ford's Proposed Re-Definition of 'Mental Model'. *System Dynamics Review*, **15**(2): 195.

Langan-Fox, J., Code, Sh. and Langfield-Smith, K. 2000. Team mental models: techniques, methods and analytic approaches, *Human Factors* **42**(2): 242-271

Langan-Fox, J.,Wirth, A., Code, Sh., Langfield-Smith, K. and Wirth, An. 2001. Analyzing shared and team mental models, *International Journal of Industrial Ergonomics* **28**: 99-112

Langfield-Smith, K. and Wirth, A. 1992. Measuring differences between cognitive maps, *Journal of Operational Research* **43**(12): 1135-1150

Maier, F and Grössler, A. 2002. What are we talking about?—A taxonomy of computer simulations to support learning, System Dynamics Review 16(2): 135-148

Markóvski, L and Goldberg, J. 1995. A method for eliciting and comparing causal maps, *Journal of management* **21**(2), p. 305-333

Morecroft, J. 2007. *Strategic modelling and business dynamics*: A feedback approach. John Wiley.

Ritchie-Dunham J. 2002. *Balanced scorecards, mental models, and organizational performance: a simulation experiment*. PhD dissertation. University of Texas at Austin, Austin, TX.

Schaffernicht, M. 2006. Detecting and Monitoring Change in Models. *System Dynamics Review*, 22(1): 73-88.

Schvaneveldt, 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*, edited by R. Schvaneveldt, Norwood, NJ:Ablex

Warfield JN. 1989. Societal Systems: Planning, Policy and Complexity. Intersystems Publications: Salinas, CA.

Wolstenholme, E. 1990. *Systems enquiry*, John Wiley