Causality and diagrams for system dynamics

Martin Schaffernicht
Facultad de Ciencias Empresariales
Universidad de Talca
Talca – Chile
martin@utalca.cl

Abstract

Polarity and causality are important concepts but have not received much attention in the system dynamics literature. The great effort it takes students to properly understand them has motivated this inquiry. In the framework of a conceptual model of interacting with complex systems, several cognitive tasks are proposed. This paper concentrates on one of them that deals with causal links' polarity. An examination of other approaches that deal with causality and use more or less similar diagram languages shows that usually causality is only very broadly defined, and where it is operationally defined, this is done with respect to events rather than behavior. In contrast to these approaches, system dynamics is about behavior rather than events. We then revisit the traditional criticism of causal loop diagrams and show a way out, but add two new criticisms related to the inability of causal loop diagrams to address behavior: in fact it seems that they are closer to the event-related definition of causality. Also, the impossibility to execute them in simulations means that executable concept-models are to be preferred: they express important information a causal loop diagram cannot represent and on top of it they render the behavioral consequences visible (as opposed to the events). In conclusion, causal loop diagrams should only be used by experienced modelers, and be banned from educational use.

Keywords: causal link, polarity, dynamic complexity

1. Introduction: polarity and causality [1]

For the last two years, I've had the opportunity to teach system dynamics as an elective for undergraduate business students at my university. This course spends a lot of time dealing with the very basic aspects like polarity and stock-and-flow thinking. For my students, it is very challenging to understand and get used to the "correct" definition of polarity; so I use series of examples where the behavior-over-time graph of two variables is used to decide which type of polarity is involved.

[1] I'm grateful to Erling Moxnes and the students of the Bergen University for their critical questions and comments

When confronted with a task like the one shown in the following figure, most of my students intuitively believe this is a case of negative polarity:

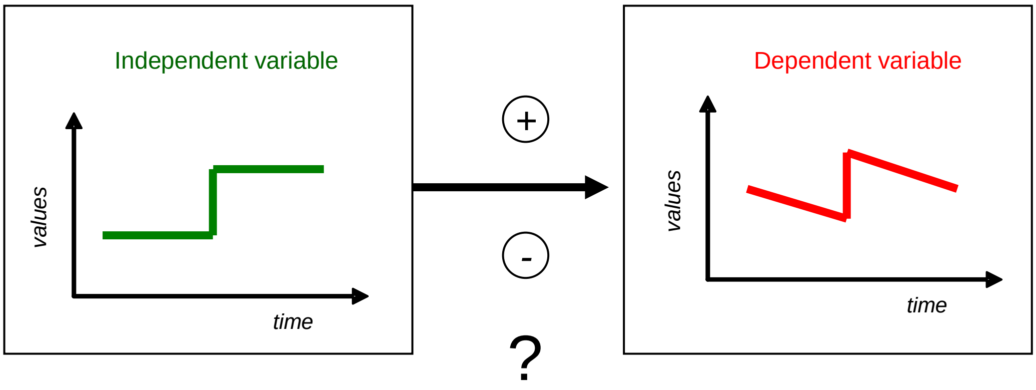

Figure 1: example of an "impossible" case of polarity

However, if one applies the "complete" definition (see Sterman, 2000), the dependent variable takes on values higher than what would have been the case; since the independent variable experienced a rise, this is a case of positive polarity!

I was troubled by this difficulty and decided to inquire into how many different configurations of causal influence I could produce to confront my students with such deceptive tasks. If we limit ourselves to "step" changes in the variables, and admit that the dependent variable may have a base behavior – a slope that is positive, null or negative, then the usual 4 combinations for two polarities become 12. If we admit "step" and "ramp" changes in the dependent variable, there are already 24 combinations. It became evident that there is a "mystery" about causal loop diagrams and polarity. How could it be that a tool meant to help you is so tricky to use?

The subsequent inquiry into the relationship between causal loop diagrams, polarity and behavior made it necessary to reflect upon the notion of causality for system dynamics. As described by (Pedercini, 2006), leading and publishing dynamicists assume the world to be such that one can specify stable causal relationships between variables in order to explain phenomena or design decision policies; causality is understood as the polarity of each link and there is widespread use of "causal loop diagrams".

The notion of *causality* and *causal diagrams* are also used by researchers in other disciplines interested in mental models and/or causality – for different purposes ranging from studying to influencing causal reasoning (Eden, 1990; Halper and Pearl, 2005a and 2005b; Johnson-Laird, 1999). However, beyond the similarities, there are differences. In usual causal diagrams, feedback loops may be identified, but they are not separately conceptualized and signaled. Also, system dynamics puts emphasis on the polarity of causal relationships (Richardson, 1991), which is one necessary condition for converting knowledge about structure into knowledge about behavior
[2]. The nodes do not always refer to variables, but also to conceptual constructs, actions and events. How do these acceptions of causality and the different types of causal diagrams relate to each other? Ando what can this mean for system dynamics? It is the purpose of this paper to contribute some elements to the answer of these questions. I believe this is worthwhile for the following reasons.

System dynamics has a well defined *normative* apparatus with rules that tell us how to decide which factors shall be part of a model, how to define the type of variable and how to quantify and validate. In a way, system dynamics is a method to enhance causal thinking. On the other side, there has been growing concern about how people fail to perceive causal relationships (the

[2] "knowledge" is used here in the sense of "best available belief".

so-called misperception of feedback; see Sterman, 1989; Moxnes, 2000; 2004) and fail to think adequately about them (stock-and-flow thinking; see Booth-Sweeny and Sterman, 2000 [3]).

The system dynamics literature has been the stage for a brief dispute concerning causal loop diagrams, in which the simple (and most popular) definition of polarity was shown to be flawed and only one of the commonly used notations for "positive" and "negative" did not fail the test (Richardson, 1997). However, the dispute seems not to have been settled, since there are still articles using the popular definition (Warren, 2004). There must be some reason for this popularity.

Also, the mental models thread seems not to have aroused investigations into the way how we think with causal relationships. System dynamics has its own definition of mental models (Doyle and Ford, 1998; 1999:114):

> "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."

This definition does not mention "causality" nor "polarity"; neither did their paper deal with ways to represent mental models. However, mental models are used to study causal reasoning and frequently use "causal maps" (Johnson-Laird, 1999). So may it be that causality is a concept that system dynamics just takes as granted, like Pedercini (2006) suggests? May it be that dynamicists simply take it for granted that "causal loop diagrams" represent articulated mental models? In the face of the reported failures to perceive and correctly think with feedback loops, there may be good reasons to study how we actually do perceive causal relationships and how we fail to, and how we *actually* do think with causal (mental) models. And in a context where the debate over the use and usefulness of "causal loop diagrams" still goes on (Homer and Oliva, 2001; Richardson, 1997; Warren, 2004), it may be worthwhile to ask what this type of diagram expresses and should or should not be used for.

This paper inquires into the meaning of causality for system dynamics by relating it to what it means for other fields and leads to a renewed critique of causal loop diagrams. The second section introduces a conceptual model of a perceiving, thinking and acting person interacting with a complex dynamic system. It introduces a sequence of cognitive operations that must be accomplished in order to appreciate the probable behavior of a multi-loop model.

The following section reviews the concepts of causality and the tools used to represent it for those who use concept mapping, cognitive mapping, causal mapping and causal diagrams. We find that they are concerned with events rather that behavior and that most do not search quantification/simulation.

Then the fourth section treats the case of system dynamics. We find that its interest for behavior distinguishes it from the other approaches. We revisit the previous criticisms of causal loop diagrams and extend the list of shortcomings: they have no notational means to distinguish between behavior (first derivative) and event (second derivative), which is not good for an approach where this difference is important. Also, they cannot represent the fact that the effects of a cause show up in a smoothed way. Even though causal loop diagrams have the comparative advantage of explicitly representing feedback loops, they allow too many misunderstandings.

So the conclusion of this inquiry is that system dynamics has a very detailed and rich notion of causality, since it goes beyond events. However, causal loop diagrams are a poor tool for modeling with this notion, and its use should be reserved to experienced dynamicists.

[3] the "Bathtub" line of work has been readily uptaken and there have been many presentations about the subject in the international system dynamics conference since the original paper.

2. Interacting with complex dynamics systems

Each person can be thought to be a constantly interacting with external systems. System dynamics conceives these systems to consist of structures that generate behaviors.

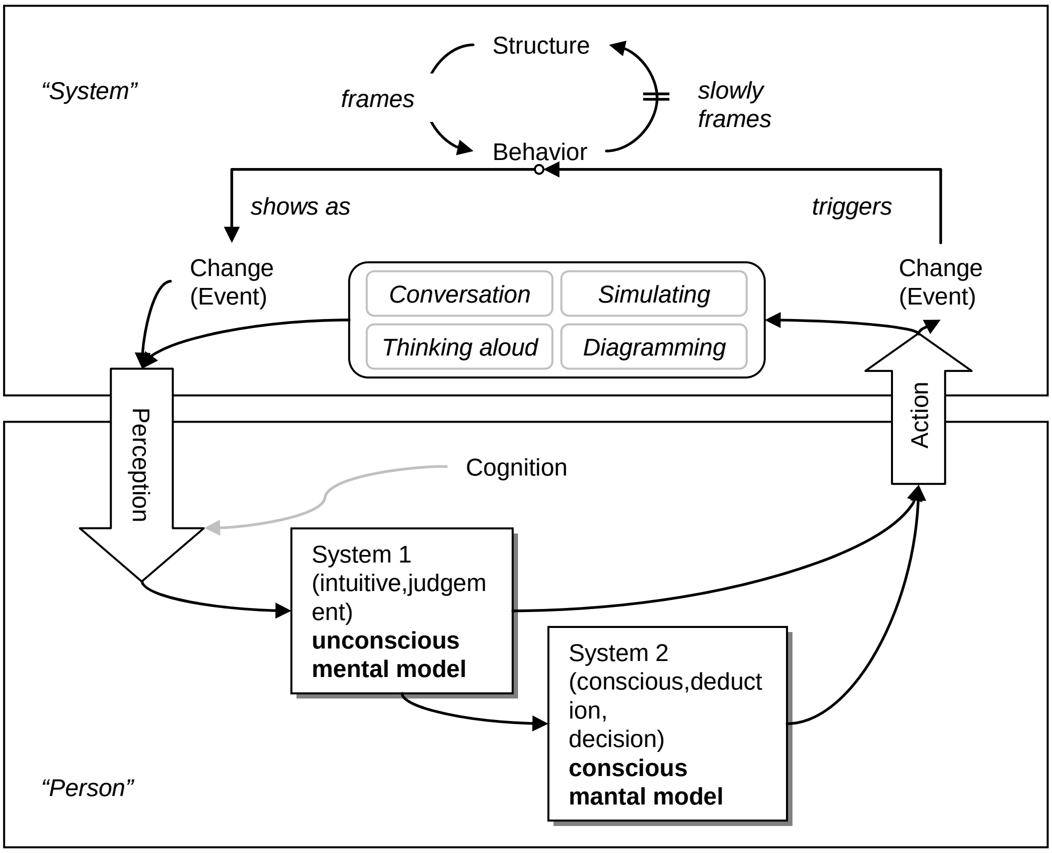

Figure 2: interaction with complex dynamic systems

As has been indicated (Reichel, 2004), other schools of thinking also recognize that behavior slowly transforms structure. The person appears as a system consisting of perception, cognition and action. According to (Kahnemann, 2002), perception is more sensitive to change (events) than to states: "a difference that makes a difference". It is thought that cognition can be decomposed into two different systems: system 1 represents unconscious, intuitive thinking, which brings into play unconscious mental models. System 2 realizes the conscious thinking, part of which can be expressed as articulated mental models. Intuitive action (and the judgment it reveals) is based on system1, while deliberate decisions are produced by system 2. The internal activity of thinking also has an effect back on perceiving; this is something we all know as "you see what you know", but it has also been experimentally found (for example, Payne and Baguley, 2006).

This perception triggers cognitive processes that many times remain in system 1 and finish with some action. At other times, system 2 may become activated and conscious thinking intervenes.

Let us assume that we find ourselves in a classroom situation where we expose students to a causal loop diagram and want to see to which degree they will be able to make an adequate statement about the probable behavior of a feedback loop's behavior.

Consider what has to happen by chaining from the final stage of thinking backward: in order to think well the feedback loop, one has to detect its polarity, which cannot be done without detecting previously the loop and each individual causal link's polarity; this in turn has to be preceded by detecting the causal link between the two variables and the variables themselves in the first place. Figure 3 illustrates this process:

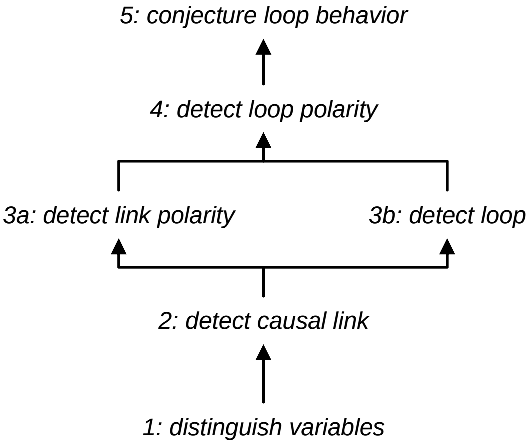

Figure 3: steps towards conjecturing loop behavior. Arrows represent precedence, not causality.

The phases drawn in higher places (with higher numbers) are closer to usual system dynamics questions, and this is where we have rules and tools. However, the lower numbered phases arouse relevant questions:

1) how do we come to consider something as a relevant variable?

2) how do we come to believe in the existence of a causal link between two variables?

3a) how do we come to see the polarity of a causal link?

3b) how do we come to perceive a feedback loop?

Even though to the trained system dynamicist this may seem ridiculously obvious, it is usually not so for other people. We can wonder if the failure comes more from perception or from cognition (and which of the two possible systems), and if the subjects fail to perceive a causal link between two variables they do perceive or if they fail to realize that there is a feedback loop (even though all participating variables and causal links have been identified).

However, for the remainder of this paper, we assume that step 1 "distinguish variables" and step 2 "detect causal link" have already been absolved, because we wish to focus on the thinking about and with the causal relationships between the variables [4]. Let us define "variable" as "an attribute of some entity that is stable in time but takes on different values over time". We will now examine the ways causality can be understood and the different diagram languages we can use to reflect upon it.

[4] It should be recognized that distinguishing variables is not a trivial activity; it is out of the scope of this paper, but the interested reader may start consulting Argyris (1993) – in an appendix concerning design causality, writes how the fact to consider one or another flow of events as one variable affects the following steps of perceiving and thinking. As for recognizing variables and links between them, there are different competing explanations for how we do it; see Gopnik et al., 2004, Allan and Tangen, 2005 and López et al., 2005. The most prominent one is associative learning, where the occurrences of events or "cues" is used to form causal attributions.

3. Causality and polarity

Some historical facets of the concept "causality"

Causality has been a subject for philosophers and scientists for a very long time (in human standards); Aristotle elaborated ideas concerning four kinds of causes:

- material: what A is [made] out of
- formal: what it is to be A
- efficient: what produces A
- final: what A is for

When we talk about causal links between variables, we are mainly interested in the efficient cause:

- what is it that caused a variable to perform a specific behavior?
- what is it that produces a certain effect in a variable we want to govern?

David Hume (1984) got to the conviction that – be there causality in the world or not – the human individual can only develop thoughts about his impressions and experiences and by consequence, humans can only attribute causes. We perceive one object A doing something and then some other object B near to A does another thing. If this repeats a sufficient number of times, we'll believe that A (somehow) causes B to do something.

Since the middle of the 20th century, psychologists have developed attribution theory, which deals with how people come to their causal beliefs in everyday situations (Heider, 1958; Kelley, 1973). System dynamics is pragmatic from its outset and has always been interested in causal beliefs that people articulate from their mental database (as observed by Pedercini, 2006).

4. Different meanings and diagram languages outside the system dynamics realm

In several fields, people explicitly use a concept of causality and rely on some form of causal diagram to reflect upon it. Since there are differences between them, it seems important to present each of them in turn and then focus on the one proper to system dynamics.

Causal links are currently used in (at least) four different ways: concept mapping, cognitive mapping, causal maps causal diagrams as used in psychology and A.I.

Concept mapping.

In concept mapping, concepts are "things usually referred to by nouns or noun phrases" that can be "linked to form propositions" (Rebich and Gautier, 2005, p. 358). In this case, the nodes in the diagram are concept phrases, for example "aerosol emissions [are generated by] industrial activities" or "longwave radiation trapping [is associated with] greenhouse effect".

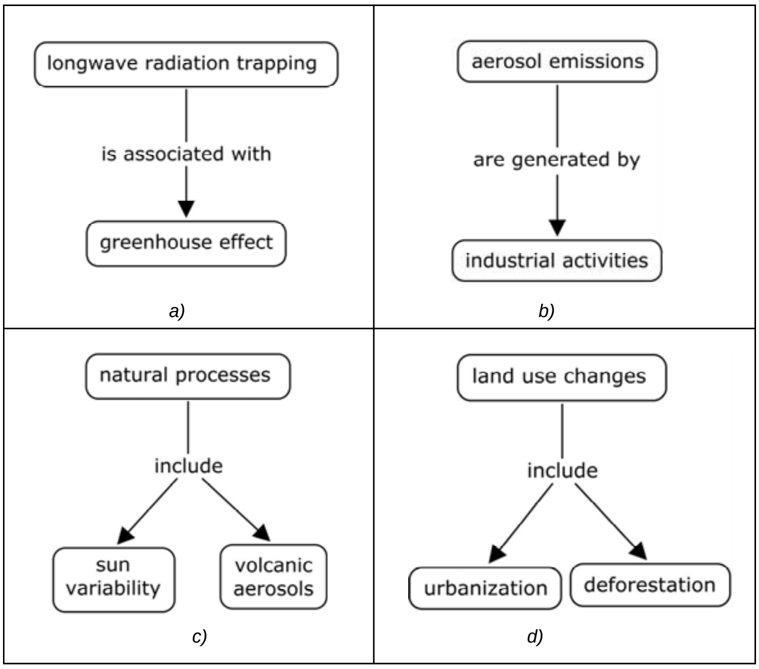

Figure 4: two examples of concept maps. Extracted from a cognitive map that is referred to in Rebich and Gautier, 2005 [5].

Note that the links have name labels that are usually verbs. In the two examples shown in Figure 4, we can see several aspects relevant from a SD point of view. In 3b), it shows that sometimes the concept statements are similar to variables ("aerosol emissions"), and at other times they are rather entities to which several attributes or variables could be attached (like "industrial activities"). Sometimes the concept points to a process, like "sun variability" in 3c); this is possible since – even though a process is something active, like a verb – one can substantiate this verb. One might as well posit the causal link in active voice, such as to make the link's arrowhead point in the same direction as the causal influence; however, there is no rule that would require this. Finally, 3d) "land use changes" is a case where the concept is a *change* (which will many times be perceived as "event").

Also, not all links refer to causality: in 3a) "is associated with" simply states that one concept has something to do with another one, and in 3c) and 3d) we see that they can create a logical order amongst the concepts. In 3b) the link "are generated by" is clearly of causal nature; however, it appears in passive voice. This is possible when links have names. One might as well have said "industrial activities [generate] aerosol emissions".

We note that concept maps are very flexible, because they do not impose many restrictions on the modeling process and the symbols: concepts may be variables, processes, entities or events. Links may be anything that is relating two or more concepts. Clearly, one may elaborate a concept map that contains only variables (as concepts) and causal links (as links and in active voice), but this would only be a special sub-class of all possible concept maps.

According to this, we conclude that concept maps are not meant to focus on causal effects; so it is not surprising that there is no particular definition of what causality is: it is just some influence.

This kind of modeling is easy to learn because it is intuitive; in particular, it does not separate structure from behavior. It is used to express and organize knowledge and also assess learning about complex situations like climate change (like in Rebich and Gautier, 2005).

[5] The diagrams have been elaborated with "CMapTools", a specialized software for creating, managing and analyzing concept maps. It can be obtained at

Cognitive mapping

As mentioned by Doyle and Ford (1999), “cognitive maps” mean different things to different authors. Tolman (1948) referred to maps that are constructed in the cognitive system; Axelrod (1976) was interested in mapping cognitive contents.

We discuss cognitive mapping as a “technique used to structure, analyze and make sense of accounts of problems” (Ackermann et al., 1996, p. 1; good introductions can also be found in Eden, 1990 and Bryson et al., 2004). This kind of modeling effort is concerned with uncovering and relating between each other assumptions, action possibilities, strategies and goals, hence the name of SODA (strategic options development and analysis). It is assumed that actions lead to outcomes over a causal link. Cognitive mapping is based on the theory of personal constructs (Kelley, 1973). Such a construct is like a chunk of discourse concerning some theme, for example climate change or a business problem. A typical chunk is about 10 words long; in addition to the positive statement (“centralize our services in Leeds”, see Ackermann et al., 1996, pp. 6-8) may be complemented by its opposite (“centralize our services in Leeds ... open local offices”, where “...” stands for “rather than”. A typical map developed from the mentioned construct might be the one shown in the following figure.

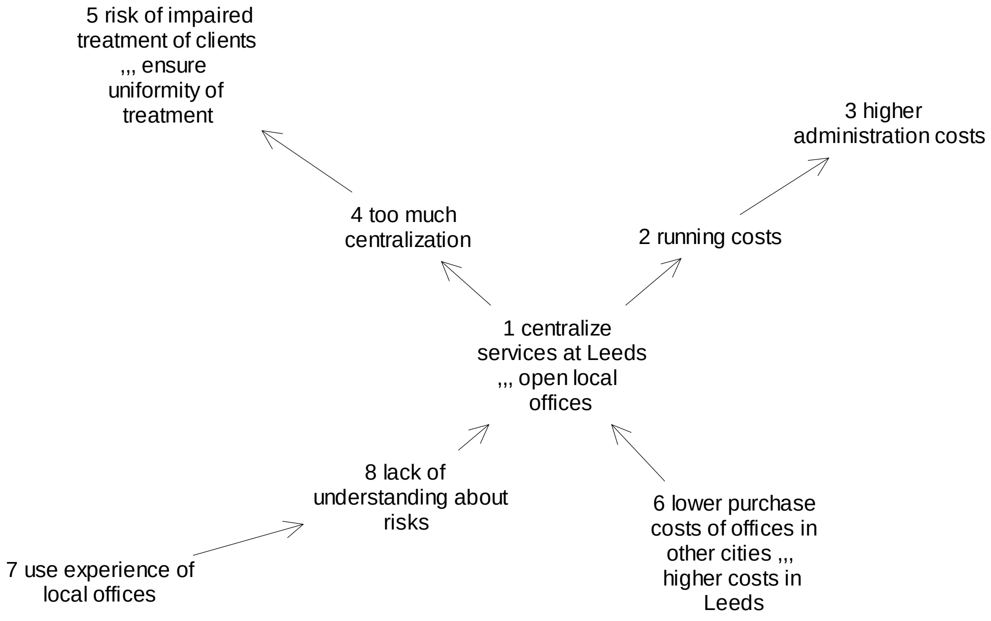
5 risk of impaired treatment of clients ,,, ensure uniformity of treatment
3 higher administration costs
4 too much centralization
2 running costs
1 centralize services at Leeds ,,, open local offices
8 lack of understanding about risks
6 lower purchase costs of offices in other cities ,,, higher costs in Leeds
7 use experience of local offices

Figure 5: example of a cognitive map (adapted from Ackermann et al., 1996, p. 10).

In these diagrams, the nodes (constructs) are numbered. This is useful later on for analyzing the map, in order to detect “heads” (no links coming in), “tails” (no links leaving), central nodes (heavily linked) and loops (amongst others) [6]. Clearly, some nodes refer to actions (1 and 7), while others mention what might be considered a variable, together with some information about its state (“lack of” “lower”, “higher”, “too much”) and sometimes dynamics. In some cases this qualification informs about an event going on in a variable, like in 3 “*higher* administration costs”.

[6] This map has been elaborated using “Decision Explorer”, a specialized software tool used for cognitive mapping, for example in SODA. It can be purchased at www.banxia.com.

The links mean “leads to” and are clearly about causation. “Higher administration costs” are the consequence of decisions taken upstream. However, it is also possible to insert “time links” denoting a “before/afterwards” relationship and “connotative links” that articulate a “has something to do with” relationship. Sometimes an action leads to an effect related to the negative pole of a construct; then the link is labeled with a “-“ sign.

Cognitive maps are elaborated to better understand a problem or an opportunity. This is why in SODA, the nodes are grouped into categories like goals, strategies, options and assumptions. The orientation is clearly causal, however it is not attempted to quantify or to simulate. This is justified by the fact that decision takers often are involved in delicate political systems that push not to disclose all the information and intentions; they are also and under time pressure, and would not be willing to use a complex tool that requires learning and long working sessions. Causal mapping is intuitive and easy to learn, since it does not separate behavior from its underlying structure, nor does it require to identify the variables. This may be worth some discussion (see Homer and Oliva, 2001), and one may wonder if it is not possible to complement cognitive mapping with system dynamics modeling; this has been done (Howick et al., 2006).

It has to be noted that in cognitive mapping, causality is assumed to link events (including actions).

Causal mapping

A causal map is a representation of causal beliefs of an individual or a group of individuals (Markóvski and Goldberg, 1995). In consists of variables and links that may indicate a type of relationship (“positive” or “negative”) and indication of strength (an integer number, often between 1 and 3).

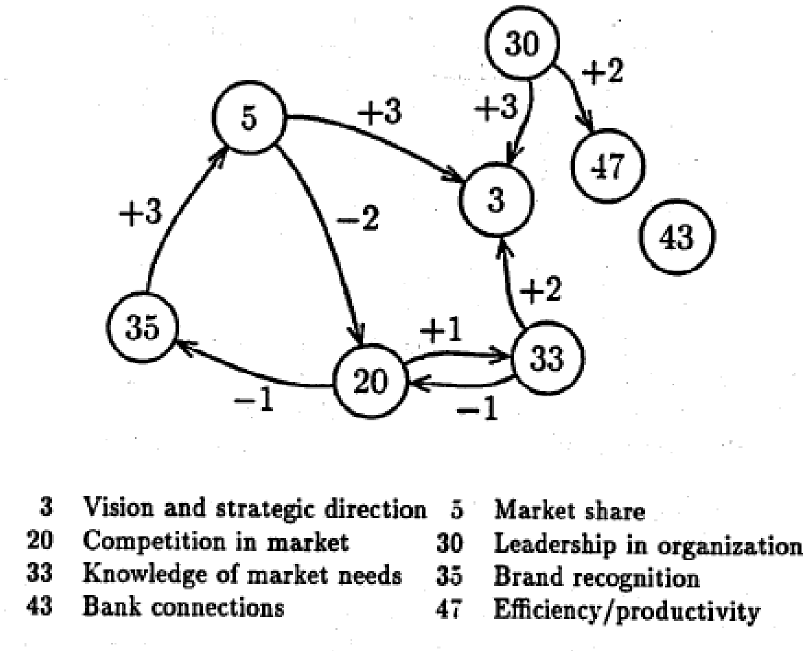
30
+2
+3
5
+3
47
3
+3
−2
43
+2
35
+1
33
20
−1
−1
3 Vision and strategic direction
5 Market share
20 Competition in market
30 Leadership in organization
33 Knowledge of market needs
35 Brand recognition
43 Bank connections
47 Efficiency/productivity

Figure 6: example of a causal map (Source: Markóvski and Goldberg, 1995, p. 307)

Just like in the case of cognitive mapping, the nodes in the diagram are numbered; however, note that now they represent variables, not the “constructs” including particular values or events.

In such diagrams, only structure is explicitly represented. Behavior has been abstracted away (no verbs in the labels). It is noticeable that the notion of *polarity* appears as “strength” of the links, understood according to what we will call the “popular” definition.

These diagrams can be represented as association matrix and analyzed for similarities and/or differences (Markóvski and Goldberg, 1995; Langan-Fox, et al., 2001; Langan-Fox, Code and Langfield-Smith, 2000; Langfield-Smith and Wirth, 1992). Even though the possibility to find feedback loops exists, these are not conceptualized. In these diagrams, causality is not explicitly based on events; rather it means “leads to” without specifying how the effect is transmitted.

Causal diagrams

Researchers interested in automated reasoning and the possibilities of its tools to theorize about cognition, often use causal diagrams (Gopnik et al., 2004; Halper and Perl, 2005a and 2005b; Pearl, 1995). Such diagrams contain variables and causal links, like shown in the following figure:

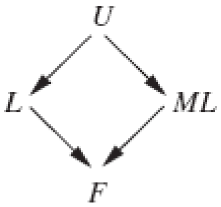

Figure 7: example of a causal diagram (Source: Halpern and Pearl, 2005a, p. 850).

The variables are meant to hold, at each moment, one of two or more discrete values. For example, Figure 7 is taken from an example where the cause for a forest fire has to be identified; the fire may have come from a lightning or a lit match dropped by somebody.

- F means “Forest fire” (F=1 means “fire”, F=0 means “no fire”);
- L means “Lightning” (L=1: “there is a lightning”, L=0: “no lightning”);
- ML means “Match Lit” (ML=1: “match lit”, ML=0:”no match lit”)
- U stands for the set of context variables, like the degree of oxygen in the air, if it is raining...

Causality means that two variables are linked by “a chain of events each directly depending on its predecessor” (Halper and Perl, 2005a, p. 844). In this sense, an event is what makes the value of a variable take on the specific value it has at a moment - for example, the occurrence of a “lightning” or its absence “no lightning”.

Strict formal rules are imposed on the causal diagram, assuring that only variables that are in the causal arc be part of the set of variables modeled as endogenous. A set of structural equations represents the way an event in one variable is caused by events in its preceding variables.

The rules and the equations make automatic treatment possible. Thus if we are in the possession of a causal diagram that satisfies the formal conditions, and we have the structural equations, and we know the context variable’s values, then we can automatically determine the cause of an event or the effect of an intervention.

However, like the example shows, the focus is on discrete events and the variable have discrete sets of values. Causal diagrams represent only structure, the dynamics of events has been abstracted away.

Summary

We may now outline the relevant findings about the different approaches, as shown in the following table:

Type of diagram	***Main use***	***Nodes***	***Links***	***Driving force behind causality***	***Degree of abstraction***
Concept maps	Structure knowledge in form of propositions	Concepts in general	Links of any type	Undefined	Low
Cognitive maps	Uncover, structure and analyze problem accounts	Constructs (phrases with entities, variables and descriptions of their values or changes	Qualitative causal links including negative and positive type	Events, actions	Low
Causal maps	Represent beliefs about the causal structure.	Variables	Qualitative causal links, including polarity (popular definition)	Undefined	High
Causal diagrams	Determine causes of events or consequences of interventions (actions)	Variables with discrete values	Quantified causal links (structural equations)	Events	High

Table 1: comparation of non-SD diagrams related to causality.

The diagramming languages that allow nodes to be simply concepts or constructs do not impose on the modeler the task of defining the variables. These diagrams are more intuitive to construct and to read. However, concept maps are not meant to aid in decision problems, and cognitive maps are developed to articulate and organize ideas such as to discover goals, strategies and actions in a world of discrete events that are not quantified; these “models” are not meant to inform about the evolution of the quantities of the variables - they don’t even force to specify variables.

Causal maps and causal diagrams force to specify the variables. This makes them more abstract to elaborate and to interpret. However, they are not intended to generate quantitative behavioral information: causal maps represent what people believe the causal structure to be and search to compare these beliefs. Causal diagrams are about what happens between events as cause and as effect.

5. Causality and behavior - system dynamics

Causality for system dynamics

For system dynamics, the main concern is to understand how structure (variables and causal links) generates behavior in a world of continuous processes, rather than discrete events. On its way, system dynamics has created specific concepts as well as symbols to represent them: the signed feedback loop, the type of variable (accumulation or flow), the delay and – where quantification comes into play, the non-linearity.

Initially, the "stock-and-flow" diagrams were used as graphical language; they had a symbol for each of the special concepts and allowed quantification, which is necessary for simulation. There is wide agreement on that the presence of multiple feedback loops makes simulation necessary in order to assess the continuous behavior.

System dynamics has a very specific definition of causality, which has to do with its conceptual universe. The world is assumed to be in continuous development, and can be described by two types of variables. If we define behavior as the different values a variable takes over time, then stock variables are *accumulators* or *stocks* with a behavior determined by their own value "just before" and the sum of all connected *flows*. The behavior of a *flow* variable is defined for a period of time; it is determined only by the currently visible value of *stock* variables and possibly some converters (or auxiliary variables). Thus there is a fundamental difference between the two types of variables, since accumulators depend on their state and the connected flow rate's quantity. *Flow* rates depend on the quantities in *stocks*.

Also, whereas in the philosophic discussion about causality (Halper and Pearl, 2005a), causality has to do with how events cause events, system dynamics deals with how behavior causes behavior (through a causal structure; in this, there is agreement between the communities). So what is *behavior* and how does it relate to *events*?

In a continuous world, a stock variable has a specific value at each point in time. Flow variables are defined for a period of time. Since perception occurs at points in time, only stock variables can be directly perceived - even though the mind computes behavior from sequences of perceptions and is more sensitive to changes that to states (Kahneman, 2002). *Behavior* is then the change of values (the slope or first derivative of a function). An *event* can only be a *change in behavior* (change of slope or second derivative of a function).

This is something the other approaches presented above do not touch upon: concept maps, cognitive maps and causal maps do not pretend to reason in quantitative terms (so the do not need an operational definition of causality). Causal diagrams deal with *events*, and *behavior* does not appear in the same way. In the example of the forest fire, there is fire or there is not – there is no need to ask for the forests fire behavior before or afterwards.

The following figure illustrates some cases:

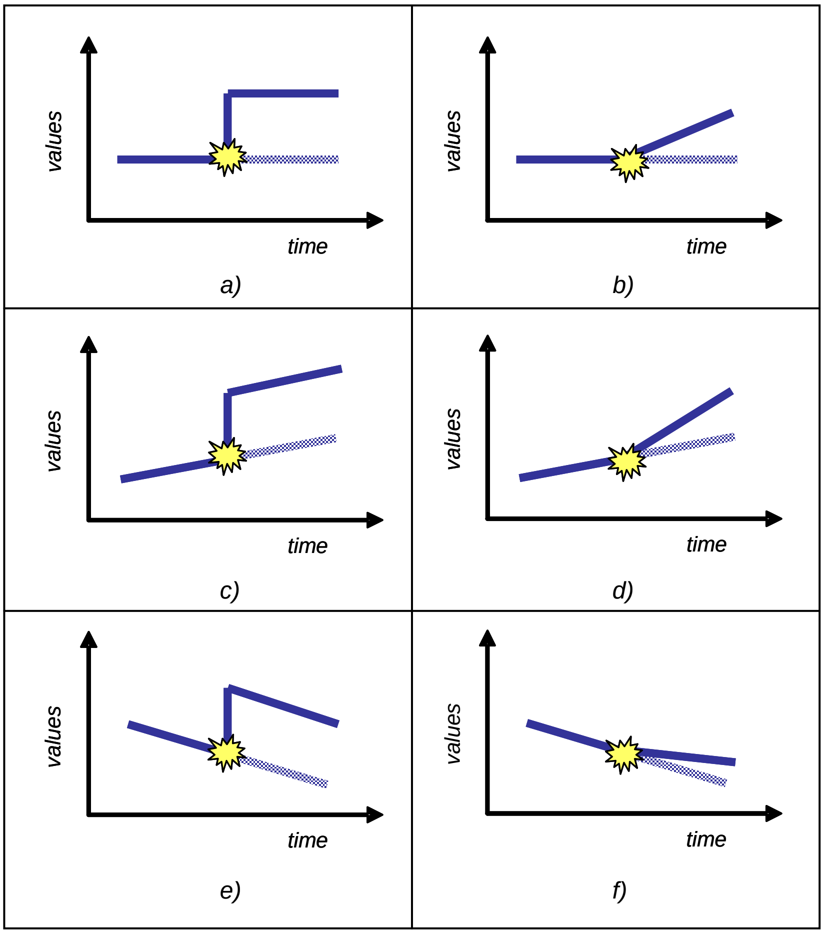

Figure 8: a "step" and a "ramp" event as a change of behavior (change of values) above different behaviors.

It seems that for system dynamics' concerns, there are two types of fundamental events: one is a vertical translation of the graph, represented as a "step" event (figures a, c and e). The second one is a change of slope, represented by a "ramp" event (figures b, d and f). Other forms like a "pulse" event are combinations of fundamental events, like two "step" (one upwards, the other downwards) [7].

While figures 8a and 8b do not look unusual, they refer to cases when the variable is static during the whole considered period of time, except the change event. Clearly there are many cases when this is not so: the employment of a country's economy may be rising (slowly) but the government wishes to push it up or rise fasted (cases c and d). Or the sales of a company are descending and management would like to give them a push or at least slow down the decrease (cases e and f). This is one characteristic that distinguishes system dynamcis from the other approaches presented above: here we care not only for events, but for behavior in general.

The very specific relationship between flow and stock variables means that between two variables, there are rules as for what kind of event causes what kind of event:

– a "step" event in a flow variable causes a "ramp" event in the stock variable; if the flow is an inflow, then the ramp has the same sign as the step; if the flow is an outflow, the ramp has the inverse sign of the step.

[7] One may wonder if a "ramp" event could not be decomposed in a sequence of little "step" events; after all, an instruction like "ramp(1,10)" is like a superposition of "step(1)" from time "10" on.

– a "ramp" event in a flow variable causes a non-linear "ramp" event in the stock variable [8]; if the flow is an inflow, then the stock's ramp has the same sign as the flow's one; if the flow is an outflow, the stock's ramp has the inverse sign of the flow's one.

We see that there is never a "step" event in a stock variable (unless we equal "ramp = sequence of steps").

Flow variables are free to react to a stock's "ramp" event in any manner, bound only by the polarity of the link.

The relationship between behavior and event, as well as the rules derived from the relationship between stock and flow variables, are inherent in the "stock-and-flow" diagram language. This means that such a diagram, thanks to these incorporated regularities, can be simulated once quantified. It then helps to understand the connection from structure to behavior.

Causal loop diagrams

Later on, the so-called "causal loop diagrams" (CLD) started to be used as a means to communicate selected insights from a simulation study (Homer and Oliva, 2001). Once this diagram language existed, it became tempting to use it also in other phases, especially for articulating causal beliefs in the early phases of modeling projects. This diffusion of CLD has lead to two debates: for once, there is argument on if simulation is always necessary or recommendable. Also, there has been criticism of the simplification of polarity. Finally, one may wonder why CLD have become so much more popular than "influence diagrams" (Wolstenholme, 1990), if the former do not distinguish between accumulators and flows, but the latter do. Maybe this can be explained by the simplicity that is won by forgetting about the difference.

In order to see the fundamentally limited character of CLD, we will now examine them from scratch.

A causal link is given between two variables when a subject believes that what happens in one of the variables will cause some consequence in the second variable. We may call the first variable "independent variable", since it does not depend on any variable that appears inside the (mental) model; the second variable may be called "dependent variable" for the obvious reason of being influenced by the first one [9]. An articulated causal link is then the smallest possible mental model.

In system dynamics, the arrow representing the causal link has the usual arrowhead to indicate the direction of influence and also a "+" or "-" indicating the link's polarity. Polarity is what determines an essential quality of the cause's effect. Each of the two variables is a placeholder for a series of quantities: the *cause* will be observable as a distinct behavior of the independent variable, and the dependent variable will display a distinct behavior as *effect*. This behavior will be an increase or a decrease. During all this time, we should not forget that both variables always "behave" in some way, so neither the cause nor the effect can be something different from a modification of behavior. Since we are thinking in terms of a continuous world, our variables have to have a behavior; the "events" that are perceived and thought of as cause and effect must then be a change in the variables' behavior.

We often use PULSE, STEP or RAMP functions to introduce such "events" and observe their effects. This is testimony of the fact that humans perceive change more easily that state (Kahnemann, 2002) and that habituation makes a constant behavior appear as state.

[8] If we define "ramp" like in the preceding footnote, then it becomes clear that he sequence of "steps" in the flow variable will cause a sequence of superposing "ramp" events in the stock variable.

[9] This is of cause only possible as long as there are no loops; however, *polarity* does not need a loop to be explained and our reflection suffices with one link.

Also, we should recall that there will be other variables: no interesting system will consist of only two variables: consequently, the variable we call “dependent” during this mental exercise may be subject to other influences. As will be shown below, this has not always been the case in the debate.

Let us assume that for one instant, there are no other variables. Now, when an increase event (▲) in the independent variable triggers an increase event (▲) in the dependent variable, then we have a case of what is intuitively classified as positive polarity:

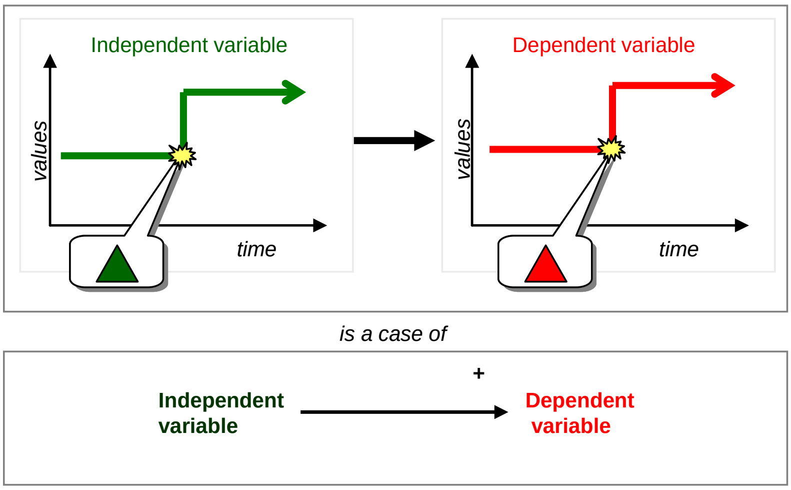

Figure 9: an example of positive polarity

When we think in terms of a variable, then the variable is a shortcut for all the instances of behavior we have seen before. In a way, Hume’s logic goes from the upper part of the figure, where the behaviors are, to the lower part which corresponds to an abstraction of the variables and a judgment of “causal link”. If we have two variables capable of two types of changes each, it is clear that there are 4 basic configurations:

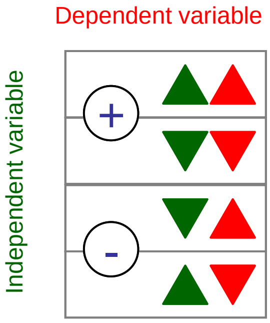

Figure 10: 4 basic configurations for 2 polarities

According to this state of affairs, the simple or “popular” definition of polarity seems rather convenient:

- *positive* (+): when the independent variable changes, then the dependent variable changes in the *same* direction;
- *negative* (-): when the independent variable changes, then the dependent variable changes in the *opposite* direction.

When both the independent and the dependent variables distinctively increase or decrease, we speak of *positive* polarity. When one distinctively increases and the other decreases, we speak of *negative* polarity.

However, this definition is easily shown to be flawed (Richardson, 1997): if we take the mini model where the birth rate influences the population with a positive link, an increase in the birth rate would cause an increase in the population. However, is it meaningful to assert that a decrease of the birth rate would bring about a decrease of population?

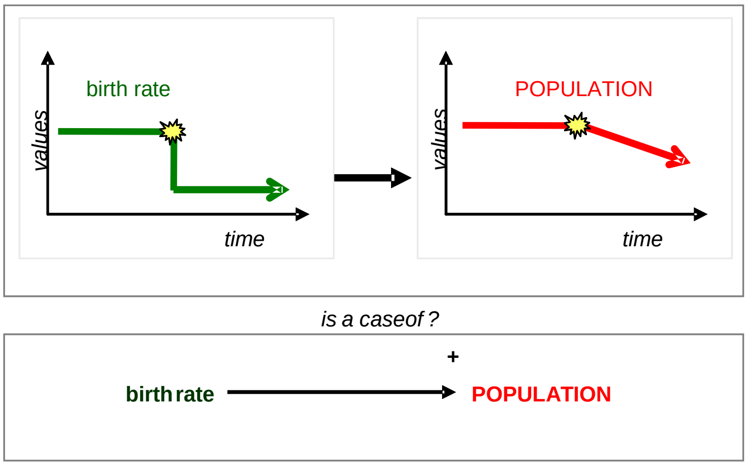

Figure 11: can the birth rate lead to lower population?

As indicated by Richardson (1997), this apparent problem is due to the popular definition and is overcome by using the full definition.

- positive (+): when the independent variable changes with a particular sign (+ or -), then the following values of the dependent variable will be above (or less) than what they would have been.
- negative (-): when the independent variable changes with a particular sign (+ or -), then the following values of the dependent variable will be less(or above) than what they would have been.

This is the "official" definition that students find in the textbooks (for instance, Sterman, 2000). With this definition in mind, there is certainly no problem with thinking that a decrease of the birth rate causes population to decrease below the level it would have had without the birth rate's decrease. One can attribute this, for instance to the effects of other variables that cause changes in population, like the death rate.

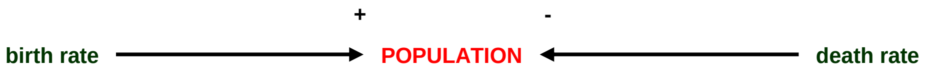

Figure 12: the birth rate can lead to lower population

In Figure 12, there is a new variable that influences the population. It is intuitively understood that the death of people reduces the population; so one can easily imagine that without births, the usual behavior of population is downwards. So if there is a sudden "step" decrease in the birth rate, the population will respond with a negative "ramp" and will be lower than it would have been. (Of cause, if one wishes to understand why a "step" cause provokes a "ramp" effect, one has to re-introduce the distinction between flow and accumulation.) There may be a rather diverse set of possibilities for the basic behavior of population, in which the birth rate influences in a different manner after the "event" constituted by its change of behavior:

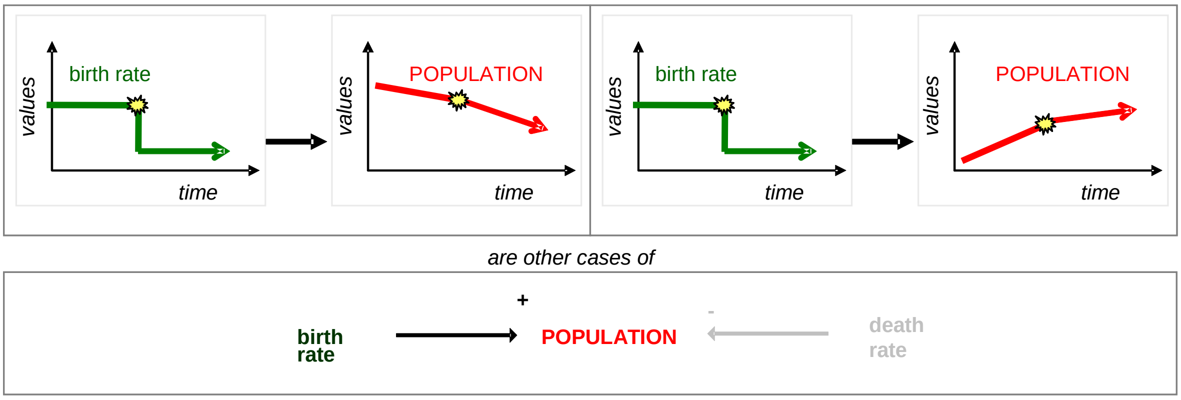

Figure 13: other cases of behavior of the same model

On the left hand of this figure, we see that a downwards step in the birth rate causes the decline of population to accelerate. On the right hand, the same cause triggers population growth to slow down. Both cases (and more if one wishes) are included in the set of behaviors that the model structure represented in the causal loop diagram can generate.

Why, then, is the simple definition so popular? It continues to be used in so-called systems thinking [10] publications and scholarly publications (Warren, 2004). The suggested reason is that it is simpler. After all, it is easier to think in terms of "events" and forget about the rest of behavior, i.e. do as if it did not exist (constant values). This would bring us back to the model without death rate, where the population is implicitly assumed to be constant when there are no births. It is easier to focus on what happens between the birth rate and population without bearing in mind the existence and effect of the death rate. We see now that this erroneous judgment is attributable to the simple way we have to think about these variables, since causal loop diagrams do not enforce the observation of the rules and regularities stated in the previous section. They allow for these simplifications – at the price of implicitly assuming false behavioral consequences. In particular, the causal arrows in causal loop diagrams appear to symbolize only events, whereas system dynamics is interested in behavior in general.

However, there is more to be criticized about causal loop diagrams, which cannot be overcome with the "complete" definition.

Not only "events" (as a special episode of behavior) possess causality; ordinary – constant – behavior also does. As soon as we acknowledge the difference between flows and accumulations, it becomes inevitable to think that the absolute value of the birth rate (flow) causes behavioral effects in the population (accumulation). This is well-known to those who master graphical integration (see for example Sterman, 2000); however, for the sake of explicitness, the following figure illustrates the case:

[10] The choice to call *system dynamics without stocks and flows and without simulation* "systems thinking" is a debatable one. For the sake of those who use the distinction and do simulate, are they not thinking? And as far as the communication with people from outside the system dynamics community is concerned, why are they supposed to understand then inside system dynamics "systems thinking" does mean something very different from what it usually does?

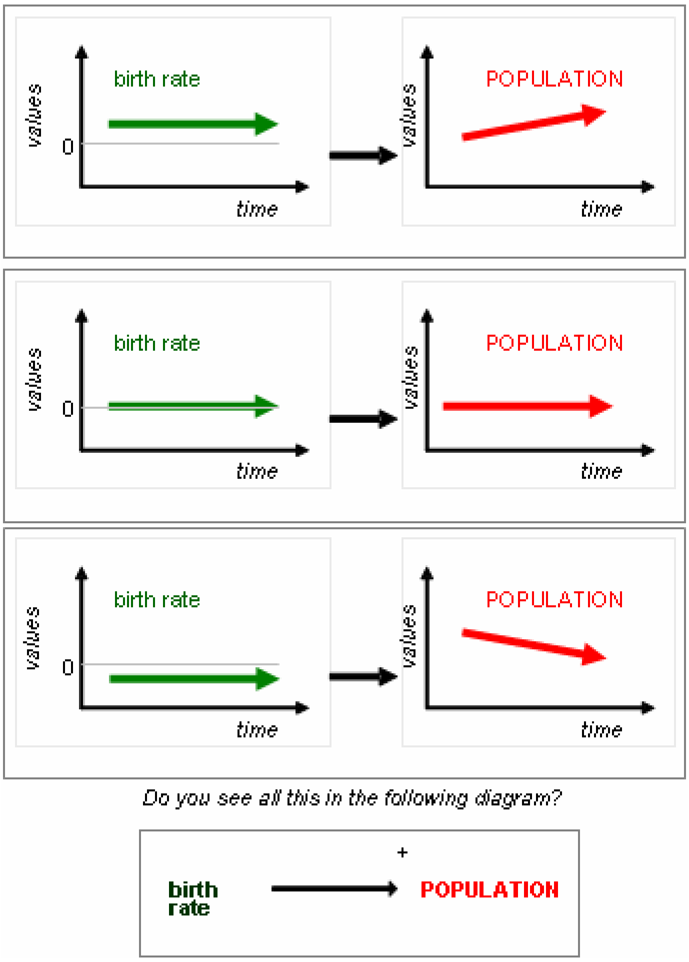

Figure 14: examples of causality without event

As long as an individual bears in mind the fact that the birth rate is a flow and the population is a stock, as well as the relationship between these types of variables, this individual may indeed "see" all these possibilities in the diagram. However, if CLD are meant to be used with people who do not master system dynamics, would this individual want to use this type of diagram?

Richardson (1997) as mentioned an alternative definition of polarity:

- *positive* (+):the independent variable *adds to* the dependent variable;
- *negative* (-):(+):the independent variable *subtracts from* the dependent variable.

If we think "the birth rate adds to population" then the usual behavior and the change event (say, a "step") are accurately reflected. However, even if we replace CLD by "influence diagrams" (Wolstenholme, 1990) and population appears as stock and the birth rate as rate, it is impossible to express the causal effect of the birth rate for population on a diagram (unless we add a new symbol category like ">", "0" and "<" to mean "positive", "zero" and "negative" respectively. Even though this may in principle be possible, it would deprive the diagrams of their simplicity. Note also that a rate that can be positive or negative corresponds to a "bi-flow"; inflows and outflows implicitly carry a restriction on the range of possible values, which causal loop diagrams do not express.

The last limitation is that causal loop diagrams do not have a means to represent the "smoothing in" of the caused effect over time. As observed by Moxnes (1998) and Moxnes & Saysel (forthcoming), people appear think statically: they assume implicitly that the effect of an intervention in a control variable will realize itself immediately. This is, of cause, an erroneous assumption that leads them into faulty decisions. It can be attributed to not recognizing stock variables as such (and not adequately taking into account the behavioral implications of this fact). Again, one would have to be a accustomed dynamicist to bear in mind all this while working with a causal loop diagram.

It has to be concluded that causal loop diagrams cannot capture what causality really means for system dynamics, since they talk only about events and not behavior in general. For system dynamics., causality includes behavior, the way the value of a variable is determined in each period of time (not the way it changes is determined, which is only a special part of it).

What are causal loop diagrams good for?

After all, it turns out that causal loop diagrams are a qualitative tool that should be used only when it is not demanded to process structure into behavior or to reason about behavior. If used with the complete or the "adds/subtracts" definition *after* having understood the behavior via simulation, it can help to make sense of the behavior found; still, if the specific notions of system dynamics help to better think about behavior, the pedagogical value of causal loop diagrams may be put into doubt (since they avoid confronting oneself with these things).

Maybe causal loop diagrams are a useful tool for exploring beliefs concerning the causal structure of a situation. However, in situations where it is important to understand the behavioral consequences of decisions (interventions), the "automated reasoning" of simulation becomes necessary or at the very least desirable. One could probably figure out some plausible guesses concerning the probable quality of behavior in cases without feedback loops or when all the loops have the same polarity; but these are extreme cases and should not be used to design a rule. This would require a notational change in causal loop diagrams (distinguish stocks and flows and add the ">,0,<"); additionally, one would have to identify the "tails" (independent variables) and trace their assumed behavior, and then trace the superposed influences through all the causal links. A lot of work: probably causal loop diagrams would be just "stock-and-flow" diagrams and they would not be so popular any longer

So it appears that causal loop diagrams are so popular because they are well adapted to intuitive thinking. But if the situation at hand requires computations that intuitive thinking is bad at, one would expect the tools to enhance this thinking (and not to have the same shortcomings). On the other side, if the situation is politically delicate or time presses, an approach like SODA may be more convenient: it does not force to separate structure from behavior and thus is easier to use.

In which way can the use of causal loop diagrams help to discover or articulate (attributed) causal links? How are they usually recognized? Let us return to the first two phases presented in the conceptual model in Figure 3 (p. 5):

1) how do we come to consider something as a relevant variable?

2) how do we come to believe in the existence of a causal link between two variables?

We will try to explore this question using the different tools on the population example.

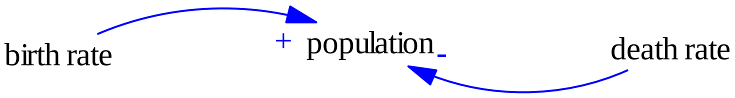

Figure 15: the population case as a causal map or causal loop diagram

We start restating the typical CLD, where most people are tempted to use the "popular" definition and thus project structural information into behavioral conclusions that can easily be mislead.

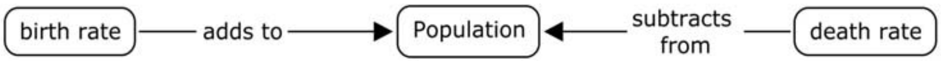

Figure 16: the population case as a concept map

When the same case is expressed as a concept map, there seems to be exactly the same structural information as in the CLD. Note that in this example, I have used the "adds to" definition (however, one can also name the links according to the other definitions). The fact of doing so may be seen as a *little* help to think adequately about the causal relationships.

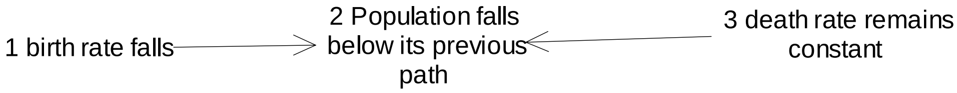
1 birth rate falls
2 Population falls below its previous path
3 death rate remains constant

Figure 17: the population case as a cognitive map

When the story is told by a cognitive map, we see that behavioral aspects are now explicitly stated in the "constructs" (and it is not necessary to compute them based on the polarity indications). This is less general than a causal loop diagram (that offers the structure for all cases of behavior without contaminating it with information about one of the possible behavior contexts).

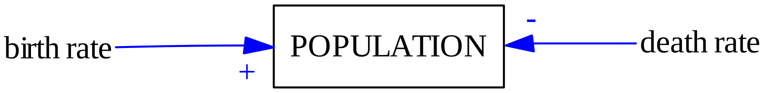

Figure 18: the population case as an influence diagram

The influence diagram recalls us that "population" is of a different type than the other variables. This is not expressed in the CLD; however, the observer has to do the thinking in order to understand what this difference means.

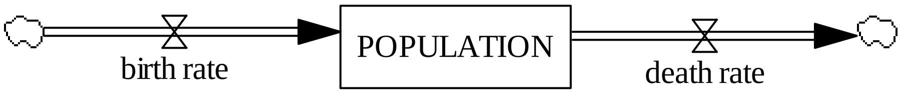

Figure 19: the population case as a stock-and-flow diagram

Seen as a stock-and-flow diagram, it is easier to see that the birth rate “adds” and the death rate “subtracts”. But even so: what difference does this make?

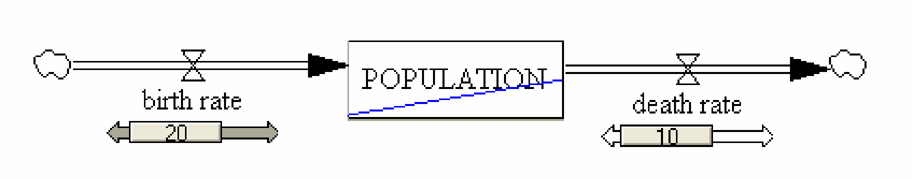

Figure 20: the population case as stock-and-flow diagram with behavioral information

Everything changes with simulation. Each of the mayor modeling software packages has its own way to visualize the variables’ behavior in the context of the stock-and-flow diagram. Now one can readily see what the current structure, together with the parameter context, means in behavioral terms.

If structure is inferred from behavior and then is used to generate behavior, using a tool that represents only the structure forces the user to carry the behavioral part on his mind. While this can be expected from an experienced user, a beginner is left with a problem. Due to the apparent simplicity of the tool, he may rapidly gain a feeling of progress; however, his mental model of what the diagram means may contain all kinds of errors concerning the model’s behavior, and since this part does not become articulated, it is not open to critique and the subsequent improvement.

It follows that causal loop diagrams should not be used without the guidance of an experienced modeler who will point out the adequate behavioral aspects.

The population case is admittedly a very simple one. Since it does not involve feedback loops, the causal loop diagram could not play out its notational advantage over concept maps or cognitive maps. Neither did we gain much by using an influence diagram. The stock-and-flow diagram was not more complex to draw, but it had the big advantage of allowing simulation, enabling the tool to show the behavior. This real-time feedback of the meaning of the structure as a huge help, also during the modeling, since it helps to discover inconsistencies and errors. It cannot possibly be offered by CLD.

In more complex cases, where there are interacting feedback loops, causal loop diagrams have one advantage over the other diagram languages: feedback loops exist as a concept and have a symbol to be represented explicitly in the diagram. However, event the most expert of system dynamicists will not be able to guide his clients through the behavioral interpretations without simulating.

In conclusion: CLD should be banned from use as decision-finding tool and limited to exploring causal (structural) beliefs – where they more expressive power than the other tools presented here:

- concept maps and cognitive do not focus on variables and causal structure and will waste mental energy; they may be even more intuitive, but their product has still to be translated in terms of variables;
- causal diagrams strive to understand cases better described in terms of discrete events;
- causal maps lack the possibility to express “feedback loop” and “delay”.

Another recommendable use might be for simplification purposes when explaining the simulation model's behavior. However, there remains a big doubt as for how neatly one can mentally separate behavior from structure: after all, we cannot infer structure without behavior and we cannot conjecture behavior without structure. In the end, the "concept model" approach described by Richardson (2006) may be more fruitful for system dynamics.

6. Conclusions

This article started remarking that despite its fundamental importance for system dynamics, polarity and causal thinking have not received much attention in the specialized literature. The apparent effort to learn the meaning of polarity was the starting point of this inquiry. An examination of other approaches that deal with causality and use more or less similar diagram languages has shown that usually causality is only very broadly defined, and where it is operationally defined, this is done with respect to events rather than behavior.

In contrast to these approaches, system dynamics is about behavior rather than events. We have proposed to understand an event as a change of behavior, the second derivative of a function. Part of being attentive to behavior is that there are specific differences between accumulators and flow variables which have implications for behavior. We have found that causal loop diagrams could be used in a way that avoids the flaws shown in previous critiques: for doing so, one has to use the full definition rather that the popular one. However, the popular definition continues to be just this: popular.

We have then found that causal loop diagrams are not able to help taking into account behavior in a way that keeps coherent with system dynamic's conceptual world, especially the relationship between flow and stock and the gradual realization of effects. It is as if causal loop diagrams are better equipped to think about events than about behavior. This would help to understand why they are so popular, just like the popular definition, and why my students find it much harder to understand the full definition.

Finally, if behavior is even harder to think with than events, tools for causal thinking should help in keeping coherence in our minds between structure and behavior. This is exactly what our modeling software does. So the diagram language of "stock-and-flow" is not only able to express important aspects of a modeled system, it also allows the computer to enhance our mind.

This is certainly not the end of my inquiry. In closing these lines, it has become my conviction that executable "concept models" have many advantages over causal loop diagrams, that should be used only by experienced modelers. One personal consequence is that I'm now very busy replacing all the causal loop diagrams of my educational materials by simple stock-and-flow models.

In the future, it should be asked how we recognize variables and causal relationships – steps 1 and 2 in Figure 3 (p. 5). Many beginners find it hard to come up with variables at the outset of the modeling process; "ask the client" only shifts the burden to other people. If we believe that humans are sensitive to change (to events), then how do we articulate our mental contents as variables? Also, if theories about causal learning are based on events, what is the status of behavior (the slope of the curve or the first derivative of the function)?

System dynamics has its advantages, since it can help to improve understanding of situations where the other approaches do not have a lot to offer. However, most adults nowadays not only ignore system dynamics, they *do* know other approaches and it is well known that most adults prefer avoiding changes. If we wish more people to become system dynamics thinkers – that is to say when we think as educators – then it may be helpful to know the answers to these questions.

7. References

Ackermann, F., Eden, C. and Cropper, S., 1996, Getting started with cognitive mapping; www.banxia.com/depaper.html (12/3/2007)

Allan. L. and Tangen, J. 2005. Judging Relationships Between Events: How Do We Do It? Canadian Journal of Experimental Psychology; Mar 2005; 59, 1; pg. 22

Argyris, Ch. 1993. Action science, Addison Wesley

Axelrod R (ed.) 1976. The Structure of Decision: The Cognitive Maps of Political Elites. Princeton University Press: Princeton, NJ.

Booth-Sweeney, L. and Sterman, JD., 2000. Bathtub dynamics: initial results of a systems thinking inventory, *System Dynamics Review* **16**(4): 249–286

Bryson, M., Ackermann, F., Eden, C. and Finn, Ch. 2004. Visible thinking – unlocking causal mapping for practical business results, Wiley

Doyle, F. and Ford, A. 1998. Mental model concepts for system dynamics research, System Dynamics Review **14**(1), (Spring 1998): 3-29

Doyle, F. and Ford, A. 1999. Mental model concepts revisited: some clarifications and a reply to Lane, System Dynamics Review **15**(4), (Winter 1999): 411-415

Eden, C. 1990. Using cognitive mapping for Strategic Options Development and Analyses (SODA), in J. Rosenhead (Ed.), *Rational Analysis for a Problematic World*, Wiley: Chchester

Gopnik, A.; Glymour, C. ; Sobel, David M. ; Schulz, L. E. ; Kushnir, T. and Danks, D. 2004. A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 2004, 111, 1, 1-30

Halper, J. and Pearl, J. 2005a. Causes and Explanations: A Structural-Model Approach. Part I: Causes, Brit. J. Phil. Sci. 56 (2005), 843–887

Halper, J. and Pearl, J. 2005b. Causes and Explanations: A Structural-Model Approach. Part II: Explanations, Brit. J. Phil. Sci. 56 (2005), 889–911

Heider, F., 1958. *The Psychology of Interpersonal Relations*. New York: Wiley

Homer, J. and Oliva, R. 2001. Maps and models in system dynamics: a response to Coyle, *System Dynamics Review* Vol. 17, No. 4, (Winter 2001): 347–355

Howick, S. Ackermann, F. and Andersen, D., 2006. Linking event thinking with structural thinking: methods to improve client value in projects, *System Dynamics Review* Vol. 22, No. 2, (Summer 2006): 113–140

Hume, D. 1984 Enquiry concerning the human Understanding [Investigación sobre el conocimiento humano], (Translated by Jaime de Salas Ortueta), Alianza Editorial, Madrid

Johnson-Laird, P. N. 1999. Deductive reasoning, *Annual Review of Psychology*; 1999; 50: 109-135

Kahnemann, D. 2002. Maps of bounded rationality: a perspective on intuitive judgment and choice, Nober Prize Lecture, December 8, 2002

Kelley, H. H. 1973. The processes of causal attribution. *American Psychologist*, 28, 107-128.

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

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

López, F., Cobos, P. and Caño, A. 2005. Associative and causal reasoning accounts of causal induction: Symmetries and.asymmetries in causal predictions and diagnostic inferences, Memory & Cognition; Dec 2005; 33, 8; pg. 1388

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

Moxnes, E. and Saysel, A. Misperceptions of Global Climate Change: Information Policies

Moxnes, E., 1998, Overexploitation of renewable resources: The role of misperceptions, Journal of Economic Behavior & Organization Vol. 37 (1998) 107-127

Moxnes, E., 2000. Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development, *System Dynamics Review* **16**(4):325–348

Moxnes, E., 2004. Misperceptions of basic dynamics: the case of renewable resource management, *System Dynamics Review* **20**(2): 139-162

Payne S. and Baguley Th., 2006. Memory for the process of constructing an integrated mental model, *Memory & Cognition*; Jun 2006; **34**(4): 817

Pearl, J. 1955. Causal diagrams for empirical research, *Biometrika* (1995), 82,4, pp. 669-710

Pedercini, M. 2006. What's behind the blue arrow? - The notion of causality in System Dynamics; *Proceedings of the 24th International Conference of the System Dynamics Society*, Nijmegen

Rebich, S. and Gautier, C., 2005. Concept Mapping to Reveal Prior Knowledge and Conceptual

Change in a Mock Summit Course on Global Climate Change, *Journal of Geoscience Education*, v. 53, n. 4, September, 2005, p. 355-365

Reichel, A. 2004. (Re-)Structuration of System Dynamics, *23rd International Conference of the System Dynamics Society*, Oxford

Richardson, 1991 *Feedback thought in social sciences and systems theory*. Pegasus Communications

Richardson, G. P. 1997 Problems in causal loop diagrams revisited, *System Dynamics Review* Vol. 13, No. 3, (Fall 1997): 247–252

Richardson, G. P. 2006 Concept Models. *Proceedings of the 24th International Conference of the System Dynamics Society*, Nijmegen

Richmond, B., 1993, Systems thinking: critical thinking skills for the 1990s and beyond, System Dynamics Review Vol. 9, no. 2 (Summer 1993):113-133

Sterman, JD. 1989. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science* **35**(35): 321-339

Sterman, John, 2000. *Busyness Dynamics – systems thinking and modeling for a complex world*, John Wiley

Tolman EC. 1948. Cognitive maps in rats and men. Psychological Review 55: 189±208.

Warren, Kim, 2004. Why has feedback systems thinking struggled to influence strategy and policy? *Systems Research and Behavioral Scienc*; Jul/Aug 2004; 21, 4; pg. 331

Wolstenholme, E. 1990. Systems Enquiry. Wiley