 doi:10.20944/preprints202107.0004.v3

Article

A Conceptual Framework for Constructing Decision Policies by Processing the Possibilities in Mental Models of Dynamic Systems with the Cognitive Theory of Mental Models

Martin FG. Schaffernicht [1], Miguel López-Astorga [2], Ramón D. Castillo [3] and Cristian A. Rojas-Barahona [4]

Abstract

This article is a theoretical contribution to mental model research, which currently has different threads. Whereas some researchers focus on the perceived causal structure, others also include decision policies and decisions. We focus on the link between recognized causal structure ("mental models of dynamic systems") and policies, proposing Johnson-Laird's theory of mental models as the link. The resulting framework hypothesizes two types of systematic mental model errors: (1) misrepresentation of the system's structure and (2) failure to deploy relevant mental models of possibilities. Examination of three experiments through this lens reveals errors of both types. Therefore, we propose that the cognitive theory of mental models opens a path to better understand how people construct their decision policies and develop interventions to reduce such mental model errors. The article closes by raising several questions for empirical studies of the reasoning process leading from mental models of dynamic systems to decision policies.

Keywords: Mental Models; Dynamic Decision Making; Systems Thinking; Learning;

Introduction

This article contributes to mental model research in dynamically complex situations, emphasizing how mental models lead to decision policies. The notion of *mental models* is ubiquitous in the system dynamics approach. Jay Forrester described them as unavoidable fundament for human reasoning, decisions, and action (Forrester, 1971). In his retrospective of the first 30 years of system dynamics, he pointed out the incompleteness, the internal contradictions, and the flaws in expected consequences of decisions, already hinting at a link from mental models to decision policies (Forrester, 1987). In the wider field of management research, mental models are how people represent knowledge (for a historical overview, see Johnson-Laird, 2004).

In system dynamics, two threads of mental model work go beyond this general notion. First, *mental models of dynamic systems* or MMDS (Doyle and Ford, 1998, 1999; Groesser and Schaffernicht, 2012; Lane, 1999) represent people's knowledge of a dynamic situation. To date, only a few published studies have examined MMDS empirically or furthered the methods for their analysis. Second, Gary and Wood (2016) proposed a layered view of mental models, where (1) perceived causal relationships underpin (2) strategies that frame (3) decision rules, which then drive (4) performance. However, their main attention was on decision rules and the role of heuristics in driving decisions. This brings their research close to "misperception of feedback" and "stock-and-flow error," where observable behavior and

[1] Facultad de Economía y Negocios, University of Talca (Chile)

[2] Institute of Humanistic Studies Abate Molina, University of Talca

[3] Facultad de Psicología, University of Talca

[4] Facultad de Psicología, University of Talca

 doi:10.20944/preprints202107.0004.v3

heuristics play a prominent role but usually do not directly elicit and analyze mental models. Yet, the term *perceived causal relationships* is conceptually equivalent to MMDS, and in either case, the focus is on knowledge structures. Hence, several directions for further research proclaimed by Gary and Wood also apply to the MMDS thread. These directions point to: (a) sharpening terminology, (b) advancing mental model measurement, (c) advancing knowledge of properties of mental models, especially errors like superstitious beliefs or cognitive blind spots; and especially (d) advancing knowledge of links between the distinct layers of mental models.

Decision rules based on heuristics link with decisions, but there remains a gap between the perceived causal structures and the strategies (or policies) leading to the decision rules. How do people use their MMDS to devise a policy (or strategy)? We propose that the *mental model theory* (Johnson-Laird, 1983, 2010) bridges this gap. This is a psychological theory of reasoning: when faced with a decision situation, people deploy and process *mental models of possibilities* (MMP) to conclude which imaginable possibilities hold true[i]. Such reasoning can be intuitive or deliberate, the difference being if only obvious possibilities are processed, or all possibilities compatible with prior knowledge are considered (Johnson-Laird and Ragni, 2019; Khemlani, Byrne, and Johnson-Laird, 2018; Ragni and Johnson-Laird, 2020a). This is amenable to mentally processing the possible behavioral consequences of a decision, and it arguably leads to rule-like decision policies.

We introduce a conceptual framework in which the variables, causal links and feedback loops in the MMDS are the raw material for deploying MMPs, whose processing yields behavioral expectations leading to decision policies. This framework implies two types of predictable mental model errors:

1) MMDS errors are boundary mismatches. They happen when an MMDS (a) leaves out relevant features of the situation like variables, causal links, feedback loops, delays, or non–linearities or when (b) it includes irrelevant features. Such errors can compromise later processing through erroneous factors and relationships.
2) MMP errors happen when a relevant mental model of possibilities, representing a behavior pattern the system can display, is neither deployed nor processed. Such errors lead to flawed conclusions and policies with "surprising" effects.

We proceed in three major steps. First, a section about important theoretical and conceptual aspects regarding mental models discusses three relevant principles of the modern *mental model theory* and clarifies its relationship with the research themes "misperception of feedback" (MoF) and "stock-and-flow error." It then introduces the proposed conceptual framework. The third section discusses the use of mental models in some studies of MoF: the herd management problem, the fishery fleet problem, and the predator-prey management problem. It analyzes the mental models mentioned in the original publications and points out the mental model errors of both types. The discussion section then argues that research using this framework is complementary to MoF studies and allows conceiving interventions tailored to reduce mental model errors. Eventually, the concluding fifth section mentions relevant limitations, research challenges, and future steps.

Theory and concepts regarding mental models

Mental models of dynamic systems revisited

The operational definition of the *MMDS* developed by Groesser and Schaffernicht (2012) added a clear yet flexible data structure to the conceptual definition (Doyle and Ford, 1998, 1999):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

> "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 stock, flow, and intermediary variables that interact in linear and mostly non-linear, delayed ways, whose structure is analogous to the perceived structure of that system."

This defines the components of an MMDS as a data structure representing an individual's knowledge regarding the causal structure of a system (equivalent to the first layer of the framework proposed by Gary and Wood, 2016) and can, in principle, include the description of its behaviors. The following discussion of the *mental model theory of reasoning* will refer to the definition of MMDS to show how both constructs are different but compatible.

The mental model theory of reasoning: reasoning with mental models of possibilities

People make decisions between contrasting descriptions of situations, alternative descriptions of events and their causes, and imaginable courses of action and their respective consequences. The mental model theory explains the reasoning process leading to such decisions (Johnson-Laird and Ragni, 2019; Khemlani *et al.*, 2018; Khemlani and Johnson-Laird, 2019). According to it, human reasoning manifests itself as assertions like "global surface temperature is rising," "if CO_2 emissions drop to zero, then the surface temperature will decrease" (Sweeny and Sterman, 2005), or "if I work more hours per day, then my fatigue will increase." The first example is a *factual* assertion, and the second and third examples are *conditional* assertions, consisting of an *antecedent* and a *consequent*. Let p and q represent antecedent and consequent, respectively. Both antecedent and consequent refer to an entity, and the *connective* "if…then" establishes a causal relationship. For instance, if p is "CO_2 emissions drop to zero" or "I work more hours, and q stands for "surface temperature decreases" or "my fatigue increases," and "if p then q" codifies both assertions. In this article, we will focus on conditional assertions.

People deploy mental models representing the possibilities implied by such assertions (Johnson-Laird, 2012). The mental model theory has developed a series of general principles, out of which the following three are relevant for this article: (1) Representation; (2) Dual-process; and (3) Modulation (for an overview of all prinxiples, see Khemlani *et al.*, 2018). The models in the mental model theory of reasoning can be expressed in different ways. The present paper will resort to the symbols used in papers such as in López-Astorga (2021).

The principle of representation

According to the *principle of representation*, humans process the meaning of assertions in factual, counterfactual, hypothetical, or fictional discourse as MMP (Johnson-Laird and Ragni, 2019). Generically, a possibility is "a subset of finite mutually exclusive and exhaustive alternatives, where each alternative is a category stipulating what is common to an indefinite number of different realizations" (p. 5)[ii]. A *conditional* implies conjunction of possibilities (Khemlani, Hinterecker, and Johnson-Laird, 2017) that hold in the absence of information to the contrary (Khemlani *et al.*, 2018). A priori, a conditional "if p then q" (where "&" stands for "and" and "¬" for "not") comprises three possibilities:

$$\text{Possible } (p \,\&\, q) \,\&\, \text{Possible } (\neg p \,\&\, \neg q) \,\&\, \text{Possible } (\neg p \,\&\, q).$$

Humans mentally represent each of the possibilities as a mental model analogous to reality in one possible world. We consider the behavior modes of variables possibilities: the values of variables may increase, decrease, or remain constant (the slope), and change may accelerate, stay constant, or decelerate (the curvature). For instance, in each of the above examples, p and q may each refer to a variable's behavior. Depending on the situation and prior knowledge, individuals may represent the behavioral possibilities using only the slope, only the curvature (Ford, 1999), or the combination (Sterman, 2000).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

The first possibility (*p* & *q*) is intuitive. However, there are two other possibilities. First, the conditional also implies that if "I work more hours" (*p*) does not happen, then the consequence "my fatigue increases" (*q*) won't happen either. And second, it may also happen that "my fatigue increases" (*q*) even though I did not work more hours (*p* did not happen): there may be other causes. Yet, (*p* & not-*q*) is impossible because, in that case, the conditional would be false. Note that according to the rules of modal logic, the second possibility is false; however, according to the model theory, natural language is not constrained by the meta-language of formal logic (Johnson-Laird and Byrne, 2002).

The principle of dual-process

The *principle of dual-process* states that two distinct but linkable cognitive processes operate in the human mind (Khemlani *et al.*, 2018; Stanovich, 2012). *System 1* is quick, intuitive, and effortless; however, it only deploys the first and most salient mental model, limiting individuals to the possibility in which both *p* and *q* hold and omitting possibilities in which *p* or *q* are not true. In their place, we write "…" as a reminder that there are omitted possibilities (Khemlani *et al.*, 2018). *System 2* is slow, reflexive, and needs effort (Byrne and Johnson-Laird, 2020; Stanovich, 2012). The gain is the deployment of all three possibilities, which is referred to as *fully explicit models* (FEM), which also represent the possibilities where *p* or *q* are not true. This is summarized in Table 1:

Table 1: System 1 deploys and holds mental models representing what is true, while system 2 stands for FEM representing also what is not.

Conditional	If *p*, then q	
Models	Mental (system 1)	Fully explicit (system 2)
	p & *q*	*p* & *q*
	…	¬ *p* & ¬ *q*
		¬ *p* & *q*

When people think intuitively about a situation, two of the possibilities are neither represented nor processed. But clearly, these possibilities can be true in the real situation. For instance, my fatigue can increase despite not having worked more hours. Consequently, people will make systematic mistakes when reasoning intuitively. Neglecting models of possibilities yield erroneous conclusions that are consistent with the premises but do not follow them (Khemlani *et al.*, 2018).

The principle of modulation

The *principle of modulation* states that prior knowledge and new facts can change the set of FEM of possibilities deployed by system 2 (Khemlani *et al.*, 2017; Quelhas, Johnson-Laird, and Juhos, 2010). This can happen in several ways: (a) prior knowledge can add concepts and relationships which were not explicit in the original assertion, and (b) new evidence can be perceived (Johnson-Laird and Ragni, 2019; Khemlani *et al.*, 2018; Ragni and Johnson-Laird, 2020a). If a new fact rebuts a conclusion, reasoners withdraw the conclusion; they also amend premises to restore consistency and try to explain the origins of the inconsistency (Khemlani *et al.*, 2018). Either way can block the deployment of a fully explicit model or lead to the deployment of a model that is not part of the default possibilities mentioned above (Table 1).

For instance, consider the conditional "if births decrease, then the population may decrease." Due to the meaning of "may," modulation adds a fourth possibility:

Possible (*p* & *q*) & Possible (¬*p* & ¬*q*) & Possible (¬*p* & *q*) & Possible (*p* & ¬*q*).

The first three possibilities are in line with the previous subsection. Possible (*p* & *q*) holds whenever births decrease sufficiently to yield a negative net flow. When births remain constant and greater than deaths, possible (¬*p* & ¬*q*) is true.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

Possible (¬*p* & *q*) is the case when births stay smaller than deaths. The fourth possibility follows from the fact that "may" also means "may not."

Consider next the conditional "if population grows, then this is exponential growth":

Possible (*p* & *q*) & Possible (¬*p* & ¬*q*) & Possible (¬*p* & *q*).

Exponential growth is one shape of growth, so (*p* & *q*) holds, and so does (¬*p* & ¬*q*). However, modulation blocks possibility (¬*p* & *q*) because it makes no sense now. Also, it deploys possibility (*p* & ¬*q*) because the population may be growing asymptotically.

When system 2 is active, it compares the deployed models of possibilities with the currently known facts. The mental model theory proposes that a conditional is evaluated as necessarily true if all models hold, as possibly true if the conclusions match the premises in some models, and as impossible if none of the models hold (Khemlani *et al.*, 2018).

Causal diagrams can represent conditionals and thus make the underlying MMDS visible, as illustrated in Figure 1. Continuing with the conditional "if I work more hours, then my fatigue will increase," the black variables and causal link represent what would be the salient (system 1) possibility and its non-salient negation. The positive polarity is consistent with these possibilities. The gray part is far less salient from the conditional, but the possibilities (¬*p* & *q*) and (*p* & ¬*q*) hold, and system 1 will deploy them.

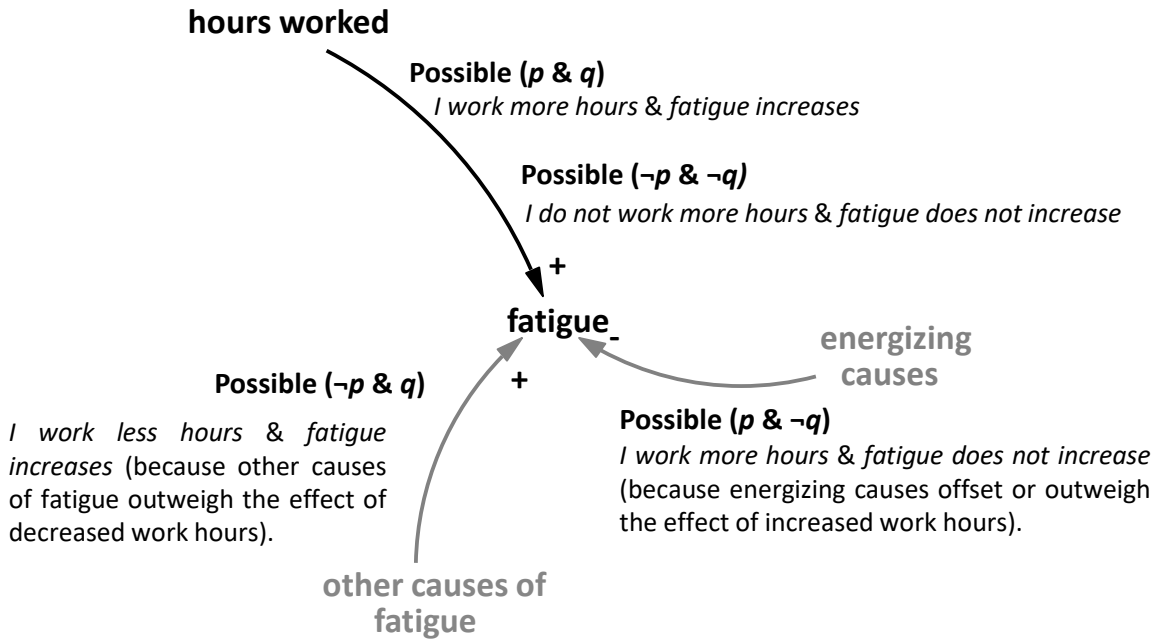

Figure 1: The relationship between mental models and causal diagrams

If the perceived MMDS does not contain the additional variables, the possibility (¬ *p* and *q*) would be omitted for its lack of salience. Note that what is not in the MMDS is not available for reasoning. If there is only one causal link from "work hours" to "fatigue," this states that this causal relationship operates always, and there are no other causal influences from other variables (Pearl, 2009; Pearl and Mackenzie, 2018). Recognizing less salient factors as relevant takes more mental effort. According to the *model theory*, they will be overlooked if there is no reason to make such additional efforts.

Based on these considerations, we propose that MMDS and MMP can be combined to study decision policies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

Conceptual framework

The conceptual framework proposed here comprises four layers. The *first layer* is the situation: an unstructured cloud of features, and each one may or may not be relevant to achieving a goal. In *layer 2*, the MMDS is an individual's representation of the situation's structure inside a conceptual boundary. Here a first error type appears: the model boundary mismatch comes in two variants: (a) relevant features may have been left out, and (b) irrelevant or even illusory features may have been included. Boundary mismatches can be detected and corrected later on (Sterman, 2002), but while they exist, what is left out of the MMDS cannot be stated in assertions, and illusory MMDS elements will lead to illusory assertions. Therefore, this error type compromises the ensuing process of deploying mental models. We refer to MMDS boundary mismatches as *MMDS errors*.

The third layer is conditional assertions and the possibilities deployed by system 1 or 2. The MMDS constrains the assertions and the MMP. We propose a basic vocabulary of behavioral features to express default conditional behavior patterns. The word labels are: greater than, smaller than, equal to, increases, decreases, remains constant, accelerates, decelerates, quicker, slower, reaches a maximum, reaches a minimum, and has an inflection. Combined, these labels can refer to elementary behavior modes like accelerating and decelerating increase, but also diverse other descriptions like increases quicker than, increases slower than, or similar. One can combine the variable names drawn from the MMDS with the descriptions to form conditionals:

If <*variable A*> <*labels*> then <*variable B*> <*label*>

The second type of error happens here: MMP errors. As the assertions are constrained by the MMDS, missing structural or illusory elements entail missing or illusory conditionals and MMPs. Furthermore, system 1 risks overlooking true possibilities; only system 2 would deploy, and modulation may fail to block flawed MMPs inherited from MMDS errors. Arguably, MMP errors lead to flawed decision policies.

Policies are the *fourth layer*[iii]. We represent them as *deontic* conditionals prescribing what to do in response to certain conditions: "if <condition> then <do this or that>." The follow immediately from the conclusions reached in the MMP layer. Therefore, leaving possible circumstances unconsidered opens the door for decisions that provoke undesired outcomes. Possibilities may have remained unconsidered because of either type of error. The four layers are illustrated in Figure 2, which also shows that each layer depends on the previous one.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

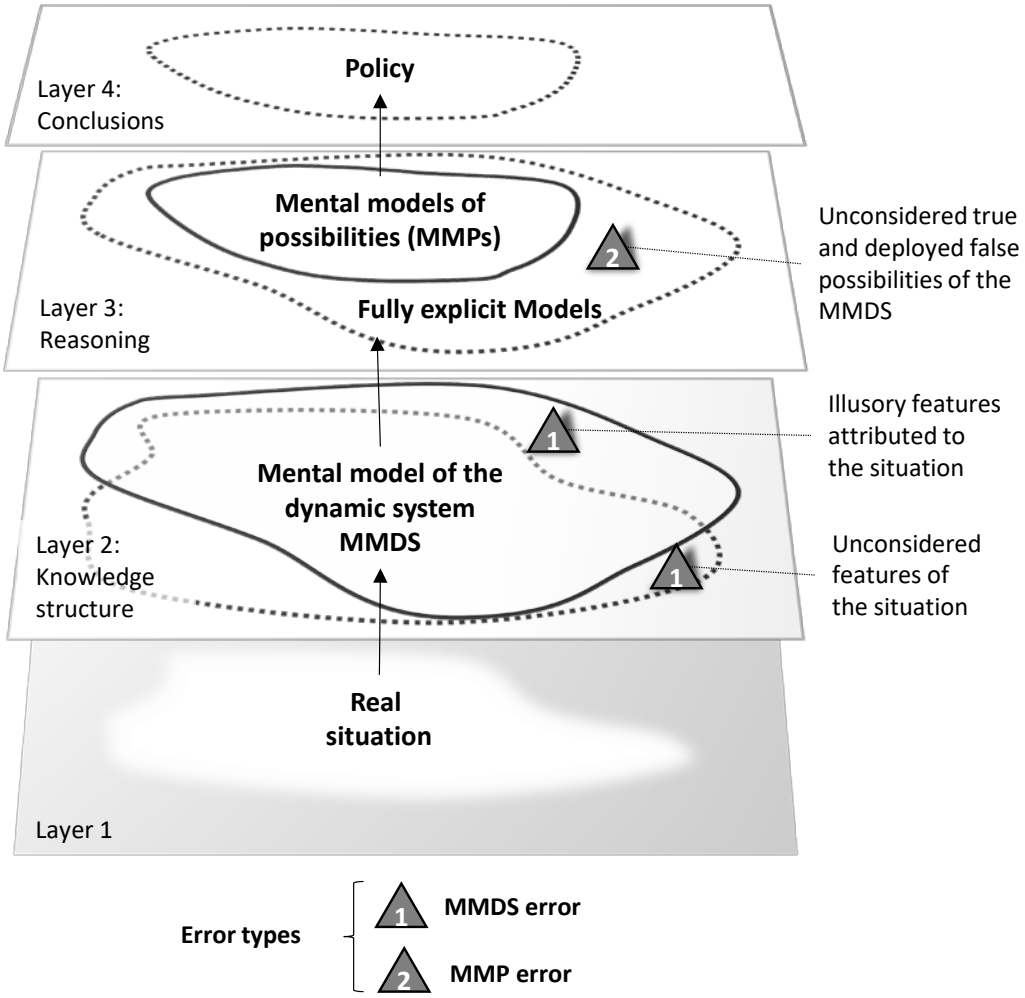

Figure 2: The error types, the relationship between situation, the mental model of the dynamic system, mental models of possibilities and policies

If people commit such MMDS and MMP errors, the ability to identify them is a first step toward correcting or even avoiding them. As people use their reasoning abilities to figure out decision policies, the ability to identify MMDS errors and MMP errors is helpful. Indeed, we found such errors in reports from dynamic decision tasks in the system dynamics literature, as the following section shows.

Using the framework in some studies of dynamic decision tasks

Three similar decision tasks based on dynamic systems

Several studies have reported mental models or vignettes from utterings articulated by participants in laboratory experiments. In the experiments discussed here, participants had to maximize the profits from a fishing fleet catching fish (Moxnes, 2000), or maximize sustainable meat production from a reindeer herd grazing lichen (Moxnes, 2000, 2004), or maintain two animal species (foxes as predators and rabbits as prey) stable by controlling the predator population (Jensen and Brehmer, 2003). Achieving growth or stability when the behavior of variables responds not only to one's decisions but also to other factors is typical for dynamic decision-making (Gonzalez, Fakhari, and Busemeyer, 2017).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

Despite the differences between the decision tasks, each of these experiments involved two interdependent resource stocks[iv]:

- Fishery: ships and fish (see fig. 1 on page 332 in Moxnes, 2000);
- Animal production: reindeer and lichen (see fig. 4 on page 336 in Moxnes, 2000; the same case was discussed in the 2004 article);
- Populations: foxes and rabbits (see fig. 1 on p. 124 in Jensen and Brehmer, 2003).

These decision problems have a similar causal structure beneath the particular names given to the respective variables (Figure 3).

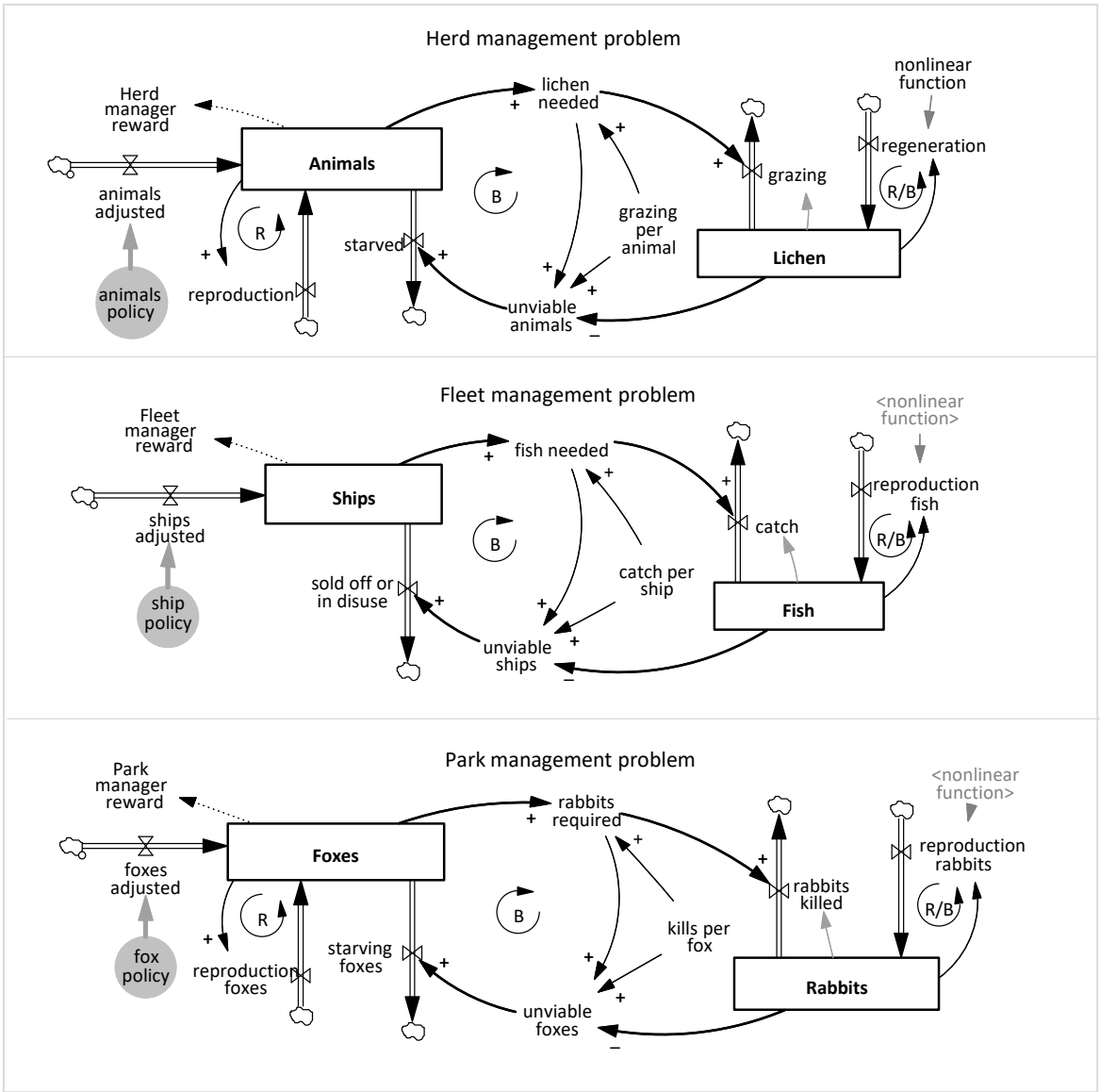

Figure 3: The core structure of all three dynamic decision problems comprises two interdependent stock resources interconnected by a balancing loop. Each of these stocks has its own regenerative structure (except ships). The participants' reward depends on the resource they can directly influence, and they have to figure out a policy for influencing it conveniently.

 doi:10.20944/preprints202107.0004.v3

In each case, participants directly control the first stock, influencing the second stock: ships catch fish, reindeer graze lichen, and predators eat prey. Therefore, sustainable solutions always depend on balancing both interdependent resources. But the situation is complex: the participants can influence the first resource (ships, reindeer, or foxes), but reindeer and fox populations also influence themselves through reproduction. They influence lichen and rabbits through consumption, and both lichen and rabbit populations influence themselves through regeneration, and lichen and rabbits influence reindeer and foxes through starvation.

In each problem, participants need to account for the regeneration structure of both interdependent stocks and the interdependence between them (all animals reproduce and all plants regenerate, only ships do not; also, the reproduction dynamics of rabbits are simplified–their interdependence with food plants is not included). Participants receive briefings containing all relevant information about the variables and causal relationships. Therefore, a system dynamicist could construct an MMDS from the briefings, and "a near-to-optimal management strategy follows easily from the perfect model" (Moxnes, 2004). However, naïve individuals do not know about the difference between stock-and-flow variables or look out for feedback loops [v].

The herd management problem

Discussing the reindeer experiment, (Moxnes, 2004) proposed the assertion "the more animals, the less lichen, and vice versa" as a static mental model, observing that this statement is "strongly supported" by historical data given to participants during the briefing: "The previous owner has increased steadily the number of reindeer from 1150 to 1900. Consequently, the lichen thickness (mm) has dropped from 50 to 24.4 mm," accompanied by graphs and a table with the historical data.

In terms of MMDS, the assertion contains two variables and possibly one causal link. Two interpretations are consistent with the assertion:

a) *There are more animals*, and *there is less lichen* & *there are fewer animals*, and *there is more lichen*

b) If *animals increase*, then *lichen decreases*

The first assertion describes how animals and lichen have changed without reference to a causal relationship-just like the uttering concerning the static mental model. However, the briefing mentions that the decrease in lichen thickness is a consequence of the increasing herd size. Therefore, the second assertion could indeed be how naïve individuals assert what is going on.

The briefing mentions the "consequence" in the text, and the graphs and the table show the variables' behaviors, which verify the second assertion because the evidence is consistent with the conditional. Participants will therefore engage system 1 and only consider the possibility that "*animals increase* and *lichen decreases*" (p & q), which is true when the grazing is greater than lichen growth. However, grazing is not always greater than lichen growth, which becomes apparent when processing the remaining FEM deployed by system 2:

- "*Animals* do not *increase* & *lichen* does not *decrease*" ($\neg p$ & $\neg q$) as well as "*animals increase* & *lichen* does not *decrease*" (p & $\neg q$) hold when the increased number of animals graze less than the lichen's regeneration, which is entirely possible (note that in this case, lichen may remain constant or increase).
- It is also possible that "*animals* do not *increase* & *lichen decreases*" ($\neg p$ & q) if grazing increases less than lichen growth.

 doi:10.20944/preprints202107.0004.v3

There are four FEM, and the third one (p & $\neg q$) contradicts the first one; therefore, system 2 concludes that the second assertion is possibly true, but not necessarily. Of course, naïve individuals may lack the prior systems knowledge needed for the modulation, which deploys the possibility that "*animals increase* & *lichen* does not *decrease*"—accordingly, they may overlook that the assertion does not always hold and thus derive a flawed policy like "when lichen decreases, I must decrease the herd size." Considering that lichen thickness is greater than half of the maximum thickness, lichen decrease toward the optimal thickness is desirable: this policy would not allow the individual to achieve a good performance.

The briefing information requires some relevant prior systems knowledge. One stock cannot directly change another stock, so it is impossible for animals to change lichen. There must be one or several flows involved. As animals can increase and decrease, there must be at least one inflow and one outflow, and the same must be the case with lichen. The inflow to animals cannot be the outflow from lichen because animals are a different substance than lichen. So, grazing must be an outflow from lichen, and it is proportional to the number of animals. Thus, the "simplified dynamic mental model" suggested on page 151 does not commit the same mistake: "since lichen has decreased historically, the historical grazing must have exceeded the growth of lichen." This assertion mentions three variables: lichen, grazing, and lichen growth; "animals" remains implicit, but "grazing" means the total grazing by all animals and therefore implies a causal link from animals in the MMDS. Additionally, gazing drains lichen, and lichen growth adds to lichen. So, a deliberate effort of thinking (system 2) is:

If *there are more animals*, then *there is more grazing*.

If *there is more grazing*, then *there will be less lichen than what would have been the case otherwise*.

If *there is more lichen growth*, *then there will be more lichen than what would have been the case otherwise*.

The last two conditionals can be combined:

If *grazing is greater than lichen growth*, then *lichen will decrease*.

The briefing defines the per capita grazing as a constant, so the latter conditional turns into:

If *animals * per capita grazing is greater than lichen growth*, then *lichen will decrease*.

This is consistent with the assertion on p. 151. Two FEM are, therefore:

Possible (*animals * per capita grazing* is equal to *lichen growth* & *lichen is constant*)

Possible (*animals * per capita grazing* is smaller than *lichen growth* & *lichen increases*)

An individual who cannot determine the size of lichen growth cannot derive the critical number of animals and can only resort to a hill-climbing approach based on outcome feedback season after season (Moxnes, 2004).

The fishery fleet management problem

In the fishery case reported by Moxnes (1998), the causal structure is analogous to the reindeer case, despite the superficial differences. Hence, participants' reasoning seemed analogous, too: "the subjects order new vessels as long as they perceive those profits […] are improving." A simple policy consistent with this data is then "if *profits change in one direction*, then *I change the fleet in the same direction*," which is again the hill-climbing logic.

The assertion mentions two resource stocks–ships and profits-but obviate the fish stock. Not accounting for this third stock is like taking a certain catch per ship as granted. The assertion behind the policy is:

If *I have more ships*, then *I will have more profits*.

System 1 then deploys the mental model of the possibility:

Possible (*I have more ships* & *I have more profits*).

In such a case, the individual prevents several FEM of possibilities:

Possible (*I* do not *have more ships*, and *& I* do not *have more profits*).

This contradicts the following model:

 doi:10.20944/preprints202107.0004.v3

Possible (*I* do not *have more ships* & *I have more profits*).

When one of these possibilities holds, the other does not. However, the current fish stock determines which one is true: if the fish population is recovering from overfishing in the past, then the catch per ship may increase and yield more profits while holding the fleet constant. Finally, if the fleet is large enough to make the total catch exceed the fish reproduction, the following model will hold:

Possible (*I have more ships* & *I* do not *have more profits*).

System 1 will not deploy these models, but even if an individual engages system 2, it takes some prior knowledge to think through the FEM. Individuals who overlook the fish stock are unlikely to consider its net change resulting from net reproduction minus catch. The MMDS error to leave the third stock and its flows unrecognized leads to an MMP error. The consequence is believing that the conditional is *necessarily* true even though it is only *possibly* true.

The predator-prey management problem

Similar results came from the predator-prey experiment. Jensen and Brehmer (2003) describe a mathematical approach based on the briefing information, starting with the realization that in equilibrium, the populations of foxes and rabbits remain constant. The mental model reasoning behind such an approach is based on a biconditional like:

If *foxes are in equilibrium*, then *rabbits are in equilibrium*.

Yet, the set of FEM in the case of biconditionals is slightly different from the one for conditional.

Possible (*foxes are in equilibrium* & *rabbits are in equilibrium*)

Possible (*foxes are not in equilibrium* & *rabbits are not in equilibrium*)

Individuals need to realize that:

If *foxes born are equal to foxes starved*, then *foxes are in equilibrium*.

This is also a biconditional: either both parts are true, or none is. It can be used to determine the fox population for which it is true, and then one can trust that the rabbits will move into equilibrium, too. However, only one out of 15 individuals came close to taking this approach, and the authors joined Moxnes, referring to the majority as following a feedback-driven policy. They also described an "ideal way to solve the task using a feedback approach" (p. 123-124). People would need to understand behavioral relationships leading to the following conditionals:

If *the rabbit population increases*, then *the fox population is too small*.

If *the rabbit population decreases*, then *the fox population is too large*.

If *the fox population increases*, then *the rabbit population is too small*.

If *the fox population decreases*, then *the rabbit population is too large*.

As before, the intuitive mental models (system 1) will only represent the obvious possibility without further deliberation. The FEM of the other possibilities will only seem plausible to individuals with sufficient systems knowledge to remember why the populations change during the year.

Mental model errors in these examples

In each of these decision problems, the MMDS missed several relevant variables and causal relationships and failed to make a difference between stocks and flows; no hint at recognized feedback loops was in the published data: these are all MMDS errors. The assertions derived from the flawed MMDS are false because they contradict the possible behaviors of the respective systems. These MMP errors are a consequence of the MMDS errors. As the only recognized variables were stocks, the term "static mental model" is convenient, and so is the notion "open-loop model" (Sterman, 1989a, b) for leaving the loops out. The assertions which made sense (given these MMDSs) lead to mental models of possibilities (system 1) which miss several fundamental behavioral possibilities, in other words: MMP errors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 August 2022 doi:10.20944/preprints202107.0004.v3

The FEM cover these possibilities, but system 2 will only kick in when there is a considerable cognitive disease (in the terms of Kahneman, 2011). If the briefing information and the intuitive MMP do not trigger such unease, only "surprising" decision outcomes could possibly do so. Yet, a series of biases may preclude this possibility. The discussion section will return to this topic.

A thought experiment in the herd management task

Still, the prior knowledge of systems thinkers makes them likely to avoid most MMDS errors and the ensuing MMP errors. Trained systems thinkers will extract more variables and causal relationships from a case description than a naïve individual because they know the basic concepts and must look for the stock variables and then the flows and then the causes of the flows to devise an endogenous explanation.

For the reindeer experiment (Moxnes, 2000, 2004) it has been argued that "a near-to-optimal management strategy follows quite easily from the perfect model" (Moxnes, 2004). Therefore, participants in such games face two challenges: (a) constructing the perfect model (MMDS) of the decision situation and (b) deriving a policy from it.

The following sequence of assertions describes a way how a systems thinker can devise a policy based on a dynamic MMDS. Suppose that prior knowledge can be expressed in the following axiom-like statements:

A system's state at one point in time is defined by the level of its stocks.

A system's behavior realizes through the impact of flows on the stocks.

Only flows change stocks.

Flows depend on the level of one or several stocks.

Inflows add to stocks.

Outflows drain from stocks.

Then the briefing information is processed into an MMDS corresponding to the "herd management problem" segment of Figure 3 (above). The following assertions are then true a priori:

Grazing drains lichen

Lichen growth adds to lichen

Reproduction adds to animals

Starving drains from animals

Animals link to grazing per animal with positive polarity

*Lichen links negative to starving when animals * grazing per animal > lichen*

Some further steps lead to the desired level for lichen. The "quite easily" in Moxnes' discussion suggests that they are produced by system 1:

If *lichen growth is maximized,* then *the highest number of animals can be sustained*.

If *I have more animals*, then *I will have more profits.*

If *lichen growth is maximized,* then *my profits will be maximized*.

If *the level of lichen is 30 mm*, then *lichen growth is maximized*.

If *the level of lichen is 30 mm*, then *my profits will be maximized*.

Set *the desired level of lichen* to 30 mm.

The following assertions describe the ideal situation. They are true (the intuitive MMP holds). System 1 is sufficient because the additional FEM do not hold. Therefore, someone who has concluded to set the desired level of lichen to 30 mm can rely on intuitive reasoning:

If *lichen = desired level of lichen,* then *maximum lichen growth*.

If *lichen approaches the desired level as quickly as possible* then *profits will be maximized*.

 doi:10.20944/preprints202107.0004.v3

Once the target for lichen is set, the following policy adjusts the herd size independently of the initial endowment of animals and lichen and without hill-climbing:

If *lichen* < *desired level of lichen*, then *I should decrease the number of animals*.
If *lichen* > *desired level of lichen*, then *I should increase the number of animals.*
If *I should change the number of animals*, then *I should change them the quickest possible*:
If *I need to* reduce *the number of animals*, then set the *number of animals to 0*.
If *I need to increase the number of animals*, then:
Set *lichen surplus* to *lichen–desired level of lichen*.
Adjust *animals* by *lichen surplus* / *grazing per animal*.

The resulting conditionals refer to what ought to be done (deontic) and represent a valid policy:

If *lichen* < *desired level of lichen*, then I should set *the number of animals* to 0.
If *lichen* > *desired level of lichen*, then I should (set *lichen surplus* to *lichen—desired level of lichen* & adjust *animals* by *lichen surplus* / *grazing per animal*)

This result suggests that the mental processing of conditionals and their possibilities yields an explanation of how naïve individuals arrive at flawed policies and how individuals with systems skills derive well-functioning policies.

Discussion

The use of mental models of dynamic systems and possibilities

The previous sections show that reasoning with a MMDS can be represented as conditionals which imply certain possibilities. MMDS errors lead to missing or illusory models of possibilities (MMP errors). Two other sources of MMP errors are (a) the neglect of true possibilities by system 1 and (b) not blocking false possibilities by modulation because of missing prior knowledge. Faulty sets of possibilities led to feeble or even flawed decision policies that failed to achieve high performance or were unsustainable.

Focusing on the link from MMDS to decision policies, we argue that the proposed framework enables mental model research in dynamic decision-making to complement research on the link between policies and decision behaviors prominently carried out by the "MoF studies. One important benefit comes from directly eliciting MMDS as articulated by their holders. This allows us to detect systematic mental model errors, where MoF studies typically depend on the researcher to ascribe such errors. Such detection is a step toward developing corrective interventions. For instance, including some contradictory evidence in briefing information could trigger system 2 and facilitate processing non-salient possibilities and modulation.

Yet, even though our framework and our results suggest that the *theory of mental models* can be consistently and fruitfully used in combination with the concepts and methods toolset so far developed in system dynamics research on dynamical decision-making, the results are theoretical. Now, empirical studies with human participants are needed to solidify this result. Several research questions need to be addressed:

- *Learning*: given the iterative nature of dynamic decision tasks, what in the MMDS and MMP changes over the iterations of a dynamic decision experiment? How do new evidence and prior knowledge interact, and under which conditions do individuals reconsider their MMDS and assertions?

 doi:10.20944/preprints202107.0004.v3

- *Perception*: do certain cues in the information fed to participants during the iterations reduce MMDS or MMP errors? For instance, if naïve individuals misperceive flow variables, can error feedback trigger system 2, and make them reflect on what causes stocks to change? Also, does data visualization, possibly combined with sound and even with haptic elements, lead to systematic variations in the errors?
- *Cognitive load, working memory, and cognitive dissonance*: the briefing usually introduces simplifying assumptions needed to avoid unnecessary complexity. Participants must keep them in their working memory during the experiment. As the brain has only limited resources, a higher demand on working memory should be expected to diminish the attention given to reasoning (Brunyé and Taylor, 2008). If this can be confirmed, does making such assumptions salient just–in–time during the iterations decrease this phenomenon? When participants have prior knowledge in similar situations, retaining such assumptions in their working memory will become even more demanding because they contradict their experience. This may lead to MMDS errors that are induced by the decision task. This would suggest that some experimental situations trigger artificial mental model errors and improve the experimental settings to avoid such problems.
- *Dynamic complexity*: the decision situations in the experiments discussed here are comparatively simple. Other situations, like "fish banks" or versions of the "market growth and underinvestment" model, include more feedback loops and more delayed relationships. Thus, results observed in studies dealing with the previous questions can be examined in increasingly complex decision tasks.
- *Transferability of insights*: decision tasks may vary in their superficial features, but they may also vary in the complexity of the underlying causal structures. To the extent where participation in experimental games makes individuals learn something, the question arises if they can transfer this new knowledge to another task. Would some kinds of MMDS or MMP errors decrease?
- *The role of prior knowledge and modulation*: the outline of the mental model theory has mentioned that the meaning of words used in the assertions can activate additional semantic knowledge. But in cases like "if births decrease, then the population may decrease," pragmatic knowledge may also have consequences. For instance, knowledge of causal link polarity (Richardson, 1986, 1997) or of population dynamics (population also changes due to deaths) allows concluding that the possibility of (*births decrease* & ¬ *population decreases*), which happens when *births* decrease but remain greater than *deaths*. A trained systems thinker will have no problem recognizing that the *population* is impacted by more than one flow variable. A sharp logical thinker may suspect that as *births* can cause all

 doi:10.20944/preprints202107.0004.v3

kinds of behavior in a *population*, they may also recognize that there must be another variable influencing *population* together with *births*. The question is, does systems knowledge make people more likely to deploy and process this possibility?

Such research will provide insights into the cognitive reasons behind phenomena like the MoF and contribute to the system dynamics literature. Also, cognitive scientists gain access to a type of integrative decisions and reasoning that concentrates on dynamic behaviors rather than assertions concerning certain states or certain events and one-off decisions.

Conclusions

This article introduces a way to analyze the structure of and the reasoning with mental models of dynamic decision situations: (1) MMDS–well known in the system dynamics field but seldom applied–contain the mental representation of the decision situation, and (2) MMP frame the reasoning of decision-makers according to the "theory of mental models" developed by Johnson-Laird and collaborators. We show how the elements of the MMDS are used as conditional assertions, implying certain sets of possible behaviors expressed as MMP. Our conceptual framework, therefore, consists of the layers of situation, MMDS, MMP, and decision policy. It fits into a more general framework consisting of perceived causal structure (here MMDS), strategies, decision rules, and results (Gary and Wood, 2016) at the link from perceived causal structure to strategies.

We show that decision-makers can theoretically commit different types of mental errors. MMDS errors are model boundary mismatches and happen when an MMDS lacks relevant elements or contains irrelevant elements. MMP errors happen when individuals fail to consider one or several true possibilities and when false possibilities are not blocked by modulation. Our analysis of three well known dynamic decision studies confirms that MMDS errors and MMP errors happen.

Therefore, we propose that combining MMDS with the *theory of mental models* is fruitful. It provides the possibility to represent how decision-makers reason with their MMDSs, and to pinpoint the errors committed due to, for instance, the MoF. These errors make a flawed policy seem correct to a decision-maker. Knowledge of these errors and the cognitive theory of mental models makes it possible to design decision environments that minimize such errors. We hope this perspective may motivate researchers to apply and advance this framework critically.

The obvious limitation of this contribution is its theoretical character. To our knowledge, no empirical studies have been carried out to test our claims in new experiments. To orient such studies, several orienting questions have been discussed.

It is now time to include real individuals as decision-makers. Some directions for empirical research have been delineated, and we hope this article may encourage empirical studies in this area.

References

Baghaei Lakeh A, N Ghaffarzadegan. 2015. Does analytical thinking improve understanding of accumulation? *System Dynamics Review* **31**(1-2): 46-65.

Baratgin J, I Douven, JSBT Evans, M Oaksford, D Over, G Politzer, V Thompson. 2015. The new paradigm and mental models. *Trends in Cognitive Science* **19**(10): 547-548.

Braine MDS, DP O´Brein. 1991. A theory of If: A lexical entry, reasoning program and pragmatic principles. *Psychological Review* **98**: 182-203.

Braine MDS, DPE O'Brien. 1998. *Mental Logic*. Lawrence Erlbaum Associates, Inc., Publishers., Mahwah, NJ.

Brunyé TT, H Taylor. 2008. Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language* **58**: 701-729.

Byrne RMJ, PN Johnson-Laird. 2020. If and or: Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **46**(4): 760-780.

Chater N, M Oaksford. 1999. The probability heuristics model of syllogistic reasoning. *Cognitive Psychology* **38**: 191-258.

Doyle JD, DN Ford. 1998. Mental models concepts for system dynamics research. *System Dynamics Review* **14**(1): 3-29.

Doyle JD, DN Ford. 1999. Mental models concepts revisited: some clarifications and a reply to Lane. *System Dynamics Review* **15**(4): 411-415.

Ford DN. 1999. A behavioral approach to feedback loop dominance analysis. *System Dynamics Review* **15**(1): 3-36.

Forrester JW. 1971. Counterintuitive behavior of social systems.pdf. *Technology Review*(January 1971).

Forrester JW. 1987. Lessons from system dynamics modeling. *System Dynamics Review* **3**(2): 136-149.

Gary MS, RE Wood. 2016. Unpacking mental models through laboratory experiments. *System Dynamics Review* **32**(2): 101-129.

Gonzalez C, P Fakhari, J Busemeyer. 2017. Dynamic Decision Making: Learning Processes and New Research Directions. *Human Factors* **59**(5): 713-721.

Groesser SN, MF Schaffernicht. 2012. Mental models of dynamic systems: taking stock and looking ahead. *System Dynamics Review* **28**(1): 22.

Jensen E, B Brehmer. 2003. Understanding and control of a simple dynamic system. *System Dynamics Review* **19**(2): 119-137.

Johnson-Laird PN. 1983. *Mental Models Towards a Cognitive Science of Language*. Cambridge University Press, Cambridge, UK.

Johnson-Laird PN. 2004. The history of mental models. In Manktelow K.I., M.C. Chung (eds.), *Psychology of reasoning: theoretical and historical perspectives*. Psychology Press, London.

Johnson-Laird PN. 2010. Mental models and human reasoning. *Proc Natl Acad Sci U S A* **107**(43): 18243-50.

Johnson-Laird PN. 2012. Inference with mental models. In Holyoak K.J., R.G. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press, New York, pp. 134-145.

Johnson-Laird PN, RMJ Byrne. 2002. Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* **109**(4): 646-678.

Johnson-Laird PN, M Ragni. 2019. Possibilities as the foundation of reasoning. *Cognition* **193**.

Kahneman D. 2011. *Thinking, Fast and Slow*. Farrar, Strauss and Giroux, New York.

Khemlani S, RMJ Byrne, PN Johnson-Laird. 2018. Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science* **42**(6): 1887-1924.

Khemlani S, T Hinterecker, PN Johnson-Laird. 2017. The provenance of modal inference. *Proceedings of the Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pp. 259-264. Austin, TX. Cognitive Science Society.

Khemlani S, PN Johnson-Laird. 2019. Why machines don't (yet) reason like people. *Künstliche Intelligenz* **33**: 219-228.

Lane DC. 1999. Friendly amendment: A commentary on Doyle and Ford's proposed re-definition of `mental model'. *System Dynamics Review* **15**(2): 185-194.

López-Astorga M. 2021. Reminiscence theory and immortality of the soul: Some mental representations in Plato's Phaedo. *Annals of the University of Craiova Philosophy Series* **48**(2): 5-14.

López-Astorga M, M Ragni, PN Johnson-Laird. 2022. The probability of conditionals: a review. *Psychonomic Bulletin & Review* **29**(1): 1-20.

Moxnes E. 1998. Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics. *Management Science* **44**(9): 1234-1248.

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.

O'Brien DP. 2014. Conditionals and disjunctions in mental-logic theory: A response to Liu and Chou (2012) and to López-Astorga (2013). *Universum* **29**(2): 221.235.

Pearl J. 2009. *Causality - Models, Reasoning, and Inference*. Cambridge Univiersity Press, New York.

Pearl J, D Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York.

Qi L, C Gonzalez. 2015. Mathematical knowledge is related to understanding stocks and flows: results from two nations. *System Dynamics Review* **31**(3): 97-114.

Quelhas AC, PN Johnson-Laird, C Juhos. 2010. The modulation of conditional assertions and its effects on reasoning. *Quarterly Journal of Experimental Psychology* **63**: 1716-1739.

Ragni M, P Johnson-Laird. 2020a. Reasoning about epistemic possibilities. *Acta Psychologica* **208**.

Ragni M, PN Johnson-Laird. 2020b. Explanation or Modeling: a Reply to Kellen and Klauer. *Computational Brain & Behavior* **3**(3): 354-361.

Ragni M, I Kola, PN Johnson-Laird. 2018. On selecting evidence to test hypotheses: A theory of selection tasks. *Psychol Bull* **144**(8): 779-796.

Richardson GP. 1986. Problems with causal‐loop diagrams. *System Dynamics Review* **2**(2): 158-170.

Richardson GP. 1997. Problems in causal loop diagrams revisited. *System Dynamics Review* **13**(3): 247–252.

Rips LJ. 1994. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press, Boston.

Schaffernicht MF. 2021. Three Generic Policies for Sustained Market Growth Based on Two Interdependent Organizational Resources—A Simulation Study and Implications. *Systems* **9**(2): 43.

Stanovich K. 2012. On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In Holyoak K., R. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning* Oxford University Press, New York, NY, pp. 343-365.

Sterman J. 1989a. Misperceptions of Feedback in Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* **43**(3): 301-335.

Sterman J. 1989b. Modeling managerial behavior - misperceptions of feedback in a dynamic decision-making experiment. *Management Science* **35**: 18.

Sterman J. 2000. *Business dynamics - Systems Thinking and Modelling for a Complex World*. McGraw-Hill.

Sterman J. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501-531.

Sterman J. 2010. Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review* **26**(4): 316-334.

Sweeny LB, J Sterman. 2005. Cloudy skies: assessing public understanding of global warming. *System Dynamics Review* **18**(2): 207-240.

[i] Other theories take a *probabilistic* approach (Baratgin *et al.*, 2015; Chater and Oaksford, 1999) or propose that human reasoning follows *formal inference rules* (Braine and O'Brein, 1991; Braine and O'Brien, 1998; O'Brien, 2014; Rips, 1994). Two decades ago, the formal rules approach seemed to be a strong competitor for the model theory (Doyle and Ford, 1998), but recent comparative research suggests that the model theory overcomes several paradoxes which the remaining two types of theories leave unanswered (López-Astorga, Ragni, and Johnson-Laird, 2022; Ragni and Johnson-Laird, 2020b; Ragni, Kola, and Johnson-Laird, 2018). Since both research streams work on mental models, we argue for a fruitful combined use.

[ii] There are three types of possibility: alethic, deontic, and epistemic. Alethic possibilities are "anything not self-contradictory" (Khemlani et al., 2018), and typical verbs are "causes", "prevents", and "allows" (Johnson-Laird and Ragni, 2019). They are relevant because prior knowledge can add causal relationships to the situation perceived by an individual (as discussed below for the principle of modulation). Deontic possibilities refer to anything permissible (Khemlani et al., 2018) and arise from speech acts indicating that something is permissible or obligatory or the opposite, using terms like "obliges", "prohibits", "permits" (Johnson-Laird and Ragni, 2019). Epistemic possibilities deal with anything consistent with knowledge (Khemlani et al., 2018). These possibilities are particularly important because default possibilities are defeasible in the face of evidence to the contrary.

[iii] We use the term "policy" for what Gary and Wood (2016) called "strategy" because the system dynamics literature has traditionally referred to the qualitative part of rules as policies. The precise meaning of "decision policy", "strategy" and "decision rule" varies across fields, and one could avoid misunderstandings by referring to generic and specific policies (for an example, see Schaffernicht, 2021). However, a thorough discussion of terminology is beyond the scope of this article.

[iv] In the fishery and the reindeer problems, there is a third stock accumulating profits and production, respectively; they represent a reward which participants try to maximize. However, there is no physical interdependence between the reward and the resource which participants influence to maximize it.

[v] Evidence from several studies suggests that individuals with a strong mathematical background (Qi and Gonzalez, 2015), a high academic level (Sterman, 2010) or who are primed towards analytical thinking (Baghaei Lakeh and Ghaffarzadegan, 2015) tend to do better in stock-and-flow thinking. But Moxnes reported results from individuals who

 doi:10.20944/preprints202107.0004.v3

were familiar with the application domain but not trained as system thinkers; Jensen and Brehmer worked with undergraduate students who were neither knowledgeable in the application domain nor systems modeling.