Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

HOW MODELING METHODS FOR FUZZY COGNITIVE MAPPING CAN BENEFIT FROM PSYCHOLOGY RESEARCH

Samvel Mkhitaryan

Department of Health Promotion
CAPHRI, Maastricht University
P.O. Box 616
6200 MD Maastricht, The Netherlands

Philippe J. Giabbanelli

Department of Computer Science & Software Engineering
Miami University
501 E High St
Oxford, OH 45056, USA

ABSTRACT

Fuzzy Cognitive Maps (FCMs) are aggregate-level simulation models that represent concepts as nodes, capturing relationships via weighted edges, and apply an inference mechanism to update the nodes' values until a desired effect is achieved. FCMs are increasingly combined with other techniques. Agent-Based Models (ABMs) use FCMs to represent the 'mind' of each agent, which governs (and is influenced by) interactions with other agents or the environment. A question continues to elude simulationists: what should be the building blocks for such simulations? FCMs can now be optimized using machine learning and quantitative data, which means that an agent's mind can be automatically modified to closely align with an evidence base. However, there are multiple ways in which an FCM can be transformed: which transformations correctly capture how individuals change their mind? In this paper, we explore these questions using psychology research, thus leveraging knowledge on human behaviors to inform social simulations.

1 INTRODUCTION

Edmonds and Moss have long emphasized that "all available evidence should be used" (Edmonds and Moss 2004) when building a model, particularly when it seeks to capture a complex phenomenon and we need to precisely understand its inner workings. For many modelers, the notion of 'evidence' often evokes quantitative data, such as time series. However, "qualitative information *can* be formally modeled in simulations where the deduction of outcomes is performed computationally" (Edmonds and Moss 2004). Examples can be found across types of models: observations from experts can be captured as graphical functions in System Dynamics (Eker et al. 2014), well-tested processes allow stakeholders to shape an Agent-Based Model by sharing their insight through games (Wimolsakcharoen et al. 2021), and a text corpus can be turned into a Fuzzy Cognitive Map using natural language processing (Pillutla and Giabbanelli 2019). While quantitative evidence may be a direct measurement of a system (e.g., sensors can record time series, patients' vital signs can be reported), qualitative evidence is often produced by individuals to express

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

their perceptions of a system. Approaches to facilitate the integration of qualitative evidence in a model thus differ based on how to involve individuals, who may represent stakeholders, subject-matter experts, or community members (Voinov et al. 2018). In a simplified dichotomous categorization, individuals are either *indirectly* involved in a modeling project when modelers use secondary sources such as articles or tweets (Papada et al. 2019; Çoban and Onar 2017; Sandhu et al. 2019), or *directly* involved when modelers elicit their perspectives through means including in-person interviews (Henly-Shepard et al. 2015) or voice-activated technologies (Reddy et al. 2019). A direct elicitation may be limited to informing the scope of a model and, upon its completion, confirm its validity. Alternatively, individuals can be the primary creator of the model and modelers have a facilitation role: this approach is known as *participatory modeling* (Voinov et al. 2018) and constitutes the application context for this paper.

Participatory modeling often uses Fuzzy Cognitive Mapping (FCM) to elicit and externalize the mental model of an individual, resulting in a graph or 'map'. The graph is labeled to represent concepts, which are connected via directed, weighted edges to capture the causal strength. Fuzzy membership functions are a defining feature of an FCM (Nápoles et al. 2020) and serve to represent uncertainty or vagueness, for example in the causal strengths. Most importantly, we stress that an FCM is not merely a picture of a system: it is a simulation model (Papageorgiou and Salmeron 2014), because it is able to update the concept values based on connected concepts and causal strengths. FCMs have been used for several decades, either in participatory modeling (Firmansyah et al. 2019) or as a soft computing tool for classification and forecasting (Felix et al. 2019). Innovations continue to be regularly presented, as evidenced by a 2010 volume on advances in tools and methods (Glykas 2010) or a special issue in 2017 (Froelich and Salmeron 2017). In this paper, we focus on two such recent innovations, which pertain to *hybrid* simulation methods and primarily target *social simulations*. First, recall that an FCM provides a transparent way to externalize the perspectives of an individual or group. If we seek to simulate *many* virtual entities who follow the complex decision-making processes of their real-world counterparts, then we necessarily need several FCMs. However, an FCM only provides a 'mind' hence FCMs would not be interacting unless they were put within bodies. The idea of equipping each entity with an FCM serving as its mind leads to the first innovation: using FCMs as the minds of agents in an Agent-Based Model (Giabbanelli et al. 2017; Davis et al. 2019). Although algorithms and simulation environments have been developed to support this hybrid approach (Giabbanelli et al. 2019), a fundamental question remains unanswered: which patterns of interactions and decision-making processes should be supported by these FCM/ABM models to capture social phenomena? In other words, what should be the *building blocks* of simulation models that equip agents with sophisticated decision-making modules in the form of FCMs? This question is important to impact simulation practices, in line with regular calls by researchers to integrate ABMs with "other simulation methodologies to represent higher level or hierarchical processes" (Naghshbandi et al. 2020).

Second, FCMs have historically been limited by an exclusive reliance on qualitative evidence: they only served to represent the *perspectives* of an individual or a group. The situation has changed with the emergence of machine learning for FCMs. Some of these methods view FCMs as neural networks and use heuristics to rewire them extensively, until the desired quantitative output from the FCM matches the one provided in a dataset. Although such approaches can yield high accuracy, the fact that they may entirely rewire one's perspectives makes them unsuitable for participatory modeling or the analysis of causal structures (Nápoles et al. 2020). The most commonly employed (Hebbian-like algorithms) thus consists of *adjusting* the causal weights of an FCM such that the simulation reproduces the target pattern (Papageorgiou 2011). However, there is a *multiplicity* of solutions that score equally on current performance metrics (e.g., accuracy): there are many ways in which a person's mind (as represented by an FCM) may be adjusted to align with a quantitative evidence base. This is a practical challenge: if one's perspective may be tweaked in seemingly interchangeable ways to account for another evidence, then which alteration is the most plausible given how individuals tend to change? Identifying the right solution is essential given the need for models combining quantitative and qualitative evidence (Mkhitaryan et al. 2020), as exemplified in reviews noting the under-representation of social datasets among primarily quantitative models (Steger et al. 2021).

In this paper, we take an interdisciplinary approach regarding the two questions above on hybrid FCM models for social simulations. Specifically, we examine these questions from the lens of psychology research. This is motivated by the fact that both of these simulation questions are echoed by the literature on human behaviors, which has long sought to identify key patterns of interactions and decision-making processes, or document how individuals can change their mind when exposed to a new evidence. Consequently, this paper seeks to promote an interdisciplinary discussion and potential synergies on hybrid FCM models by examining the relevance of existing frameworks and experiments in psychology.

The remainder of the paper is organized as follows. In section 2, we briefly cover the functioning of FCMs as well as the two innovations that motivate this paper (FCM/ABM and FCM/Machine Learning hybrids). Section 3 is devoted to FCM/ABM while section 4 covers FCM and Machine Learning, thus allowing modelers with interest in one specific question to focus on the corresponding section. Brief concluding remarks are provided in section 5.

2 BACKGROUND: FUZZY COGNITIVE MAPS FOR SIMULATIONS

2.1 Fuzzy Cognitive Maps: A Primer

A cognitive map describes the perceived behavior of a system, as seen by an individual or group. An FCM provides a simulation model for a cognitive map, articulated around two components, which we now introduce intuitively. The *structure* of the map consists of nodes, to describe salient constructs in the problem space (e.g., the depth of a lake, the amount of fish or birds), and edges, to characterize causal effects (e.g., the more the birds and the fewer the fish). Among the many graph-based simulation approaches (e.g., System Dynamics, Bayesian Networks), FCMs stand out by their approach to causal effects: they represent the strength of causation rather than rates or probabilities, and they allow any structure to emerge (e.g., loops). The *dynamics* of the system consist of iteratively updating the nodes' values until a subset of interest stabilizes (i.e., the 'output' nodes). Each node is updated based on the neighbors that causally impact it (i.e., visiting each incoming edge) and the strengths of these effects.

Mathematically, "an FCM is a nonlinear dynamical system akin to a neural network. Its structure is a fuzzy signed digraph with feedback, where nodes are fuzzy sets taking values in [0, 1] and edges are fuzzy rules" (Giabbanelli et al. 2019). The state of an FCM at step (or 'iteration') t can be represented by $F^t = (V^t, E, F)$, where V^t is a vector of n concept nodes, each having a value $V_i^t \in [0,1]$; E is a set of edges, whose causal weights are encoded by an adjacency matrix A in which positive weights specify a causal increase (if i increases then j increases) and negative weights specify a causal decrease (if i increases then j decreases); and f is a clipping (or 'transfer') function that keeps the value of an updated node within its range [0,1]. We note that there are different ways to denote an FCM, for example by separating the set of concept nodes from their values (Felix et al. 2019). The update of an FCM can be specified as follows:

$$V_i^{t+1} = f\left(V_i^t + \sum_{j=1, j\neq i} V_j^t \times A_{j,i}\right), \tag{1}$$

Note that Equation 1 was proposed in the early 2000's, as a modification of the original equation used for FCMs (Felix et al. 2019). An essential difference is the introduction of the term V_i^t, which ensures that an input in the model (i.e., a concept without incoming edges) would keep its value. Equation 1 is performed as long as the set of output concepts $S \subseteq V$ has not stabilized (given a user-specified threshold ε) and we perform less than a user-specified maximum number of iterations (to prevent a chaotic attractor):

$$\text{Apply Eq. 1 while} \begin{cases} \exists i \in S \text{ such that } |V_i(t+1) - V_i(t)| > \varepsilon, \text{ and} \\ t < t_{max} \end{cases} \tag{2}$$

Figure 1 shows three consecutive applications of Eq. 1. For details about the implementation of FCMs or their use in simulations, we refer the readers to our WinterSim'17 article (Lavin and Giabbanelli 2017).

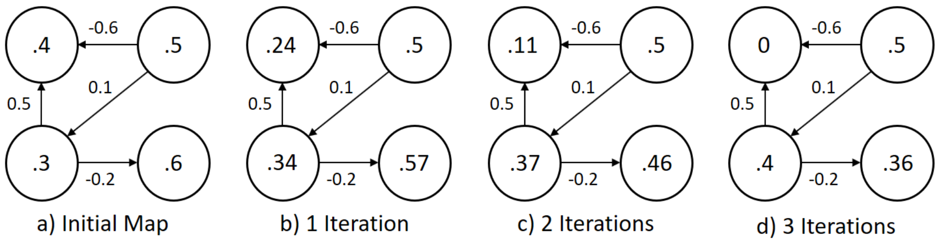

Figure 1: Node values are updated using Eq. 1 over 3 iterations, where f is the hyperbolic tangent function. Nodes are usually labeled with concept names: we instead display their values to illustrate the operations.

2.2 Using Fuzzy Cognitive Maps as the Minds of Agents

Agents *perceive* their surroundings, consisting of some other agents and parts of the environment. Then, they *reflect* on these perceptions and decide on a course of *actions*. The hybrid approach of ABM-FCM supports these mechanisms by equipping each agent with an FCM. Concept values are different since they capture the specific traits of an individual (e.g., different levels of income or varied beliefs) and the FCM structures may also be different as individuals do not necessarily think alike. An iteration of the whole model is decomposed into two steps. First, agents interact, for example by moving in the environment and being in contact with others. These interactions have an initial effect on specific concept nodes of the FCM; for example, the level of 'interest in healthy foods' may increase to register an agent's initial excitement from a conversation on nutrition. In other words, some concepts in an agent's FCM may change due to interactions. Second, the FCM is updated to model how an agent reflects on these perceptions given their specific context. For instance, an initial interest in healthy foods may disappear due to the mediating causal effect of factors such as lack of time and low income. Alternatively, the initial interest may tilt an agent's behavior if it provides the impetus to start a positive cycle. Once the FCM has stabilized, the agent has arrived at a conclusion and the new behavior is then reflected in their next interactions. Simulation environments for an FCM-ABM (Giabbanelli et al. 2017) model often include several FCMs, an ABM, and two components working at the interfaces of these approaches: an 'assignment' module to identify which agents get which FCMs, and an 'influence' module (shown in Figure 2) to specify the initial perceived effect of interactions. Section 3 focuses on the creation of behavioral templates for these specifications.

2.3 Adjusting a Fuzzy Cognitive Map: the Use of Machine Learning

The causal weights of an FCM are initially elicited from participants, in the context of a participatory modeling study. The perceptions captured in the FCM may be combined with a separate, quantitative dataset. This represents how perceptions may be adjusted if participants were exposed to an evidence base, and it improves the validity of a model as it takes into account additional measures about a system. Machine Learning provides the tools to transform the weights based on quantitative data. The transformation may be radical and computational intensive, by ignoring the initial weights entirely and optimizing an FCM by applying population-based search heuristics or viewing it as a neural network. Alternatively, the transformation may be more gradual, "to preserve the human knowledge while adjusting parts of the network that should be learned from historical data" (Nápoles et al. 2020). Such gradual approaches include methods known as 'Active', 'Balanced Differential', or 'Nonlinear' Hebbian Learning methods (AHL, BDHL, NHL). For example, participants are surveyed about a specific topic to provide initial edge weights, as usual for an FCM; in addition, a longitudinal survey can tell us how node values change and their typical values (e.g., mean). An NHL uses these initial edge weights, concept values, and ranges to adjust the edge weights based on two parameters (learning rate, decay coefficient). Admissible solutions may be found for several values of these parameters, such that there are different ways to adjust the edge weights while meeting the objectives. For example, Figure 3 from an ongoing study (Mkhitaryan et al. 2020) shows two ways in which the value of each edge can be changed to align with quantitative data.

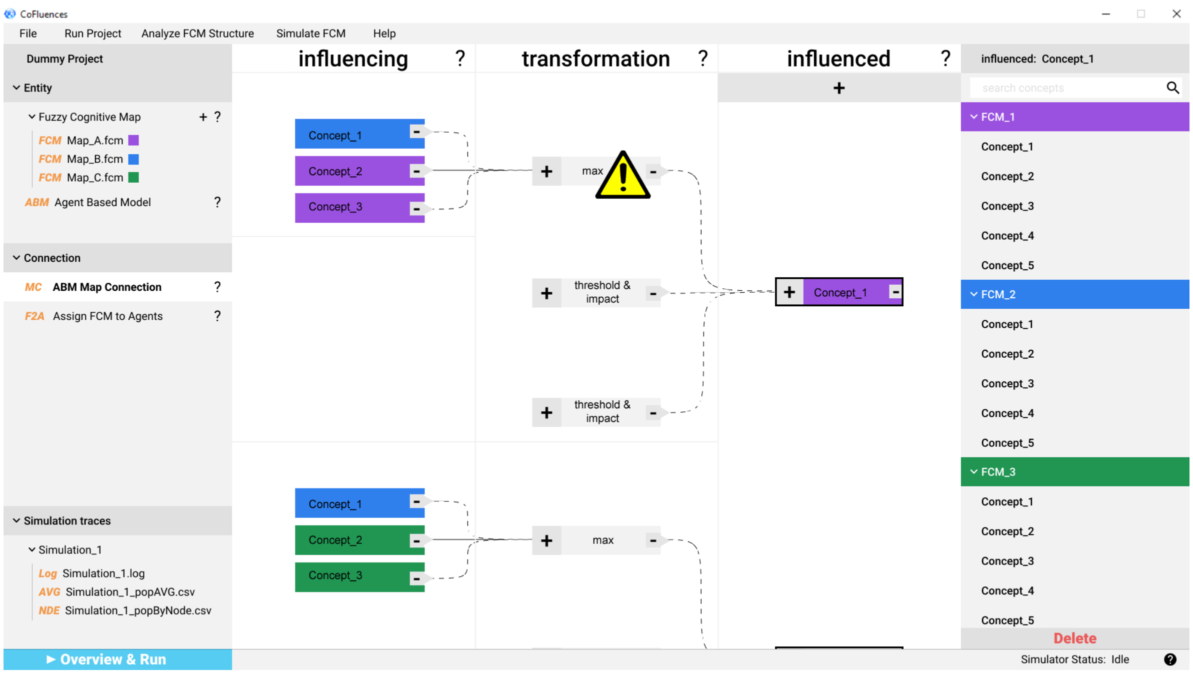

Figure 2: To connect the ABM and FCM modeling paradigms, we specify which concepts of an FCM may be initially influenced by interactions ('influenced' panel), which concepts of other agents' will influence them ('influencing' panel), and the nature of this influence ('transformation' panel).

3 CREATING A HYBRID SIMULATION OF SOCIAL SYSTEMS: WHICH BEHAVIORAL TEMPLATES DO WE NEED?

In this section, we review frameworks and models from psychology to guide innovations in simulation modeling regarding hybrid FCM-ABM models of social systems. In subsection 3.1, we focus on salient parameters involved in how individuals resolve the changes brought to their mental models from an interaction. Subsection 3.2 then shifts to the interaction itself, by identifying the parameters that shape social influences. These two sets of parameters are combined in subsection 3.3 from a simulation perspective.

3.1 Cognitive Dissonance Theory

Festinger proposed one of the most influential theories in social psychology to explain how people respond to incompatible elements in their mental models (Gawronski 2012). His theory of cognitive dissonance posits that concept nodes are either in congruence (i.e., consonant cognitions) or in conflict with each other (i.e., dissonant cognitions). Such nodes includes beliefs, desires, emotions, and behaviors (Sakai 1999). For an individual, the incompatibility of nodes is postulated to be a major source of discomfort (Festinger 1957; Harmon-Jones and Mills 2019), not only psychologically (as believed in the original theory) but also physically (Kitayama et al. 2013). Individuals thus seek to minimize the experienced discomfort by minimizing the dissonance (Festinger 1957). Figure 4 illustrates this process via the interaction of two individuals (Leo and Sofia), adapted from Festinger (1957), together with the four possible actions that Leo can take to minimize the dissonance. The process of cognitive dissonance applies to *all* agents directly involved in the interaction and also to observers (Sakai 1999). In our example, Sofia is also exposed to Leo's nodes, which may be in dissonance with her own elements of cognition.

The specific course of action taken by an individual depends on the level of discomfort (hence, the level of motivation to change), which is proportional to the *magnitude of dissonance*. In turn, the magnitude of dissonance of a node and the rest of an individual's mental model is a function of the *number* and the

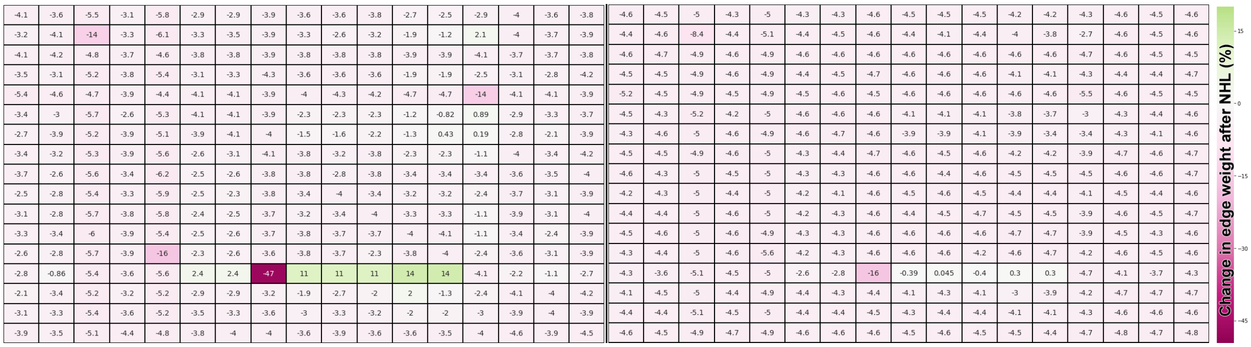

Figure 3: The two adjacency matrices depict how edge weights may be modified by NHL. Both matrices result in aligning the same initial FCM with the same evidence base, but they are obtained for different values of the NHL parameters. The left matrix minimizes the total change, at the expense of large changes. The right matrix minimizes the largest change, at the expense of changes in many edges. *This high resolution figure can be zoomed in, using the digital version of this article.*

importance of nodes that are consonant and dissonant with the focal element. For example, in Figure 4, being creative when writing (E4) may be less important to Leo than his beliefs on smoking (E2). The likelihood of change also depends on the *resistance to change* of a node, which is driven by its responsiveness to reality and the extent to which it is consonant with other nodes in the mental model (Festinger 1957).

Shultz and Lepper (1999) and Sakai (1999) proposed two models for cognitive dissonance process. The first model is based on an artificial neural network based that operates on the principles of constraint satisfaction (Shultz and Lepper 1996), which states that a system tends to a state where constraints are satisfied. Similar to FCMs, the network is composed of units (i.e., such cognitive elements as knowledge and attitudes) that are connected by weighted causal edges. Causal connections between units are excitatory, inhibitory, or nonexistent. The system reaches the optimum solution where the constraints are satisfied through specific rules (Shultz and Lepper 1999). Sakai's Multiplicative Power-Function model (MPF) provides an alternative model for cognitive dissonance, but it focuses on the formalization of the magnitude of the dissonance rather than the process of dissonance and the minimization thereof.

Although researchers agree that cognitive dissonance motivates cognitive changes, there are debates on the nature of motivation underlying these changes (Harmon-Jones and Mills 2019) and how to model them. For example, Sakai (1999) provides an important critique to Shultz and Lepper's model.

3.2 Social Influence Network Theory

The Social Influence Network (SIN) Theory describes how people evaluate and assimilate their own attitudes and that of the others on an issue. The SNI theory includes three theoretical constructs: *person's attitudes*, *susceptibilities*, and *interpersonal influences* (Friedkin and Johnsen 2011). The theory builds on several models, such as French's model of social power (French 1956) and DeGroot's consensus formation model (Berger 1981). It also relates to rational choice model of group decision making (Lehrer and Wagner 2012), the model of group decision making (Graesser 1991) and Anderson's weighted averaging model of information integration (Anderson 1981).

French's model builds on a widely accepted notion that people seek to minimize tensions in interpersonal relations by shifting their attitudes to the average attitudes of the people influencing them. French introduced the idea of social structure in interpersonal influences by postulating that people will adjust their attitudes only to those who *directly* influence them. A key implication of this social structure is that attitudes may not converge towards the average of the *group*, but rather towards the average among specific individuals (Friedkin and Johnsen 2011). French's model makes two assumptions: all connections are equally influential (i.e., homogeneity of interpersonal influences), and the resistance to social influence

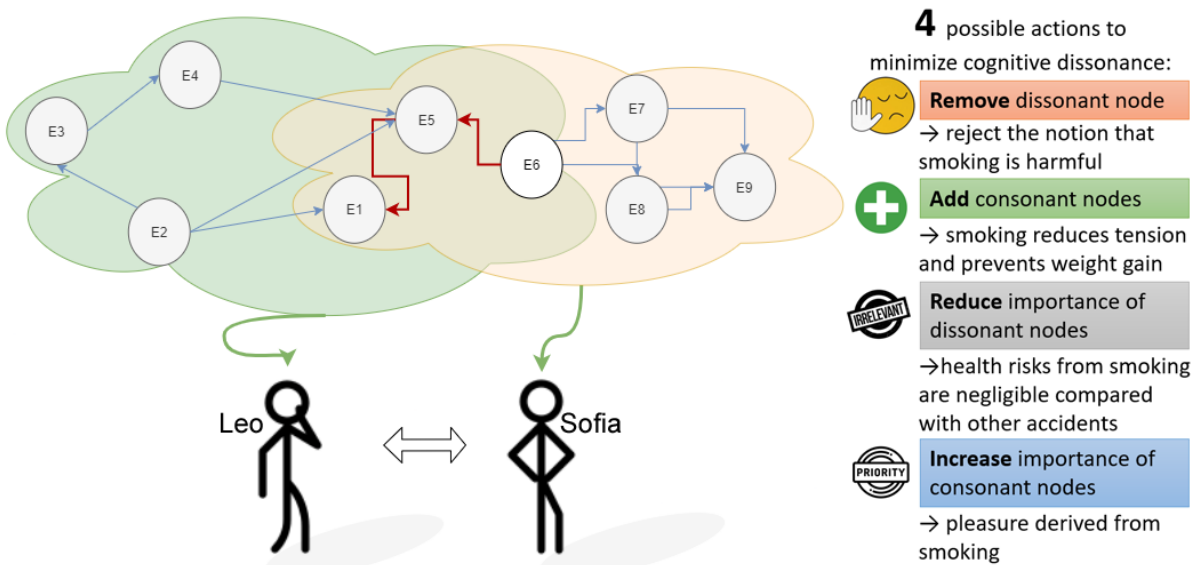

Figure 4: Leo is a habitual smoker (E1) who believes that smoking is cool (E2), it helps him to relax (E3), and be more creative when he needs to write an essay (E4). In conversation with Sofia, Leo learns that smoking is bad for health (E6). Now, this new node (E6) is within Leo's mental model, but it is incompatible (i.e., in dissonance) with existing nodes (e.g., with the fact that Leo is a habitual smoker).

is proportional to the interpersonal influence. Harary (1959) extended French's model by relaxing this second assumption, thus allowing self-weights to differ from interpersonal weights. DeGroot (1974) later also relaxed the first assumption to examine the conditions under which a group consensus is reached. Finally, Friedkin and Johnsen (1990) built on DeGroot's work to propose the general model, captured by Equation 3 and extended in their subsequent works (Friedkin and Johnson 1999). In this model, people come into contact through social interactions and share their perspetives, such that mental models are shaped during interactions. Formally, the perspectives held at time t regarding a given concept (y_i^{t+1}) depends on one's perspective before interactions (y_i^1), their susceptibilities to interpersonal influences (a_{ii}), and the attitudes of interacting peers (y_j^t) weighted by their relative influences on the focal individual (w_{ij}). The lack of resistance to interpersonal influence is specified as $1-a_{ii}$, where $(1-a_{ii}) = 1-w_{ii}$. This model permits analysis of social processes both in a group with minimal social structure as well as in a complete network (Friedkin and Johnsen 2011). This equation resembles Equation 1, as they both compute the next value of nodes of a network based on the previous value and connected nodes.

$$y_i^{t+1} = a_{ii} \sum_{j=1}^{n} w_{ij} \times y_j^t + (1-a_{ii}) \times y_t^1 \tag{3}$$

3.3 Takeaways for Hybrid Simulation Models of Social Phenomena

Current frameworks and case studies consider that an agent will change the weight of some FCM concepts as a result of its interactions (Davis et al. 2020). While studies in psychology confirm this change, they also point out that changes exist in *other agents*, for a variety of reasons that are not captured in current frameworks (e.g., based on the *magnitude* of dissonance), and are not limited to changing a concept weight. Based on these studies, we have the following four takeaways for simulation methods:

- There are four possible actions to minimize cognitive dissonance (Figure 4; right). Two of them cannot be achieved through current frameworks, which only allow changes in the values of existing concept nodes. Consequently, the frameworks may be expanded to accept the creation and deletion of nodes. This is a challenge for the interpretability of FCMs, which is one of their current strengths: if nodes are automatically created during a model, then they need to be give a name, and at present we lack methods to automatically infer meaningful new labels.

- In present frameworks, changes are only dictated by the current value of (*i*) an individual's influenced node and (*ii*) the influencing nodes among peers. However, the 'magnitude of dissonance' highlights that some nodes are more important than others. An extended FCM may thus be needed, such that each concept node is equipped with a pair of values to track its current weight and its importance.
- At present, simulations consider that only individuals who are influenced can change, thus viewing the process as unidirectional (Figure 2). However, all agents involved in the interaction (and observers) can change. Modeling specifications for ABM/FCM may thus need additional elements, to specify the reciprocal influence as well as the effect on observers.
- Social Influence Network Theory (Equation 3) posits that changes take into account susceptibilities to interpersonal influences as well as the relative influences of peers. Neither of these aspects are currently included in ABM/FCM specifications, which can thus be expanded.

4 WHEN MACHINE LEARNING CHANGES AN AGENT'S BEHAVIOR: WHICH CHANGES TO EXPECT?

To understand how machine learning algorithms may automatically update an individual's perspective, we start by examining experiments from psychology research in which individuals change. These experiments are often conducted in a group setting, such that each individual is exposed to the 'evidence' formed of others' perspectives in the group. Based on the changes observed in these situations, we then identify which trajectories may be preferable for the automatic update of a model using machine learning.

4.1 Evidence From Social Psychology Research

In an early experiment, Stoner (1961) gave short stories to each participant and asked them to make a binary choice as the protagonist, often offering a safe choice and a riskier one (i.e., entails more risk for the protagonist). Once participants chose, they gathered as a group, discussed the story, and chose as a group. The choice of a group tended to be riskier than the earlier choices made by participants. This notion has since been known as the 'risky shift': a group's decisions are riskier than the average decisions of its individual members before the group met (Figure 5). Later experiments have broadened this to the *group polarization hypothesis*, stating that the decisions made by a group exacerbate the initial tendencies of its members: greater risk is obtained when members were inclined towards risky choices, and even safer choices are found when members tended towards safe choices. Experiments have examined the distinction between the shifts in the group's response and the shifts of its individuals, thus assessing whether polarization only occurred at the level of the group or also at the individual level. In experiments, participants first expressed their opinions on each item (e.g., on Likert type scales), then had a group discussion to arrive at a consensual decision on each item, and finally filled again their individual questionnaires (Sunstein 2009). These experiments found differences between the group's response and the initial opinions of its participants, such that the group's response after discussion was more strongly in the direction of the initial predispositions of the participants (Zhu 2013; Lamm 1988). These differences depend on the task: they are generally larger in opinion related tasks than in judgment related tasks, for which participants aim at being 'objective'. Analyses of the individual responses to the same items following the group discussions also suggest shifts in individuals' own attitudes (Moscovici and Zavalloni 1969).

In early experiments on group polarization, participants were instructed to reach a consensus during their group discussion. However, groups may adopt other procedures such expert panels with moderated discussions. Complementary experiments have thus examined whether group polarization could be an artifact of a specific group procedure. In one set of studies, the authors focused on the group discussion process of 'deliberation', which is defined as "a combination of careful problem analysis and an egalitarian process in

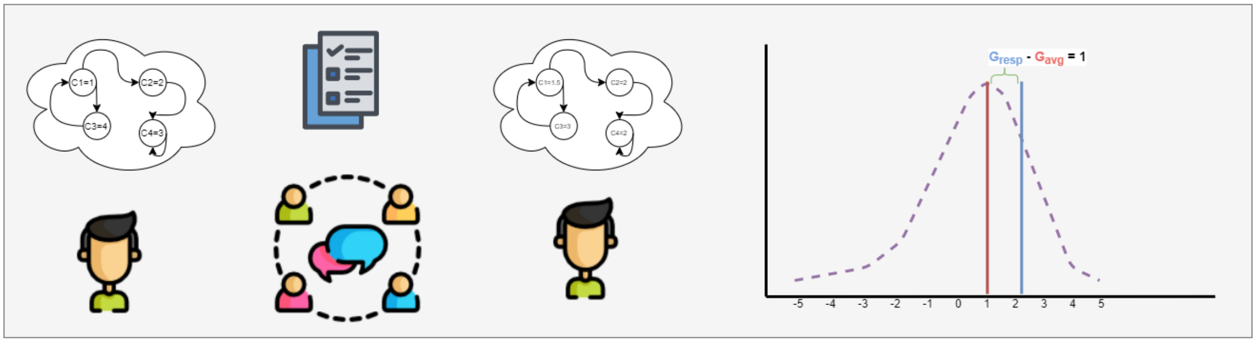

Figure 5: The experimental set-up of a group polarization study.

which participants have adequate speaking opportunities and engage in attentive listening or dialogue that bridges divergent ways of speaking and knowing" (Burkhalter et al. 2002). The group discussions in the deliberation studies often follow certain normative (e.g., respect for differing perspectives) and procedural standards (e.g., exposure to balanced information) (Schneiderhan and Khan 2008; Steenbergen et al. 2003; Dryzek and Niemeyer 2006; Niemeyer and Dryzek 2007; Gastil and Dillard 1999). Once again, experiments found significant changes in opinions following deliberation process (Price et al. 2006; Grönlund et al. 2009). However, the changes in opinions and attitudes following deliberation are not always in line with the group polarization hypothesis. More specifically, while some experiments show polarization effect, others reveal shifts in opinions towards the average (i.e., moderation) (Isenberg 1986; Grönlund et al. 2009). Note the ability to compare findings is limited by a methodological difference: deliberation research compares the *average* responses before/after deliberation, whereas group polarization research compares the response from the *group* with individual responses before discussion (Myers and Lamm 1976; Isenberg 1986).

Changes in opinions and attitudes in groups may also vary depending on the group composition. In some participatory modeling approaches, ensuring the participation of diverse groups in the discussion groups is essential to ensure that the many facets of a problem can be covered (Mendoza and Prabhu 2006). Grönlund et al. (2015) looked at the polarization effect after deliberation in groups with people with similar predispositions and compared it to groups with mixed composition. Group polarization was expected and confirmed to occur as a result of the reinforcing effect of the exposure to similar attitudes and opinions prevailing in the group of "like-minded" people during the deliberation process. In contrast, mixed groups with diverse initial opinions and attitudes were expected to depolarize (move towards the average) as a result of listening to opposing opinions. Results confirmed this depolarization effect (Luskin et al. 2002; Setälä et al. 2010; Grönlund et al. 2010). Although these results may appear to contradict earlier experimental findings on group polarization hypothesis, we stress that they are not fully comparable due to the procedural difference between group discussion and deliberation.

The studies above have examined differences on items in a questionnaire. However, a broader array of changes may take place when individuals are exposed to new evidence. In the case of cognitive dissonance, they may add or remove nodes together with their edges (Figure 4; right). Such structural changes are also possible when the evidence reinforces an individual's mental model. In contrast to studies on group discussion and deliberation, there is a relative paucity of studies examining how an entire map may change when exposed to evidence. Our recent study shows that new nodes and edges can appear when exposed to different forms of evidence, and the amount depends on the nature of the evidence (Giabbanelli and Tawfik 2021). Datasets to study these effects are manually collected and tend to be small, which limits the depth of the analysis and the ability to draw generalizable conclusions. Limitations due to sample sizes are gradually addressed with the deployment of instructional systems that automatically prompt and track changes in one's mental model (Giabbanelli and Tawfik 2020).

4.2 Takeaways for Machine Learning Applied to Mental Models

At present, machine learning algorithms change the weights of FCMs by assuming that individuals may change their mind about anything, in any direction. These algorithms also tend to either only adjust existing weights (i.e., no new nodes or edges allowed) or re-create the map entirely (i.e., new nodes/edges may appear but existing ones are not taken into consideration). The evidence summarized in the previous section paints a very different situation, and thus highlights the need to either develop new algorithms or carefully select among the solutions generated by current ones. In particular, we have four takeaways:

- A *polarization* effect may happen when the evidence is in line with initial tendencies. In an FCM, this means that positive weights can increase and negative weights can decrease.
- If the evidence goes against initial tendencies, a *depolarization* effect suggests that an individual converges towards the average. To operationalize this effect in an FCM, optimized edge weights may be halfway between the individual's initial weights and those dictated by the evidence.
- The nature of the *task* is not captured by current machine learning algorithms, since there is no categorization attached to edges such that some are matters of opinion (hence an individual may tolerate a large change) and others are established judgments (hence an individual may be less inclined to change). For now, modelers can perform this categorization manually and determine an appropriate range of changes accordingly. In the next generation of machine learning algorithms for FCMs, this categorization can become a useful feature.
- *Structural* changes can happen in reaction to new evidence, but further research is needed on this topic. For example, are new nodes more peripheral or central in a map? Are new nodes more likely to impact dissonant or consonant concepts? Are some structures of thoughts (e.g., cycles) more likely to emerge after exposure to evidence and further reflections? Such structural questions can constitute a fruitful research agenda for social psychologists, modelers, and network scientists.

5 CONCLUSION

Advances on modeling and simulation with Fuzzy Cognitive Maps (FCMs) have largely been independent from psychology research, even though these advances are concerned with topics such as how perceptions are shaped by social interactions (ABM/FCMs models) or how perceptions may change when exposed to new evidence (machine learning for FCMs). To guide the next generation of methodological advances, we proposed to reconcile domain expertise with simulation methods by identifying missing aspects in current simulation frameworks based on findings from psychology research. We provided specific lists (totaling 8 items), which can contribute to setting the agenda for future methodological innovations with FCMs.

REFERENCES

Anderson, N. H. 1981. *Foundations of information integration theory*, Volume 578. Academic Press New York.

Berger, R. L. 1981. "A necessary and sufficient condition for reaching a consensus using DeGroot's method". *Journal of the American Statistical Association* 76(374):415–418.

Burkhalter, S., J. Gastil, and T. Kelshaw. 2002. "A conceptual definition and theoretical model of public deliberation in small face—to—face groups". *Communication theory* 12(4):398–422.

Çoban, V., and S. Ç. Onar. 2017. "Modelling solar energy usage with fuzzy cognitive maps". *Intelligence systems in environmental management: theory and applications*:159–187.

Davis, C. W., P. J. Giabbanelli, and A. J. Jetter. 2019. "The intersection of agent based models and fuzzy cognitive maps: a review of an emerging hybrid modeling practice". In *2019 Winter Simulation Conference (WSC)*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 1292–1303. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Davis, C. W., A. J. Jetter, and P. J. Giabbanelli. 2020. "Fuzzy cognitive maps in agent based models: a practicial implementation example". In *2020 Spring Simulation Conference (SpringSim)*, 1–11: Institute of Electrical and Electronics Engineers, Inc.

DeGroot, M. H. 1974. "Reaching a consensus". *Journal of the American Statistical Association* 69(345):118–121.

Dryzek, J. S., and S. Niemeyer. 2006. "Reconciling pluralism and consensus as political ideals". *Am J Pol Sci* 50(3):634–649.

Edmonds, B., and S. Moss. 2004. "From KISS to KIDS–an 'anti-simplistic'modelling approach". In *International workshop on multi-agent systems and agent-based simulation*, 130–144. Springer.

Eker, S., J. Slinger, E. van Daalen, and G. Yücel. 2014. "Sensitivity analysis of graphical functions". *System Dynamics Review* 30(3):186–205.

Felix, G., G. Nápoles, R. Falcon, W. Froelich, K. Vanhoof, and R. Bello. 2019. "A review on methods and software for fuzzy cognitive maps". *Artificial intelligence review* 52(3):1707–1737.

Festinger, L. 1957. "A theory of cognitive dissonance Row".

Firmansyah, H. S., S. H. Supangkat, A. A. Arman, and P. J. Giabbanelli. 2019. "Identifying the components and interrelationships of smart cities in Indonesia: Supporting policymaking via fuzzy cognitive systems". *IEEE Access* 7:46136–46151.

French, J. R. P. 1956. "A formal theory of social power.". *Psychological review* 63(3):181.

Friedkin, N. E., and E. C. Johnsen. 1990. "Social influence and opinions". *Journal of Mathematical Sociology* 15(3-4):193–206.

Friedkin, N. E., and E. C. Johnsen. 2011. *Social influence network theory: A sociological examination of small group dynamics*, Volume 33. Cambridge University Press.

Friedkin, N. E., and E. Johnson. 1999. "Influence networks and opinion change". *Advances in Group Processes* 16(1):1–29.

Froelich, W., and J. L. Salmeron. 2017. "Advances in fuzzy cognitive maps theory". *Neurocomputing* 232.

Gastil, J., and J. P. Dillard. 1999. "Increasing political sophistication through public deliberation". *Polit Commun* 16(1):3–23.

Gawronski, B. 2012. "Back to the future of dissonance theory: Cognitive consistency as a core motive". *Soc Cogn* 30(6):652–668.

Giabbanelli, P., M. Fattoruso, and M. L. Norman. 2019. "Cofluences: Simulating the spread of social influences via a hybrid agent-based/fuzzy cognitive maps architecture". In *Proceedings of the 2019 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 71–82.

Giabbanelli, P. J., S. A. Gray, and P. Aminpour. 2017. "Combining fuzzy cognitive maps with agent-based modeling: Frameworks and pitfalls of a powerful hybrid modeling approach to understand human-environment interactions". *Environmental modelling & software* 95:320–325.

Giabbanelli, P. J., and A. A. Tawfik. 2020. "Reducing the Gap Between the Conceptual Models of Students and Experts Using Graph-Based Adaptive Instructional Systems". In *International Conference on Human-Computer Interaction*, 538–556.

Giabbanelli, P. J., and A. A. Tawfik. 2021. "How Perspectives of a System Change Based on Exposure to Positive or Negative Evidence". *Systems* 9(2):23.

Glykas, M. 2010. *Fuzzy cognitive maps: Advances in theory, methodologies, tools and applications*, Volume 247. Springer.

Graesser, C. C. 1991. *A Social Averaging Theorem for Group Decision Making*. Ph. D. thesis, University of California, San Diego, CA.

Grönlund, K., K. Herne, and M. Setälä. 2015. "Does enclave deliberation polarize opinions?". *Political Behavior* 37(4):995–1020.

Grönlund, K., S. Himmelroos et al. 2009. "The challenge of deliberative democracy online–A comparison of face-to-face and virtual experiments in citizen deliberation". *Information Polity* 14(3):187–201.

Grönlund, K., M. Setälä, and K. Herne. 2010. "Deliberation and civic virtue: Lessons from a citizen deliberation experiment". *European Political Science Review: EPSR* 2(1):95.

Harary, F. 1959. *A criterion for unanimity in French's theory of social power.*, 168–182. Univer. Michigan.

Harmon-Jones, E., and J. Mills. 2019. *An introduction to cognitive dissonance theory and an overview of current perspectives on the theory.*, 3–24. American Psychological Association.

Henly-Shepard, S., S. A. Gray, and L. J. Cox. 2015. "The use of participatory modeling to promote social learning and facilitate community disaster planning". *Environmental Science & Policy* 45:109–122.

Isenberg, D. J. 1986. "Group polarization: A critical review and meta-analysis.". *J Pers Soc Psycholy* 50(6):1141.

Kitayama, S., H. F. Chua, S. Tompson, and S. Han. 2013. "Neural mechanisms of dissonance: An fMRI investigation of choice justification". *Neuroimage* 69:206–212.

Lamm, H. 1988. "A review of our research on group polarization: Eleven experiments on the effects of group discussion on risk acceptance, probability estimation, and negotiation positions". *Psychological Reports* 62(3):807–813.

Lavin, E. A., and P. J. Giabbanelli. 2017. "Analyzing and simplifying model uncertainty in fuzzy cognitive maps". In *2017 Winter Simulation Conference (WSC)*, edited by W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 1868–1879. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lehrer, K., and C. Wagner. 2012. *Rational consensus in science and society: A philosophical and mathematical study*, Volume 24. Springer Science & Business Media.

Luskin, R. C., J. S. Fishkin, and R. Jowell. 2002. "Considered opinions: Deliberative polling in Britain". *British Journal of Political Science*:455–487.

Mendoza, G. A., and R. Prabhu. 2006. "Participatory modeling and analysis for sustainable forest management: Overview of soft system dynamics models and applications". *Forest Policy and Economics* 9(2):179–196.

Mkhitaryan, S., P. J. Giabbanelli, N. K. de Vries, and R. Crutzen. 2020. "Dealing with complexity: How to use a hybrid approach to incorporate complexity in health behavior interventions". *Intelligence-Based Medicine* 3:100008.

Moscovici, S., and M. Zavalloni. 1969. "The group as a polarizer of attitudes.". *J Pers Soc Psychol* 12(2):125.

Myers, D. G., and H. Lamm. 1976. "The group polarization phenomenon.". *Psychological bulletin* 83(4):602.

Naghshbandi, S. N., L. Varga, A. Purvis, R. Mcwilliam, E. Minisci, M. Vasile, M. Troffaes, T. Sedighi, W. Guo, E. Manley et al. 2020. "A review of methods to study resilience of complex engineering and engineered systems". *IEEE Access* 8:87775–87799.

Nápoles, G. et al. 2020. "Deterministic learning of hybrid fuzzy cognitive maps and network reduction approaches". *Neural Networks* 124:258–268.

Nápoles, G., J. L. Salmeron et al. 2020. "Fuzzy cognitive modeling: Theoretical and practical considerations". In *Intelligent Decision Technologies 2019*, 77–87. Springer.

Niemeyer, S., and J. S. Dryzek. 2007. "The ends of deliberation: meta-consensus and inter-subjective rationality as ideal outcomes". *Swiss political science review* 13(4):497–526.

Papada, L., N. Katsoulakos, I. Doulos, D. Kaliampakos, and D. Damigos. 2019. "Analyzing energy poverty with Fuzzy Cognitive Maps: A step-forward towards a more holistic approach". *Energy Sources, Part B* 14(5):159–182.

Papageorgiou, E. I. 2011. "Learning algorithms for fuzzy cognitive maps – a review study". *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(2):150–163.

Papageorgiou, E. I., and J. L. Salmeron. 2014. "Methods and algorithms for fuzzy cognitive map-based modeling". In *Fuzzy cognitive maps for applied sciences and engineering*, 1–28. Springer.

Pillutla, V. S., and P. J. Giabbanelli. 2019. "Iterative generation of insight from text collections through mutually reinforcing visualizations and fuzzy cognitive maps". *Applied Soft Computing* 76:459–472.

Price, V., L. Nir, and J. N. Cappella. 2006. "Normative and informational influences in online political discussions". *Communication Theory* 16(1):47–74.

Reddy, T., P. J. Giabbanelli, and V. K. Mago. 2019. "The artificial facilitator: guiding participants in developing causal maps using voice-activated technologies". In *International Conference on Human-Computer Interaction*, 111–129.

Sakai, H. 1999. "A multiplicative power-function model of cognitive dissonance: Toward an integrated theory of cognition, emotion, and behavior after Leon Festinger.". In *Convention of the Japanese Psychological Association, 48th, Oct, 1984, Osaka, Japan*. American Psychological Association.

Sandhu, M., C. D. Vinson, V. K. Mago, and P. J. Giabbanelli. 2019. "From associations to sarcasm: mining the shift of opinions regarding the supreme court on Twitter". *Online Social Networks and Media* 14:100054.

Schneiderhan, E., and S. Khan. 2008. "Reasons and inclusion: The foundation of deliberation". *Sociological Theory* 26(1):1–24.

Setälä, M., K. Grönlund, and K. Herne. 2010. "Citizen deliberation on nuclear power: A comparison of two decision-making methods". *Political Studies* 58(4):688–714.

Shultz, T. R., and M. R. Lepper. 1996. "Cognitive dissonance reduction as constraint satisfaction.". *Psychological review* 103(2):219.

Shultz, T. R., and M. R. Lepper. 1999. "Computer simulation of cognitive dissonance reduction.".

Steenbergen, M. R., A. Bächtiger, M. Spörndli, and J. Steiner. 2003. "Measuring political deliberation: A discourse quality index". *Comparative European Politics* 1(1):21–48.

Steger, C., S. Hirsch et al. 2021. "Linking model design and application for transdisciplinary approaches in social-ecological systems". *Global Environmental Change* 66:102201.

Stoner, J. A. F. 1961. *A comparison of individual and group decisions involving risk*. Ph. D. thesis, Massachusetts Institute of Technology.

Sunstein, C. 2009. *Going to extremes: How like minds unite and divide*. Oxford University Press.

Voinov, A., K. Jenni et al. 2018. "Tools and methods in participatory modeling: Selecting the right tool for the job". *Environmental Modelling & Software* 109:232–255.

Wimolsakcharoen, W. et al. 2021. "An agent-based model to support community forest management and non-timber forest product harvesting in northern Thailand". *Socio-Environmental Systems Modelling* 3:17894–17894.

Zhu, D. H. 2013. "Group polarization on corporate boards: Theory and evidence on board decisions about acquisition premiums". *Strategic Management Journal* 34(7):800–822.

AUTHOR BIOGRAPHIES

SAMVEL MKHITARYAN, MPH, MSc, is a behavioral data scientist at the Faculty of Health, Medicine and Life Sciences at Maastricht University. As a Ph.D. candidate, he focuses on exploring the potential of systems analysis in health behavior research. Current projects include the effect of peer influence on adolescent smoking, or the use of hybrid simulation techniques to model health behavior interventions. His email address is s.mkhitaryan@maastrichtuniversity.nl.

PHILIPPE J. GIABBANELLI, Ph.D., is an Associate Professor in the Department of Computer Science & Software Engineering at Miami University (USA). He holds a PhD from Simon Fraser University and taught at several nationally ranked American universities. His research interests include network science, machine learning, and simulation and applied to human health behaviors. His email address is giabbapj@miamioh.edu. His website is https://www.dachb.com.