

Organizations as Universal Computing Machines: Rule systems, computational equivalence, and organizational complexity

March 31, 2008 · Academic

Mihnea Moldoveanu, Mihnea Moldoveanu

Moldoveanu M, Moldoveanu M. Organizations as Universal Computing Machines: Rule systems, computational equivalence, and organizational complexity. *Emergence: Complexity and Organization*. 2008 Mar 31 [last modified: 2016 Nov 30]. Edition 1. doi: 10.emerg/10.17357.66d2beeebdfba021dfd96f49a17d4bcc.

Abstract

Organizations can be—and, have been—modeled as rule-based systems. On a reductive view, the resulting models depict organizations as cellular automata (CA) that carry out computations whose inputs are the initial and boundary conditions of a lattice of elements co-evolving according to deterministic interaction rules and whose outputs are the final states of the CA lattice. We use such models to refine the notion of the complexity of an organizational phenomenon and entertain the notion of an organization as a universal computer that can support a wide variety of CA to suggest ways in which CA-derived insights can inform organizational analysis. We examine the informational and computational properties of CA rules and the implications of the trade-off between their informational and computational complexity to the problem of ‘organizational design’ and show how the discovery of operational rules could proceed in the context of an empirical framework.

Introduction: Organizations as rule-following systems

Ever since the early work of Simon (1947), March and Simon (1958) and Cyert and March (1963) organizations have been frequently modeled as rule-following systems. Standard operating procedures (Cyert & March, 1963), routines (Nelson & Winter, 1982), socio-cognitive and cognitive scripts (Newell, 1990) and algorithms (Moldoveanu & Bauer, 2002) have all been posited as ways of understanding the dynamics of organizations through the decomposition of macro-level phenomena that seem to follow a pattern or regularity into micro-level phenomena that follow simple rules, along with some aggregation mechanism for the combined activity of the micro-level phenomena in question. Rule-based thinking about organizations has been used to understand the micro-analytics of individual behavior (March, 1994), the phenomenon of organizational ‘culture’ and ‘common law’ (March, and Simon, 1958) and to study the problem of organizational design (Simon, 1962), as follows:

The logic of using rule-based systems to study organizational phenomena

Appropriateness-to-a-rule as a criterion of ‘correct’ behavior (March, 1994)

One can model individual rationality either in a maximization-based framework common to economic theorizing (Elster, 1981; see also Sen, 1997) as the pursuit of those actions and activities that maximize the present value of some utility metric, or as adherence to sets of rules for thinking and acting (March, 1994) or for processing information (March & Simon, 1958). As March points out, it is possible to understand maximization-based logic in a rule-following framework, by positing that individual agents simply follow the ‘rules of instrumental reason’ (such as ‘consistency of preferences over time’ (see Sen, 1993 for a critique)) when they act ‘rationally according to the maximization framework. It is also possible to argue that individuals follow a maximization-based logic in adhering to certain rules (including rules for thinking and computation (Moldoveanu, 1999)), as rules allow for the predictable pursuit of ends and rule-following allows individuals to justify their actions to each other as legitimate. Even though there is no established ‘matter-of-fact’ about rule following as a general modeling hypothesis, the rule-following framework has produced a deep and fruitful research tradition in organization studies (which can be traced back to March and Simon, 1958; Cyert & March, 1962; Nelson & Winter, 1982) and can be understood as a foundation of the mind-as-computer metaphor (Gigerenzer, 1991) that lies at the foundation of the ‘cognitive revolution’ in cognitive psychology (Simon, 1992). It is to this tradition that this article contributes an understanding of the relationship between micro-level rules and macro-level rule dynamics.

Appropriateness-to-a-rule as a modeling assumption for organizations: Explaining an observed phenomenon in terms of the instantiation of a particular rule or rule set (March, 1994)

Rules (or norms) have also been frequently used for understanding organizations of individual agents, even without direct

reference to the rule-following proclivities of the individuals that constitute them. Organizations can be understood as symbolic systems, systems of meaning as well as systems of action and choice, all of which can be 'reduced' to proclivities to follow certain rules, certain rule sets, or, certain rules sets in certain conditions. Meaning, for instance, is created through collective action that embodies the rules and norms of organizations as a whole, and *organization* is about the collective pursuit of such meaning (Weick, 1995). By this account, organizational rule systems confer *meaning* upon "meaning". As organizational rules, individuated and exemplified by collective organizational action, become self-evident to members of the organization, they form a background of shared assumptions, categories and heuristics that are common knowledge (Lewis, 1969) among members of the organization, thus enabling members of the organization to pursue collective goals efficiently and reliably (Kreps, 1990). The meaning-giving activity of rules and rule systems thereby recedes into unconsciousness (and sometimes oblivion), which does not, nonetheless weaken the basic argument that rule systems—applied to the organizational level—can be used to understand the production of collective patterns of behavior.

Appropriateness-to-a-rule as a design assumption for organizations

Which rules are "best" for designing rule-bound systems—and, why? Herbert Simon (Simon, 1962, 1996) argued for the development of the 'sciences of the artificial'—of those phenomena and systems that are created by humans. These phenomena and systems include organizations, institutions, and the technologies which are used by organizations and institutions to carry out their work. The problem of organizational and institutional design becomes the problem of selecting those rules and systems of rules that are most likely to produce a particular desired outcome. Examples of successful rules include the 'modularization heuristic' and the 'separability heuristic' (Simon, 1962) for creating more survivable evolutionary entities in the presence of external 'noise' that disrupts the smooth internal functioning of these systems. Organizational culture can be understood as a system of rules and meta-rules for selecting certain kinds of behaviors, beliefs, theories, models or cognitive maps over others (Moldoveanu & Singh, 2003), and thus the Simonese framework can be applied to the problem of designing successful cultural systems (once a criterion for organizational success has been agreed upon).

The uses of rules in organizations and institutions

The fact that we can ask about 'best' rules depends on the fact that there are *many* kinds of rules and rule systems. It is important to distinguish between different kinds of organizational rule systems, and to ask: is it possible (or even conceivable) that they can be accommodated by a modeling framework? If so, what is the framework that can contain (i.e., represent, allow us to do computational and thought experiments with) the different kinds of rules and rule systems there are? And there are many. For instance:

Rules of conduct, comportment and coordination supply scripts and algorithms that allow individuals in an organization to interact successfully in coordination games (Schelling, 1960), argumentation games (Alexy, 1988; Toulmin, 1981), or morality games (Ruth Lakoff, 2001). An example of rules in this class include the cooperation principle (Grice, 1975) that is a pre-condition for participation in a dialogue (demanding that we understand the speaker's contributions as being relevant and informative to the discussion, on pain of being a non-contribution), and the principles of discourse ethics (Alexy, 1988) demanding consistency and setting standards for minimal responsiveness for speech acts (on pain of the latter becoming unintelligible collections of syllables, and thus outside of the realm of the dialogue being carried out).

Meta-rules govern the uses of rules and the ways in which rules are adapted in response to anomalies. The problem of organizational adaptation (Levinthal & March, 1993) is essentially the problem of adapting rule sets to different contexts, changing contexts and changing environmental conditions. The theory of adaptive algorithms (including evolutionary algorithms (Bruderer & Singh, 1996)) supplies a rich set of formalizable systems of meta-rules. Algorithms for search (March, 1991) can be used to study different patterns of organizational exploration in the same way as task-performance-simulating algorithms can be used to study the phenomena of organizational exploitation of a newly discovered opportunity or opportunity set.

Para-rules are rules for adjudicating among unforeseen incongruencies or tensions between rules (Cyert & March, 1963). Even without changing contexts, varying contexts and changing environmental conditions, the rule-based paradigm needs to access a special class of rules (we will call them para-rules) for adjudicating conflicts between different rule sets and thus resolving the important problem of prescribing action in situations when opposing rule systems claim equal cognitive or moral jurisdiction. Such para-rules are often ways of ascertaining truth or validity of empirical or moral rule systems, or of theories that organizations might collectively hold about the environment. They can constitute the 'epistemology of the organization' (Moldoveanu, 2002).

Ortho-rules are rules for resolving ambiguities and uncertainty about the correct application of a particular rule (Cohen & March, 1972). Classical epistemology turns up many indeterminacies and ambiguities in the establishment of the validity of the empirical claims of a theory or model: most of its core results are negative results: about the impossibility of a deductive basis for inductive logic (Hume, 1949), the ambiguity of confirmation (Hempel, 1951), the impossibility of constructing a rule system that can be used to prove its own validity (Putnam, 1985; following Gödel, 1931). Nevertheless, individual agents can and do bridge the logical (and meta-logical) divide between word and action (or event or object): they interpret events according to

particular rules, and act according to the reconstructed reality that they have created (Weick, 1995). What makes this process intelligible is the pursuit of certain rules (or habits) of cognition and interpretation, which supplies the impetus for the study of implicit cognition and perception.

There is, then, a way of studying organizations and organizing in which *rules rule*. They matter so deeply that they provide the very foundation for the meaning of speech acts (Wittgenstein, 1952), the recognition of patterns of organization, the successful coordination of communicative and strategic action and the design, planning and enactment of organizational tasks (Moldoveanu & Bauer, 2004). It may then, interest one, as a student of human action and organizational behavior, to come up with a representation of rules and rule-bound behavior that is at once general enough to accommodate the rich variety of rules and rule-based patterns of behavior and precise enough to allow us to consider detailed questions concerning rule design and its effects on macro-level organizational behavior.

The rule-based study of rule-bound systems

Stephen Wolfram (1985, 1993, 2002) suggested that *computation* can provide a powerful model for physical processes, and that *this* fact has important consequences for the predictability and tractability of problems in physics. He points to the following interesting relationship between computation and physical dynamics: “On one hand, theoretical models describe physical processes by computations that transform initial data according to algorithms representing physical laws. And, on the other hand, computers themselves (which we use to run the said computations—author’s note) are physical systems obeying physical laws” (Wolfram, 1985.) He explores the implications of this duality for the frequency of intractable or not-decidable-in-a-finite-time problems in physics, as follows:

1. First, he shows that computations can be modeled by the long-run evolution of cellular automata (CA), and therefore that predicting the evolution of a CA is equivalent to predicting the evolution of the dynamical system that the rule set embodied in the CA models;
2. Next, he argues that some calculations may be computationally irreducible, in the sense that the only way in which one can predict the evolution of the CA that corresponds to them is to follow the long-run evolution of that CA: there are, in these cases, no short cuts;
3. Finally, he shows that there are large classes of CAs (‘universal computers’, capable of simulating any other finite computational device of the same dimension) for which the problem of predicting the long-run evolution is intractable; moreover, such CAs are far more frequent than are CAs allowing ‘computational shortcuts’.

The ‘surprising insight’ is that, in the physical world, there are many more intractable problems than there are tractable ones. It seems at least *prima facie* interesting to ask: to what extent does any of this make a difference to the problems faced by organizational modelers? *This* in the previous sentence has, to be sure, *two* parts: the *first* has to do with the basic modeling move that Wolfram makes (physical systems as general purpose computers, their evolutions as computations, computations as states of a CA, CA as a generalized model for the evolution of a physical system). The *second* has to do with the issue of *complexity*: suppose organizations *could* be thought of as universal computers, what—if anything—could we say about the complexity of organizational phenomena understood as intermediate and final steps in (sometimes very lengthy) computations?

Let us take seriously the basic model of organizational states as states of a CA—with (potentially very complex) rule systems in place. In the ‘worst case’—one in which we do not agree on just what the ‘right’ rules that govern interactions should be—we can simply assume that the CA which models the organization is the CA corresponding to the total set of elementary particles in the organization, interacting according to a rule system that corresponds to the fundamental laws of physics. (As Wolfram is quick to point out (Wolfram, 1985), the modeler’s basic choice of ontology (‘particles’ versus ‘fields’ should not matter from the point of view of the validity of a CA model, as CAs can be used to model discrete-form versions of the partial differential equations that are used to model the space-time distribution and evolution of a field). It is not difficult to see that the problem of predicting the evolution of an organization over a significantly long period of—in general—intractable: even the two-body problem is intractable for certain kinetic energy levels (Casti, 1991). So, we have good reason to ask: why—and, when could we reasonably—harbor any confidence that we can *predict* anything at all about the evolution of an organization over periods of time that matter?

Wolfram’s CA-based approach offers both a sobering answer to such questions and the prospect of a framework that allows us to ask fundamental questions about predictability of organizational phenomena in the first place. In particular, if very large classes of formal CAs represent ‘universal computers’ and there is no reason to *ex ante* exclude them as valid models for organizations, then, the probability that a particular organizational phenomenon is computationally irreducible will *ex ante* be high. This means that, for a large class of organizational phenomena, *the best predictive model of the phenomenon is the phenomenon itself*. To figure out where a computationally irreducible rule set is going to ‘take’ a particular patterns, one has to

allow the evolution of that CA to ‘play out’ over whatever length of time one is interested in making predictions over. One might argue, however, that to do so means to abandon any predictive effort in the first place, because as *pre*-dicting a behavior can at best an event that is *simultaneous* to the behavior being predicted. One might also argue that most of the predictive ‘successes’ of social sciences are illusory, which is not an untenable argument given the false conflation of prediction with explanation that has colored the epistemology of the social sciences (see Friedman, 1953).

To all this one might object with a recitation of the many predictive successes that we experience everyday, relating to macroscopically observable but microscopically rule-bound patterns. For instance, ‘wanting’ to flip on a light knob can be understood as a desire coupled to a successful prediction (“if I extend my arm, catch the knob between my thumb and index finger, twist the knob, etc... then I will have turned on the light”) which relates to a potentially very complex pattern of behavior (all of the elementary particles making up the parts of my body engaged in the activity, all of the elementary particles that constitute the light and the knob, etc). *Countless* other examples can be constructed, and it is this countlessness that one usually rests arguments about the simplicity and ubiquity of predictive behavior on. However, as we have learned from Georg Cantor and his followers, *uncountability* comes in different classes, that can be distinguished according to their relative densities or measures. There are *uncountably many* natural numbers, for instance, in the sense that one could count them only by an endless counting process; whereas there are uncountably many real numbers in the sense that no mapping of these entities to a natural number counter is possible. Not surprisingly—and, relevantly—the *size* of the former class is *infinitesimally small* compared to the size of the latter. By a similar argument, it could well be that the class of predictable macroscopic patterns of behavior is infinitesimally small compared to the class of fundamentally unpredictable patterns of behavior (those modeled by computationally irreducible CAs). Given such an argument, the predictive successes of everyday life could be understood either as *very low probability accidents*—not likely, because of the very high reliability with which we can produce them—or—more likely—as carefully *contrived* scenarios in which we have learned to design predictive *short cuts*. It could be, for instance, that there are *many* possible paths by which all of the elementary particles in one’s body can interact with one another to produce the complex macroscopic behavior ‘turning on the light knob’, but, by a long process of design and feedback, the mind-body system has learned to constrain the set of initial conditions in a way that drives the CA which describes the entire system repeatedly to the same outcome. If this argument is even remotely plausible (and we have good reason to believe that it is), then, it makes more sense to be perplexed (and, as researchers, intrigued) by our predictive successes (how are they possible and when are they likely?) than by our predictive failures (if the world is described by law-like relationships between events, why can we not make more accurate and reliable predictions?). The science of ‘organizational complexity’, then, is turned upside down as its ‘organizing question’ changes from ‘whither complexity?’ to ‘whither simplicity?’. Why, when and how is it possible to make accurate predictions reliably, given that many the systems we make predictions about are likely to be computationally irreducible and thus the problem of predicting their evolution is likely to be uncomputable?’

The nature and structure of bounds to predictive intelligence

Of course, that is not how organization theorists usually think about the problem of complexity—just the opposite, in fact. They start from a (hypothetical) state in which the researcher or observer can make ‘perfect’ predictions about the evolution of the system under consideration and posit ‘sources of difficulties’ that he will encounter when trying to predict more accurately or more reliably; such as:

- *Bounded computational capabilities*. The principle of bounded computational capabilities has been a foundational principle of organizational research since Herbert Simon’s early work on bounded rationality. An agent (individual or organizational) is bounded in his or its ability to predict the future by the number of computations per unit time it can perform, and by the time available to perform these computations. In the case of computationally irreducible rule-bound systems, computational problems are particularly relevant, as the ideal predictor has to essentially build a system that has at least the computational complexity of the phenomenon that he is trying to make predictions about, and then (in order to predict), he has to ‘run the clock that measures ‘real time’” faster than the clock that clocks the phenomenon of interest;
- *Bounded informational capabilities*. Bounded rationality also has a second, *informational* component. Simon argues that upper bounds on short term memory (or short term memory utilization) are fundamental invariants of human behavior—and, by extension, of organizational behavior as well (Simon, 1990). Even if *local* rule systems could be designed with arbitrarily high complexity to alleviate the computational costs of using these rules to simulate macro-behavior, it is still the case that short-term memory bounds would limit the space in which these rules could be searched for. These ‘difficulties’—in the classical formulations of the problem of prediction—are what accounts for the predictive failures of the researcher or observer to instantiate a perfect simulation of Laplace’s demon—that figment of the 19th century imagination of (some) positivists who thought that knowledge of microscopic laws seamlessly translated into knowledge of the macroscopic patterns these laws give rise to.

By contrast, the ‘new science’ that Wolfram (2002) points to is called upon to explain predictive successes—rather than

predictive failures. To do so, however, we need a different foundational model of both 'organizations' and of (the process of) 'organization'. Moldoveanu and Bauer (2004) have proposed that a useful model for the organization from the point of view of studying 'organizational complexity' is a generalized computational device, such as a Universal Turing Machine (UTM). Organizational tasks are conceptualized as computational processes, with 'software' (or, the algorithms that 'organize' organizations) provided by the plans and strategies of managers or the models used by researchers to represent organizational phenomena, and the hardware provided by the physical embodiment of the organization. Tasks can be understood as the processes by which algorithms run on the underlying computational structure. They are states of the computational device that models organizations.

Now, by Wolfram's argument, the states of this computational device can be modeled by the instantaneous states of a CA, and thus, a CA can be used to model what organizations (*tout court*) do—and, ask: under what conditions can we make valid predictions about the long-run dynamics of organizational processes? Because there is a straightforward type-type reduction of organizational tasks to computational processes and of computational processes to the evolution of CAs, we can ask: what sorts of organizational tasks lend themselves to predictable CA instantiations, and, under what conditions can we expect to be able to make competent (accurate and reliable) predictions? Such a question can take many different forms. One can ask, for instance, a parallel question about the evolution of language (what are the attributes of words and concepts that 'cut the world at the joints?' (Nozick, 2002)—and model the rules by which we assign concepts to representations as the rules used to program the evolution of a CA; or, about the evolution of 'ways of knowing' (what are 'good discovery rules' that account for predictive success in a particular science?); or, about the evolution of reasoning more generally ('what is it in the nature of the received 'laws of thought' that makes logical inference, for instance, a valuable building block for the causal inferences that underscore predictive successes?). These problem statements can then be interpreted in the language of CAs in order to come up with representable—albeit potentially uncomputable—CA states that model the evolution of organizational patterns of behavior.

Cellular automata (CA), universality and computational equivalence

The computational modeling of organizational processes relies on the success of a series of reductions: of rule-based organizational phenomena to rule-based computational processes running on universal computational machines (such as Universal Turing Machines), of universal computational devices to CAs, and, quite often, of complex rule sets to simple rule sets. These reductions, in turn, allow us to reduce the 'complexity' of an organizational phenomenon to the rule-complexity, run-time-complexity and relative compressibility of the cellular automaton patterns that simulate that phenomenon, 'computation' writ large to the evolution of a cellular automaton embodying a particular rule set, the analysis of complexity to a typology of rule sets and of the evolved CA patterns associated with them, of macro-organizational processes to large scale cellular automaton patterns and of micro-organizational processes to local rule sets. As with any reductive strategy, we can ask:

Is the strategy general enough to capture phenomena of interest? The principle of computational representability (of in-principle computable functions) or allows us to represent any organizational process or phenomenon that can in principle be represented. It states that natural processes can be understood as embodiments of computational processes. It is reasonable to ask if any phenomenon can properly be understood as a computational process, and, of course, an answer is not easily forthcoming: like any modeling device, the principle of computational representability is not refutable, because it is a *paradigm* (Kuhn, 1962; 1990). It is, however, possible to inquire about the limits of the applicability of the move in particular cases.

A fundamental theorem of computation (Boolos & Jeffrey, 1993) states that, if a process is simulable at all, then it is simulable on a Universal Turing Machine. A representable phenomenon is one that can be modeled using functions mapping independent to dependent variables. These functions are computable: algorithms that compute them on a computational device converge to an answer after a (potentially large but finite) number of iterations—which can be thought of as point computations. Moreover, if a function is computable, then it is Turing-computable: it can be implemented on a Universal Turing Machine (Boolos & Jeffrey, 1993). Wolfram (2002) shows how to build CA representations of Turing machines, and thus how to reduce general-purpose computational processes to CA processes. Thus, by using the basic strategy shown in Figure 1, we can use this approach to reduce any organizational phenomenon that can in principle be symbolically represented to a set of CA processes.

Are the resulting findings objective, or, at least, inter-subjective, i.e., do they depend on a particular embodiment of the computational device used for simulation? The significant point about Universal Turing Machines is that they are the most general computational devices available. They can be used to simulate the operation of *any other computational device*—such as a cellular automaton. Thus, if a process is simulable or representable on a particular CPU architecture (a Pentium, say) using the local set of rules associated with programming it, then it can be simulated using the (irreducibly simple) instruction set of a Turing machine. Our strategy thus allows modelers using different computational devices and different simulation languages (such as 'theories of the organization', or 'modeling heuristics' or 'models' simpliciter) to reach agreement about important quantities (such as 'complexity') by translating their models into a more universal rule set and then comparing their findings.

Are the resulting distinctions, models and findings useful and relevant to the study of organizations? We argue that using the basic computational strategy outlined in this paper will allow us to draw important new distinctions in the study of organizations, such as those between the local complexity of the rules of interactions between individual agents and the rules of evolution of macro-organizational processes and the informational complexity of

Fig. 1: Figure 1

The process of representing organizational phenomena as computational phenomena

local rules and the computational complexity of the processes by which global patterns are produced from simple rules. It will allow us to produce models of organizational phenomena in which organizational complexity functions as both a bound and a driver or the evolution of organizational processes (Moldoveanu & Bauer, 2004). It will also allow us to generate explanations of novel phenomena such as the dynamics of rules in organizations (March *et al.*, 2000) which allow us to look not only at the differential birth and death rates of rules in organizations, but also at the content of rule systems, so as to be able to ask new empirical questions of the data, such as:

1. What kinds of rules are likely to proliferate in different environments?
2. What kinds of rules are survivable across a large set of different environmental and organizational settings?
3. What causal mechanisms relate changes in organizational behavior patterns to organizational rule sets?

We begin by attempting to establish simple rule-based cellular automata as the basic building blocks for models of organizational phenomena. We do this by showing how organizational phenomena can be understood in terms of individual agents interacting according to *locally simple rule sets*, how globally 'complex' or 'unfathomable' phenomena can be decomposed into parsimonious models that use simple, local rule sets, how the underlying local rule sets can be captured, generally and parsimoniously, by CA processes. The main import of the argument for 'doing organization science' is then outlined and amounts to a research strategy (or explanatory strategy) that can generate both deep explanations for currently acknowledged 'stylized facts' about organizations and auditable predictions of the evolution of complex organizational phenomena.

Representing rule-bound systems in terms of cellular automata

Cellular automata models are rule-based systems consisting of:

1. A set of interacting, local elements ('cells'), 1. each of which can take on one of N possible states or values ('colors' in many simulations using cellular automata);
2. A set of rules of interaction rules that pre2. scribe the state of each individual element as either a deterministic function of the color combinations of the neighboring elements, or a deterministic function of some statistical ensemble of the states of the neighboring elements, and;
3. A set of initial and boundary conditions for 3. the states or values of the interacting elements of the CA.

A CA model starts out by specifying the *rules* governing the local evolution of cells, and a set of initial conditions and boundary conditions for the two-dimensional space in which the cells are laid out. The basic strategy of the CA modeler is to start out with a macroscopic discernible pattern, to postulate a dynamical process in which the global pattern is generated through the interaction of individual elements using simple interaction rules and limited state-alphabets (for instance, the numbers of possible colors a particular cell can take on), to postulate a set of initial and boundary conditions (number of interacting cells of the CA, initial states of these elements) and to attempt to reproduce some salient aspect of the macroscopic pattern from knowledge of the micro-level rules, the initial conditions, and the applicable boundary conditions. 'Better' models are locally simple and explain more globally 'complex' stylized facts: they explain 'more with less' and thus run parallel to the 'stream of progress'—at least in some accepted views of scientific theorizing (Friedman, 1953), which attempt to make precise the norm, 'attempt to explain a lot by a little'—commonly known as 'Ockham's razor'.

How have CA models been used in organization studies, and how might they be used?

There already exists a rich literature that applies CA models to the explanation and prediction of organizational phenomena

(see, for instance, Lomi & Larsen, 2001). Researchers have produced sophisticated models of organizational behavior starting from postulates of bounded rationality and severe cognitive limitations at the level of individual agents making up the organization (modeled by a restricted set of local decision rules or heuristics), which allow them to model these agents using restricted state alphabets and simple, local, rule sets. By ‘simple’ one usually means (often without stating), admitting of low-Kolmogorov-complexity representations (Li & Vitanyi, 1993). Individual memory limitations—‘informational limitations’—are handled by the ‘small alphabet’ restriction and the restriction on the informational complexity of the rules of interaction, and individual computational limitations are handled by restricting the ‘search for rules’ to a space of computationally tractable ones. Individual agents, for instance, may be assumed to not be able—or, willing—to solve NP-hard problems (Moldoveanu & Bauer, 2004)—indeed, they are assumed to *conceptualize* their predicaments in terms of *P-hard* rather than *NP-hard* problems (Moldoveanu, 2006).

But, individual cells are not always *individual agents*—i.e., stylized models of humans—in CA modeling. For instance, McKelvey (1999), and Rivkin (2000) use CA models to represent coupled organizational activity sets. Organizational states are encoded using binary-state (‘logic gate’) N-tuples of elements that are, on average, mutually N-wise coupled, and explore the dynamics of such ‘nets’ by allowing them to evolve and measuring certain characteristics of their long-run dynamics (such as their period in the case of periodic dynamics).

CAs may also be used to ‘look inside’ the thinking of individual agents, whose local thinking and reasoning patterns are sometimes modeled using CAs. Rubinstein (1996) models the limited cognitive capacities of boundedly rational players engaged in iterated dominance reasoning to solve game-theoretic decision problems. The local rules of the CA device now act as propagators from a set of initial conditions (the ‘problem formulation’) to a ‘solution’ set that embodies both accuracy goals and time constraints. CAs can be set up to evaluate arbitrary logical expressions according to certain formulas and thus can act as propagators in the process of resolving a ‘well-structured problem’ (Simon, 1973); i.e., of proceeding, by self-evident steps (*modus ponens* and *modus tollens*) from a set of initial conditions, to an ‘answer’.

In view of these parallel developments, we have reason to ask: is there a ‘best, or, better way’ to proceed with the development of CA models in organizational research, such as:

1. *Downward* into a further decomposition of models of individual agents themselves as evolving complex entities?
2. *Upward* into models of larger organizational sub-units as interacting elements?
3. *Inward* towards more sophisticated models of individual cognition and reasoning?
4. *Sideways* into using CA elements to represent basic entities—such as ‘activities’, or, elementary behaviors—that can plausibly be conceived as being linked by deterministic interaction rules?

It is clear that CA models are sufficiently general to be used as representational devices for many different types of ‘ontologies’. Wolfram (2002) realizes this, and suggests that CA models are far more general than any one of these modeling applications. He constructs CA models for general computational processes and argues that any natural process—once understood as a computational process—can be simulated as a CA. (He focuses, to be sure, on deterministic, rather than quantum computation; and, the extension of his arguments to quantum computation is not straightforward, but, the basic computational framework for studying organizational processes *can* be extended to include quantum computers as the basic modeling devices).

Thus, it is not necessary for the organizational modeler to impose unrealistic modeling conditions on the evolving entities (as is often done by economic analyses of financial market behavior) simply in order to ‘fit’ the studied phenomenon into a CA lattice model. Rather, CAs can provide general enough model of any organizational process that is representable in algorithmic form (such as the form of a deterministic relationship between input and output, or independent and dependent variables). The modeler’s problem is transformed, from the problem of how to get organizational phenomena that are already agreed-upon by researchers as ontologically ‘valid’ to fit into the straitjacket of ‘cells’ and ‘evolution rules’ to choosing from a multitude of local rules and interacting entities that together can produce particular patterns.

Reduction of ‘complex’ rules, tasks and behaviors to ‘simple’ rule sets

It is possible for either the modeler or the organizational designer, to trade off between the size of the ‘alphabet’ that defines the number of possible states that a cell can take on (this defines the local, informational complexity of each interacting element) and the complexity of the local interaction rules (these define the computational sophistication of each interacting element).

Thus, it is possible, as a modeling (or design) assumption, to make trade-offs between the computational complexity of local rules and the informational complexity of local elements or cells. When the individual cells model individual agents within the organization, this exchangeability between computational and informational complexity becomes a critical capability for the

modeler, as it provides an explicit way to make trade-offs between the use of short-run memory and the use of computational power, a key design trade-off for any computational system.

Reduction of organizational phenomena to computationally simple, evolving local rule sets

If it is possible to reduce any organizational phenomenon of interest to a computational process, and that computational process to a CA model, then it becomes possible to reduce any organizational process that can be simulated to a corresponding CA model. CA models are structurally rich enough to allow us to tailor assumptions about individual elements to different cell-level alphabets and interaction rule sets of varying complexities. We can then apply CA models to the understanding of rule-based interactions between *individual agents* within organizations (in harmony with the dominant paradigm in the theory of organizations that has emerged from the writings of Simon (1947), Cyert (Cyert & March, 1963), March and Simon (1958), Newell (1990) and Nelson and Winter (1982)). Thus, we will focus on models that consider *individual agents* (rather than activity sets, or the neurons of the pre-frontal and dorso-lateral cortex of interacting agents, for instance), as the basic elements of CAs. In spite of this specialization of the argument of the paper, we note that it is possible (and often desirable) to apply CA decompositions and models to the understanding of intra-personal processes (involving the interaction of various motivational systems or of various computational 'nodes' that carry out calculations) or to that of inter-organizational processes (involving the characterization of whole organizations as 'cells' in a CA model); i.e., to think of a multi-resolution decomposition of organizational 'reality' (as seen through the lens of computable models thereof), with CA models applying at multiple levels of analysis.

Which rules emerge, and why do these rules emerge? Modeling and predicting the dynamics of organizational rules.

If we take seriously, as a modeling assumption, the picture of organizational behavior as macro-scopic behavior bound by rules that determine local agent-agent interactions, then individual behavior, on this view, can be understood as a purposive attempt to *correspond* or *conform* to a set of rules: of grammar and discourse ethics for modeling the production of speech acts and discursive behavior (Alexy, 1988); of social norms in the production of social behavior (March, 1994); of the rules of rationality in making decisions and implementing choices based on those decisions (Elster, 1982), of the Bayesian normative logic of epistemic rationality in the production of judgments about uncertain future events (Earman, 1992); of the rules of computation in producing models and simulations of ambiguous phenomena (Turing, 1950). The ability to construct CA models that effectively reduce any of these rule systems to a set of N-color (N-complex fundamental alphabets), K-neighbor (K-complex contingencies) rule sets allows us to speak of any organizational phenomenon as the instantiation of the macroscopic dynamics of such an underlying rule set.

An interesting question for the modeler, is: which rules will emerge as the best models of organizational phenomena? Why do these rules, rather than other rules, emerge over time? How and why do fundamental modeling rules change as a function of time?

The existence and importance of 'universal rules': rules that can replicate patterns generated by any other rules Wolfram (2002) notes that, of the 256 basic 2-color, 2-neighbor rule regimes there are some rules that can properly be called 'universal rules'. Universal rules are rules that, given enough time, can mimic the macroscopic pattern that is produced by *any other* single rule. They may require more computational effort to do so—they may, in other words, require that the rules in question be used to perform more computations than those required to achieve that pattern with the rule systems that they emulate; but they are guaranteed to produce the ensemble behavior (represented in a CA by the two-dimensional, N-color pattern of the CA lattice that emerges after M iterations of the fundamental CA algorithm) that is produced by the rules that they emulate.

There is a direct analogy between the existence of such universal rules and the existence of simple yet 'powerful' models of organizational behavior (such as rational choice-based models; or game-theoretic models) that can essentially model any organizational phenomenon. The 'no-fat' modeling approach of game theorists (Camerer, 1994) consists of (a) finding an organization-level or industry-level regularity (a macroscopic pattern, in the language of CAs), (b) positing a set of agent-level strategies, payoffs and conjectures (including conjectures about other agents and about other agents' conjectures) and (c) aggregating these agent-level micro-rules to simulate the evolution of the macroscopic phenomenon that is to be explained.

A "powerful" explanation of an organizational phenomenon is one that reduces macroscopic patterns to microscopic (agent-level, in this case) behavioral patterns (modeled as rules for behavior, or rules for choice, or rules for forming judgments, or rules for aggregating judgments in an actionable mind-set). Not surprisingly, game theoretic (more generally, rational choice-theoretic) explanation of multi-agent phenomena in organizations are popular (as both academic models and self-models, or social identities), because they seem to be able to explain anything. They can be thought of as 'universal rules': given enough computations, they will be able to explain macroscopic behavioral patterns in most organizational contexts. There is, then, a direct application of Wolfram's result regarding rule universality to the problem of predicting the kinds of rules that will emerge over time in explanations of organizational phenomena. Now, if we assume that explanatory models become interaction models—by a double hermeneutic by which individuals come to instantiate and follow as behavioral norms the rule sets they

use to explain their own behavior—then it is likely that universal rules—and the associated universal micro-models of agent behavior—will emerge as dominant rules of interaction because of their virtually unlimited applicability to a large variety of organizational scenarios and different organizational contexts. (This modeling assumption is not far-fetched, can be understood as a direct extension of Wittgenstein's (1953) 'no private language' argument.

Empirical studies of organizational rule systems (March *et al.*, 2000) have focused on empirical regularities that regard the incidence of rule births, rule deaths and rule transitions, as a function of changes in the environment of the organization, but, have not—as yet—focused on the structural and semantic features of the rules that emerge as dominant rules, or of the rules that regularly supersede other rules in the history of organizations. The concept of rule universality (together with the posited mechanism for the emergence of universal rules after repeated realizations of the organizational 'game' allows to make predictions about the substantive attributes of dominant rules: we look for the emergence of universal rules in organizational situations that reward the ability of organizational modelers and designers to create local rule sets with predictive capability over a large number of different situations (characterized by the macroscopic patterns of the resulting CA models).

Measuring the trade-off between global computational load and local informational depth in the selection of rules: a new characterization of organizational complexity. We can seek to understand the evolution of different sets of micro-rules within the organization as the embodiment or instantiation of trade-offs among four fundamental quantities that quantify and qualify both the local complexity of micro-rules and the global complexity of the evolving patterns:

1. **The informational or algorithmic complexity of the local rule sets.** This quantity (which we shall refer to as *local informational depth*) is the Kolmogorov complexity (see Li & Vitanyi, 1993) of the local CA rules, and can be defined as the length (in bits or M-ary units of information) of the minimum-length algorithm that can execute these rules. It is a measure of local 'random access memory'—or short term memory required by an element of a CA to act according to 'its' local rule system: if the agent's memory is not capable of supporting at least the number of states that the rule system that governs its local interactions refers to, then it will not be able to 'sustain' local interactions governed by that rule. If the CA element is an individual agent, then the local informational depth of the rules of the CA measures the short-term memory required in order for the agent to 'get along' with other agents.
2. **The computational complexity of the local rule sets.** This quantity measures the number of calculations required at the level of each CA element for the implementation of the local CA rules, as a function of the number of CA neighbors on which the state of the original CA element depends, and the number of possible states of the neighboring elements on which the state of each CA element depends. We shall refer to this quantity as the *local computational load* of the CA model (Moldoveanu & Bauer, 2003; see Cormen *et al.*, 1993 for a pedagogical introduction to computational complexity for algorithmic problems. For a 2-state CA model (each element can take on the value of '0' or '1') in which the state of each element depends linearly on the individual states of each of K neighbors, the local computational load for each CA element is bounded from above by 2K. Nonlinear dependencies of the state of one CA element on (including 'memory effects') will increase the number of operations required locally to compute the next state of the CA element from knowledge of the states of the neighboring elements. If each CA element is taken to model an agent (in a multi-agent model of an organization), then local computational load measures the computational power required of an agent to successfully interact with other agents according to the local rule sets.
3. **The global computational complexity of the resulting CA pattern produced by the iterative application of a set of micro-local rules or rule sets.** This quantity, which is defined as a function of the total number of variables (*CA elements x states*) that determine the evolution of local CA elements, measures how 'difficult' it is for a particular micro-rule set to replicate a particular large-scale pattern of the CA. It is a measure of the computational complexity of producing a simulation of an organizational phenomenon starting from a set of agent-level models (such as rational choice models or game-theoretic models). This quantity measures the relative efficiency of using micro-rule systems (organizational 'common law' and statutes is often written in terms of such rule systems) in order to predict the global evolution of the organization. It also measures the relative difficulty for the modeler of explaining an organizational phenomenon based on a metaphysical commitment to methodological individualism (i.e., to the explanation of social phenomena by reduction to individual-level phenomena) and a set of models of individual behavior.
4. **The global informational depth of the resulting CA pattern.** It is a measure of the relative 'compressibility' or 'reducibility' of the global pattern exhibited by the CA. Here, Wolfram's (2002) 4 complexity regimes are illuminating and can provide useful qualitative discriminators for the study of organizational phenomena. If, for instance, the global informational depth of a CA pattern is the pattern itself, then the pattern is incompressible. There is no 'short-cut' to producing the pattern, other than the iterative application of the local rule sets that produced it (regime 4) to the particular initial conditions of the CA. (Different initial conditions will produce different end results at the macroscopic level. Thus, regime 4 behavior has something in common with dynamical regimes characterized by sensitive dependence on initial conditions, such as chaotic dynamical systems). *Organizational example:* speculative bubbles (positive feedbacks in trading behavior which are not based on underlying, inter-subjectively agreeable 'facts of the matter') are structures or patterns that may or may not occur at particular space-time instances, depending on the kinds of information (i.e., initial conditions) that conditions the actions of the various traders. If, on the other hand, a macroscopic pattern is easily discerned and emerges from *any* initial conditions, then the CA model can be reduced to a set of rules that predict macroscopic evolution from microscopic rule systems without depending solely on the microscopic rule systems or varying with initial conditions (regime 1). *Organizational example:* Bertrand-Nash or Cournot-Nash equilibria, when they accurately represent oligopolistic behavior, exemplify the convergence of prices, from *any* initial conditions, to a set of prices that is predictable from knowledge of the producers' cost functions. Wolfram's 'in-between' regimes (2 and 3) are regimes in which there are macroscopic patterns (thus they are compressible) whose topological features (regime 3) or structural properties (regime 2) can be predicted via 'short cuts' from knowledge of the structure of the micro-rules via short-cuts (intermediate rules) whose implementation to simulate the macro-pattern is computationally lighter than the implementation of the micro-rules to simulate the same pattern. To use an intuitive organizational example for regime 2: Oscillatory behavior (of under-supply/over-supply) in feedback-regulated production systems (Sterman, 2000) is regular macroscopic behavior determined by microscopic rules of 'local rationality' that governs the behavior of individual agents; and here is one for regime 3: the topology of inter-organizational networks (Moldoveanu *et al.*, 2003) exhibits certain consistent regularities in situations in which network agents follow certain (micro-locally well-defined) strategies for sharing or withholding information, even though the precise structure of these networks (i.e., the positions, in the network lattice, where these topological structures will occur) cannot be specified in advance, as they sensitively depend on initial conditions. One may know that a particular network will evolve towards a large set of 'center-periphery sub-networks, without being able to figure out where,