



High-Fidelity Mathematical Models of Social Systems

**Joe Jeffrey
Northern Illinois University**

DRAFT

We present a comprehensive and systematic conceptual and mathematical framework with sufficient rigor and expressive power to support creation of models of social systems covering the entire range of phenomena in such systems, from individual actions to actions of groups at all levels, and encompassing all types of phenomena actually encountered in human systems: biological, psychological, economic, sociological, political, and cultural. The conceptual framework is formalized via entity specifications, a formalism incorporating certain concepts similar to frames, combined with technique of mathematical logic for representing relationships. The formalism is then used to derive a method for mathematically defining and calculating the similarity of any two social systems, and the complexity of a social system. This measure enables the mathematical definition of rate of social change and rate of growth of complexity of systems.

This paper presents a comprehensive and systematic conceptual framework and mathematical formalism with the capability of representing all aspects of the structure and dynamics of social systems, i.e., systems in which the actors are individual persons acting in organized, cohesive groups, ranging from families or small teams to nations and supra-national organizations. Actors may also include such cohesive groups themselves, and may include actors at multiple levels. By “high-fidelity” we mean that formal descriptions of the phenomena that are accurate, detailed, representations of all entire range of the situations, processes, and events in the system, at every level of detail.

The basis of the framework is the articulation of the concept of a community by Putman¹. That articulation is based on a more general formalism for describing large and complex entities – processes, objects, and states of affairs, due to P. G Ossorio². The formalism was designed to enable the formal description of reality at any level, i.e., giving descriptions of objects, states, and processes at any scale, and has been successfully used for that purpose in a number of software development efforts. The central intuition of the formalism is that “things” – objects, processes, or states of affairs – are specified by giving a formal name, the logically necessary immediate constituents of the thing, and the relations between them, in the same way that relations are specified in predicate

¹ A. O. Putman, “Communities,” in *Advances in Descriptive Psychology*, V. I, K. E. Davis ed., JAI Press, Greenwich, CT, 1981.

² P. G. Ossorio, “*What Actually Happens*”, University of South Carolina Press, Columbia, SC, 1978. Republished by Descriptive Psychology Press (www.descriptivepsychologypress.com), Ann Arbor, MI, 48104, in 2005.

calculate expressions, i.e., by formal name. Constituents themselves can have further elaboration in terms of constituents and relationships.

For example, nations make treaties governing trade; some of the constituents of the nations are commercial firms; some of the constituents of those firms are divisions; some of the constituents of those divisions are departments; and some of the constituents of those departments are individual human beings. A set of descriptions may comprise many levels, from the level at which the actors are nations and the processes are the kinds of processes described in ordinary language national actions, to the level at which the actors are individual persons engaging in individual actions. There is no limit to either upper or lower level actor; a set of descriptions may cover processes directly involving agents at any level, whether supra-national, biological, or at any level in between.

PRELIMINARIES: FORMAL DESCRIPTIONS OF REALITY AT MANY LEVELS

We begin with the mathematical formalism for representing “things” in the world, at any level or of any kind. We create a formalism responsive to the intuition described above a descriptions that are order pairs, the first component of the pair being the name and the second the description; the description consists of a list of the constituents (by name) and the relationships that must hold between them for the “thing” to be what it is.

The idea of formal name is the same as that used in mathematical logic: individuals, variables, and relationships are identified by identifier that carry no other information: “x”, “125,” “less,” etc. Examples include:

- agent123, Charlie, minimal_element, etc.
- less, above, next to, angle, etc.
- married(John, Mary)
- son(John, Akil) & sister(Akil, Holly)
- north(Canada, Illinois)
- etc.

In mathematical logic and computer programming, we commonly use names of individuals and relationships that remind a reader of the intuitive meaning of the names, but commonly used names have no special status, i.e., are simply formal names. Thus, $m(J, A)$ and $r1634(I279, I543)$ are equally valid formal names for the condition with the customary English-sounding formal name $\text{married}(\text{John}, \text{Mary})$

Nothing prevents the use of entire sentences as formal names, although that is not customary. In that case, although the sentence may carry a good deal of meaning to a reader of the description, it is still used as nothing more than a formal identifier, i.e., a string of letters over an alphabet. For example, the sentence, “The rise of inflation in 1920’s Germany led to the rise of National Socialism,” which would ordinarily be taken as a statement that could be true or false, may be used as the formal name of a state of

affairs; further information about that state of affairs would be given by a specification of its immediate constituents and their relationships.

We need to be able to give formal descriptions of three kinds of entity: objects (structures); processes (mechanisms), and states of affairs. State of affairs is the generalization of the concept of a state. Examples include a society with defined roles and social practices; physical mechanisms, which include component objects and physical processes; biological system such as a cell, which has components at several levels, and processes involving those components each level; and international diplomacy foreign relations, in which the entities are nation-states, NGOs, and supra-national organizations and companies, and processes involving them. (Since a state is a special case of a state of affairs, and "state" is the customary term, we will use "state of affairs" and "state" interchangeably in this paper.)

Processes have processes and objects as constituents: the steps of the process and the elements involved in it. Constituents of states of affairs may be processes, objects, or other states of affairs. The core of a formal description is thus a list of the immediate constituents and a list of the n-place relations among them, each specified by a formal name as in predicate calculus. There is no restriction on the kind of relationship or constituent that can be so represented – entities and relationships that cannot be defined or measured physically or mathematically can be formalized in this way.

1. ENTITY SPECIFICATION

An *entity specification* (*ES*) consists of an ordered pair (N, σ, D) , where:

- N is the (formal) name of the entity including, optionally, a list of alternate names and/or a numerical ID.
- σ is a list of formal identifiers of processes of which this entity is a constituent.
- D is a set of *paradigms*, the major varieties or descriptions of the entity. Celebrating a wedding anniversary (a custom in Western and other societies) has several paradigms, i.e., major varieties all recognizable as instances of the same social practice. In a different realm, DNA transcription has two major varieties, eucaryotic and procaryotic.

Each paradigm of D is an ordered triple (C, R, E) , where:

- $C = \{(C_i, T_i)\}$, in which C_i is the constituent and T_i is the constituent's classification, an element of the set $\{P, O, S\}$, representing "object," "process," or "state."
- R is the set of n-ary *relationships* that must hold between the named constituents. Any relationship may be included, not only those definable in terms of physical locations or quantities. Equations specifying quantitative relationships, including differential equations, are formal relationship names.
- The constituents and their relationships specify the structure of the entity. Additional information specifies particular instances of the entity. Identifying an actual instance requires the specification of which actual "things" (processes,

objects, and states of affairs) filling the roles named by the constituents. This information we term the *eligibilities* for the entity:

The eligibilities E are a set of ordered triples (c, i, r) , in which

- c is the name of the constituent;
- i is the name of the actual individual;
- r is the rule, or condition, under which i takes the role of c in this object.

Processes

Processes are multi-step changes in objects and configuration of objects; a process changes a state of the objects to a new state. In addition, processes may occur in many versions, i.e., combinations of the stages that are all ways of the specific process occurring.

Thus, the $\{(C_i, T_i)\}$ for a process include:

1. Two constituents that are the before-state and after-state.
2. A subset identifying stages, i.e., in which $T_i = P$. Some stages may be accomplished via two or more alternatives; these alternatives are included in this subset.
3. A subset identifying the object, i.e., $T_j = O$.
4. A subset identifying the versions of the process. Each of these constituents C_k is a state of affairs, i.e., $T_k = P$, and the constituents of C_k are stages.

Relationships between stages specify the time relationship between them: sequential, parallel, overlapping, interspersed, etc.

Objects

Objects have only object constituents, and in that sense are simpler than entities in general or processes; each constituent of an object of Type O.

Objects provide clear illustration of the use of multiple paradigms. An automobile can be described as having constituents {drive train, electrical system, fuel system, cooling system, exhaust system, suspension, frame, steering, passenger compartment}. It can also be described as having constituents {front half, rear half}. Similarly, the large subunit of a ribosome is commonly described as having a roughly spherical main body and three lobes (i.e., with 3 constituents), but it is also described as comprised of two rRNA chains (5s, 23s) and a number of proteins.

Relationship to frames

Those familiar with the concept of frames will notice that ESs are similar to frames. They differ in two important ways.

First, ESs are considerably more rigorously defined. The criterion for defining the constituents of an ES is that they are logically necessary for the thing being described to be what it is. By contrast, the criterion for defining a frame is only that it represent “things commonly found together³.” Frames do not distinguish between things that are necessary to the definition of the object, process, etc. being described and those that are commonly encountered, empirically. In addition, in actual use the criterion for a frame to have a slot is often weakened to be merely “is a related concept.”

Second, while frames are based on the recognition that “things” in the world come in “wholes” or “chunks,” and so the key idea is to identify a thing and its parts. Frames do not include specification of a set of relations $R_1 \dots R_m$ between constituents that define the thing, although some frame-bases systems allow specification of relationships. (Interestingly, while clearly a refinement of frames, Ossorio’s work predates the introduction of frames by several years⁴.)

“Incomplete” descriptions

Most situations and processes in the real world are too complex to be specified completely, i.e., down to the level of actions by individual persons. Suarez, for example, uses the example of a Mexican-American agent acting in a situation (state of affairs) involving the groups “Latin Americans,” the U. S., Spain, Europe, the European Union, Canada, and NAFTA. Each of these is an enormous group, and full descriptions would involve hundreds of millions of object, process, and state ESs, an obvious impossibility.

ESs are designed for handling incomplete specification. Each name of a constituent object, process, or state identifies a constituent. The constituent need not be, and in almost all cases will not be, atomic, i.e., not further describable in terms of constituents, but the possibility of further description of an entity *does not mean that a given set of entity specifications is not incomplete*, nor that the set cannot be used for simulation.

For example, the state of affairs named earlier, “The rise of inflation in 1920’s Germany led to the rise of National Socialism,” identifies an entity with two constituents: “the rise of inflation in 1920’s Germany,” “National Socialism,” and the relationship “led to the rise of.” Each of these constituents has further Descriptions in terms of constituents and relationships, and a set of descriptions down to the level of the famous image of a woman with a wheelbarrow of Deutsche marks to buy a loaf of bread would be enormous. It is not, however, necessary for simulation or analytical treatment.

2. DESCRIBING SOCIAL SYSTEMS

³ Minsky M (1975), “A Framework for Representing Knowledge”, in Winston P, ed., *The Psychology of Computer Vision*. New York: McGraw-Hill, NY.

⁴ P. G. Ossorio, *State of Affairs Systems: Theory and Technique for Automatic Fact Analysis*, RADC-TR-71-102, Rome Air Development Center, 1971.

Entity specifications may be used to apply to any description, simplistic or complex, naive or sophisticated, human or machine (biological or mechanical). Names of entities are formal, the constituents may be of any sort, and the relationships between them equally so. The name-and-description methodology does not require, and in fact was designed to avoid the necessity of, definition, because so many essentially human things are not definable in the traditional sense. Since we want a formalism for *social* systems, we present a conceptual framework that has been used successfully to articulate behavioral phenomena in a number of contexts, one in which the central concepts are human action and human communities. In Section 3 we formalize this framework with Entity Specifications.

To apply the general formulation of entity specification to social systems, we must first define social system. For this, we use particular formulations of human action as social practice in communities, immediately below. We then formalize these conceptualizations with entity specifications. The conc

Communities and Practices

Behavior

We use the concept of *intentional action*, formulated by P. G. Ossorio, to describe the actions of agents in social systems, i.e., human beings:

An *intentional action description* is an 8-tuple $\langle I, W, K, Kh, P, PC, A, S \rangle$ **(Formula 1)** where:

- I is the name of the actor
- W (want) is the goal of the action, the state of affairs to be achieved by engaging in the action.
- K (know) is an item-by-item list of the facts, concepts, or perspectives needed to engage in the action. Thus, to formalize Thus, to say, “X knows F,” F (the name of the item) is one of the set that is the value of this parameter in a particular instance.

The states of affairs identified in specifying an action may be for any kind of “size,” ranging from simple attributes such as size or color, or, large, complex, non-physical ones. For example, to say that a person enlisted in the Army to fight for his or her country is to say that the person is acting on the distinction of fighting for one’s country, and that their goal is that state of affairs. Similarly, a suicide bomber acts to bring about the (perceived) good of a religious or ethnic group.

It should be noted that here again we articulating something (concepts) that are more traditionally considered “mental constructs” or other, vaguer, terms. By “concept” we mean a distinction. When a person enlists in the Army, he or she is also acting on the concept named “Army” – i.e., the distinction of Army vs. other things.

- Kh (know-how) is the item-by-item list of the *skills* needed to do this behavior, that is, abilities to achieve results without engaging in a series of steps – i.e., or other, smaller, actions.
- P (process) is the actual process involved in the action – the “process aspect” of it
- A (achievement) is the actual result achieved. As with W and K, the values of this parameter may be any state of affairs.
- PC is the set of personal characteristics of the actor relevant to the action: the traits, attitudes, abilities, values, preferences of the individual. Each is identified by a name, and defined by the observable *pattern of choices* involved (rather than “internal states.”) “Value” means a state of affairs the individual will act to bring about (given an opportunity).
- S (significance) is the larger action this one is part of. The P parameter identifies the Process; the process is the set of steps that can be carried out to accomplish the action, and each step is itself an intentional action. Thus, commonly an action is a part of (a step in) carrying out another, larger, action. For example, in chess we may have the following:
 - Jie moves her pawn to K-4
 - By doing this, Jie is playing the Ruy-Lopez opening.
 - By doing *this*, Jie is playing the opening game in chess.
 - By doing *that*, Jie is playing a game of chess.

Playing a game of chess is (for chess players) *intrinsic* – i.e., done for its own sake, not because it is part of something else.

With the term “chess players” we come to the second major part of this conceptualization of human behavior: community.

Intentional action is the general case; a particular special case is all-important in describing human behavior: *deliberate action*. Deliberate action is the representation of choice in the presence of alternatives: the actor

- knows what they are doing, and
- want to do it.

Referring the W and K parameters above, this means one item is the following complex item: $(IA_k, \{IA_1, IA_2, \dots, IA_n\})$, that is, a specification of the set of actions from which the actor chooses, and the particular one chosen, IA_k .

To describe choice, however, we need to include a further concept: reasons for engaging in each of the potential actions $\{IA_1, IA_2, \dots, IA_n\}$. As with the W parameter, we treat conceptualize reasons as observable states of affairs with, tautologically, motivation power. Thus, for each IA_i , there is a set of order pairs (R_{ij}, w_{ij}) , in which R_{ij} is the motivationally significant state of affairs, and w_{ij} represents the “weight” of the reason.

Communities

Our interest, however, is not actions *per se*, but social systems. Practices – what is done in the system – are one important aspect of a social system, but all social scientists recognize that they are not the only aspect.

The most comprehensive formulation of the intuition of cohesive social groups – social systems – is the parametric formulation due to Putman⁵:

A **Community** is a 6-tuple $\langle M, P, Cp, S, C, W \rangle$ **(Formula 2)**
where:

M = Members
P = Practices
Cp = Choice Principles
S = Statuses
C = Concepts
W = World

- P (Practices) denotes the set of social practices – the actions – of the community. Practices encompass everything that a member of that community can do, *as* a member of that community; whatever they do will be engaging in one or more of the practices of the community.
- Cp (Choice principles) denotes the set of values or priorities specific to the community. The choice principles govern which of the available community practices are carried out, the priorities of the practices, and the particular way in which a practice is carried out. A value is a state of affairs that individuals act to achieve; a value of a community is a state of affairs that shared by the members.
- S (Statuses) denotes the recognizable positions in the Community. These are all the positions identifiable by members of that community, both formal and informal. “President”, “Senator”, “husband”, “child”, “suicide bomber”, “respected leader”, “doctor”, “farmer”, etc., are all names of positions in different communities. Central to the concept of a Status is that each Status has associated with it certain practices that are intrinsic to a person in that Status, i.e., are engaged in simply because the person has that position. A person in that status will therefore always have motivation to engage that practice. For example, a chess player (a person who has that position in the community of chess) has sufficient reason to engage in the practice of playing a game of chess.
- C (Concepts) denotes the distinctions acted on by the members of the community in one or more practices.
- W denotes the community’s World. This is not what it may at first seem, a list of all the items available to members. Rather, it is a specification (by formal name) of the single, all-encompassing “thing” that all of the objects, processes, and

⁵ Anthony O. Putman, “Communities,” in *Advances in Descriptive Psychology*, *ibid.*

concepts (distinctions) are part of. Thus, we hear of “the art world,” “the business world,” etc. These are worlds, of distinct communities, with these respective names.

Practices are actions; engaging in an action is engaging in a social practice. The intentional action parameters therefore are the parameters for specifying practices. However, the P (Process) parameter in the intentional action parameter is complex: specifying a process requires specifying $\langle \text{SO}, \text{E}, \text{G}, \text{C}, \text{V} \rangle$, where

- SO denotes the stages comprising a task analysis of the practice: the other, smaller, practices necessarily or optionally involved in carrying out this one. In some cases a task (a Stage) may be accomplished in more than one way; those alternatives are the Options.
- E denotes the *elements*, the logical roles, of persons or objects, in the practice
- G denotes the *eligibilities* of individuals for each role, specified by lists of individuals or formal rules
- C denotes the *constraints* on stages/options: formal specification the dependence of the occurrence of combinations of Stages/Options on either on another Stage/Option or on named states of affairs being the case
- V denotes the *versions* of the practice: the sets of Stage/Options that are considered, by members of this Community, to be valid instances of this practice.

Combining the process description above and the intentional action specification, we have the *social practice canonical form*: Each SPCF consists of a set of *paradigms* P_i , the major varieties of the action, and each where each P_i consists of

$\langle I, W, K, Kh, (\text{SO}, \text{E}, \text{G}, \text{C}, \text{V}), \text{PC}, \text{A}, \text{S} \rangle$ **(Formula 3)**

a single format that has proven useful in development of several knowledge bases covering a number of actions^{6,7}.

(The SPCF was originally devised by A. O. Putman to extend the Ossorio’s Basic Process Unit⁸.)

Each parameter of the SPCF is given by a formal, not merely discursive, name. The description formally encodes both the logical structure of the practice and the actual individuals involved in instantiations of it. What actually takes place is an instantiated version of a practice, that is, a specific set of Stage/Options recognized in the community

⁶ "MENTOR: Replicating the Functioning of an Organization", in *Advances in Descriptive Psychology*, Vol. III, pp. 243-270, K. E. Davis, ed., JAI Press, Greenwich, Connecticut, 1983.

⁷ "LDS/UCC: Intelligent Control of the Loan Documentation Process" (with T. Schmid, H.P. Zeiger, & A.O. Putman), *Proceedings of the Second International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems*, University of Tennessee Space Institute, Tullahoma, Tennessee, June, 1989, ACM Press, 1989, pp. 573-591.

⁸ Peter G. Ossorio, “*What Actually Happens*”, University of South Carolina Press, Columbia, SC, 1978, pp. 52-53.

as a way of engaging in this practice, with an actual individual person or object filling the logical roles (the Elements).

Specifying the parameters of a community, including its practices, this specifies the social system involved, completely (at that level of detail). What actually happens in that community is then governed by the following principles:

- Every Member occupies one or more Statuses in any community of which they are a member.
- Practices are everything there is to do in the community; every action in which a person engages, that is, everything a members does, when acting as a member of the community, is a social practice of the community. Every action, of every person, is an instance of engaging in one or more social practices of some community. (The limiting case of a community is the community whose members are all persons – the human community.)
- Each Status has one or more (typically very few) social practices that are *intrinsic* to that Status, i.e., practices engaged in by members that position simply by virtue of being in that position. Chess players play chess; mountain climbers climb mountains; agent-based modelers create agent-based models of systems. Pragmatically, the intrinsic practices provide the “fuel” for a system of human actors, that is, the actions that do not need any other “cause”, but rather are done by an actor in that status simply when there is an opportunity.
- The choice principles of the actor’s community function as priorities on the person’s actions. Choices by agents reflect by individual preferences (in the usual ordinary-language sense of that word) and priorities among actions reflecting community values; members of a community are expected to act in accordance with the community choice principles, and every culture (a particular kind of community) has practices that enforce their principles, i.e., practices members of the community engage in with respect to a member who has violated a principle. As a number of authors have pointed out, these community principles are quite significant and quite different from one community to another, leading to markedly different actions.

Example: A set of social practices descriptions of celebrating a wedding anniversary, illustrating further description of stages as practices. Only the stage-options are shown. Names illustrate the use of ordinary English sentences as formal names.

<p>Couple celebrates their wedding anniversary</p> <p>Paradigm 1: Couple buys gifts for each other</p> <p>Paradigm 2: Couple dines out at a nice restaurant</p> <p>Paradigm 3: Couple goes on cruise together</p> <p>Couple dines at a nice restaurant</p> <ol style="list-style-type: none"> 1. Couple goes to goes to restaurant <ol style="list-style-type: none"> a. (Option) By car b. (Option) By train c. (Option) Walking 2. Couple gets seated 3. (Optional) Couple examines menu 4. (Optional) Couple examines wine menu 5. (Optional) Couple orders wine 6. Couple orders food 7. Couple eats meal together 8. Couple pays 9. Couple departs restaurant <p>Couple eats meal together</p> <ol style="list-style-type: none"> 1. Husband eats meal 2. Wife eats meal 3. Husband and wife converse <p>Person eats meal</p> <ol style="list-style-type: none"> 1. (Optional) Person eats salad 2. (Optional) Person eats soup 3. Person eats main course 4. (Optional) Person eats dessert 5. (Optional) Person drinks wine

A *multi-community system (MCS)* is an entity in which at least two of the constituents are communities.

A *complex community system (CCS)* is an entity in which at least two of the constituents are communities *and* at least one member of a community is a member of at least two communities.

The community formulation provides the capability of formally describing all aspects of any human society, organization, company, team, or entire culture, at any level of detail. Describing a large organization, or a nation or culture, of course involves a very substantial amount of intellectual labor, and the hierarchical nature of the approach and the formalism are of great benefit in this regard, for descriptions need not be complete to be the basis of mathematical and software endeavors.

3. ENTITY DESCRIPTIONS OF PRACTICES AND COMMUNITIES

We have formulated the general case of behavior as $\langle I, W, K, Kh, P, PC, A, S \rangle$, and deliberate action, in which the person knows what they are doing and wants to do it, as the case in which one of the items in the specification of W and K is the pair $(IA_k, \{IA_1, IA_2, \dots, IA_n\})$

Looking carefully at the definition of each of the parameters, we note that we have:

- I : an object, specified by name
- W : a state of affairs, specified by name
- K : a set of state of affairs, each specified by name
- Kh : a state of affairs, specified by name
- P : a process, specified by name
- PC : a set of state of affairs, each specified by name
- A : a state of affairs, specified by name
- S : the large action: a state of affairs, specified by name

Therefore, each intentional action specification (including deliberate action specifications) is an entity specification: an action is an entity whose constituents are of the types specified.

In the case of communities, in the same way, each community is an entity, and each community specification is an entity specification, whose constituents are the Members, a set of object names; Practices, a set of state of entity descriptions; Choice principles, a set of state of affairs entities; Statuses, a set of states of affairs entities; Concepts, a set of state of affairs entities; and World, a single entity.

Communities and intentional action, stated in Entity Specification form, provide a mathematical representation of human behavior in the human context, at all levels. A social system is a formal entity consisting of (formal) immediate constituents, with n -ary relationships between them, and elaboration of constituent entities, at any level, via entity specifications.

4. MEASURING SIMILARITY AND COMPLEXITY

In order to simplify certain details of the definition below, we first define the *structural complexity* of an entity A , which has a Description with N constituents A_1, \dots, A_N and K relationships as

$$SC(A) = \sqrt{N^2 + K^2 + \epsilon \cdot \sum_{i=1}^{NA} SC(A_i)^2}$$

ε is an experimentally-determined multiplier modulating the impact of complexity of constituents, sub-constituents, etc. (Preliminary work indicates a value of approximately 0.7 for ε .)

We can now mathematically define the degree of similarity between any two entities – i.e. any two entities -- mechanisms, structures (objects), or states of affairs -- based on their constituents and relationships. The definition is designed to correspond to the following intuitions:

1. If we have two entities A and B, and the only available data for A and B is their properties (number of members, speed, per capita income, life expectancy, etc.), any measure of similarity must be based on the differing values of the attributes of the entities.
2. Structure among constituents is directly related to similarity of entities. The similarity measure should take into account how different the relationships between A's constituents and B's constituents are, in the following ways: a) If the A's constituents and B's constituents have the same relationships, but to different degrees, A and B should be more dis-similar than if the relationships had very similar values; if the A-constituents have more relationships than those of B, A and B should be more dis-similar; and if the A-constituents and B-constituents have different relationships, A and B should be more dis-similar.
3. When the constituents of A and B themselves have ESs, the measure should recursively include the structural similarity of the constituents.

Accordingly, we define the distance between two specifications as follows:

1. Let A and B be any two entity specifications. Let the constituents of A be A_1, \dots, A_{NA} with the K relationships r_1, \dots, r_K ; and let the constituents of B be B_1, \dots, B_{NB} with the L relationships r_{K+1}, \dots, r_L . Let the properties of A and B be p_1, \dots, p_M , and the relationships between A-constituents or B-constituents as r_k , where the properties and relationships are any properties and relationships of interest.

When constituents A_1, \dots, A_m have relationship r_j , the constituents form an ordered pair, and in general, for an n-ary relationship, the constituents that have the relationship are an n-tuple. Denote the number of A-tuples by NAT, and denote the number of B-tuples by NBT.

2. Let the value of property i of a constituent be represented by $p_i()$, and $r_i(t)$ denote the value of the ordered tuple t of A- or B-constituents satisfying relationship r_k . For example, in a family an important fact is the love relationships between each member (ignoring for the moment the varieties of love relationship). If there is a strong love relationship between family members A_1 and A_2 , we represent this by the notation $\text{loves}(A_1, A_2) = 0.9$ (on a 0 to 1 scale).

We create two matrices, P and R to represent the properties and relationships of A and B. Let P denote the matrix with M columns and N_A+N_B rows, whose values are the values of each property p_i .

P:

	p_1	...	p_M
A_1			
...			
A_{N_A}			
B_1			
...			
B_{N_B}			

The *property distance* between A_i and B_j is given by

$$PD(A, B) = \sqrt{\sum_{i=1}^M (p_k(A_i) - p_k(B_j))^2}$$

Let R be the matrix with L columns and $N_{AT}+N_{BT}$ rows whose entries are the values $r_k(t)$. If a constituent does not have property p_i , or a tuple does not have relationship r_k , leave the corresponding entry of the matrix blank.

R:

	r_1	...	r_K	r_{K+1}	...	r_L
A-tuple ₁						
...						
A-tuple _{N_{AT}}						
B-tuple ₁						
...						
B-tuple _{N_{BT}}						

3. If any column of P or R contains a value < 0 , re-scale the values of the column by adding the negative of the absolute value of the minimum value of the column to each value in it.
4. Normalize the values of P to the range 1 to 10, by setting

$$p_i(A_j) = 10 * (p_i(A_j) + 1) / p_{max_i}, \text{ where } p_{max_i} \text{ is the maximum value of column } i.$$

Normalize the values of the properties of A and B similarly.

5. Set each empty entry of P to 0.

The values of the property matrix P are now between 0 and 10, 0 indicating the component does not have the property of that column.

Similarly, normalize the values of R, the matrix with K+L columns representing relationship values for constituents of A and B, to the range 0 to 10.

6. When A and B have constituents, the similarity between A and B must include similarity of their constituents. Any such calculation is affected by the order of the constituents. For example, suppose we are comparing two organizations A and B, and A has a large and complex marketing department and a small, simple shipping department, while B has a large and complex shipping department and small, simple marketing department. The calculated similarity between A and B will be quite different, depending on whether the two marketing departments and two shipping departments are compared, or whether A's marketing department is compared to B's shipping department. Since the numbering of the indices of the constituents of A and B is arbitrary, the constituents of each must be re-ordered so that the distance comparison have a consistent basis.

Therefore, re-order the constituents of A and of B, from maximum SC (as defined above) to minimum.

7. The **distance between two entities A and B** is comprised of two components, the property distance and the structural distance:

$$d(A, B) = \sqrt{PD(A, B)^2 + SD(A, B)^2}$$

The *structural distance* SD(A, B) is defined recursively as follows:

Let MC = max(NA, NB) and MT = max(NAT, NBT).

Then if both A and B have Descriptions, i.e., specified constituents and relationships, then

$$SD(A, B) = \sqrt{(NA-NB)^2 + \sum_{i=1}^{MC} PD(A_i, B_{\kappa(i)})^2 + \sum_{j=1}^{MT} \sum_{i=1}^L (r_i(ta_j) - r_i(tb_{\kappa(j)}))^2 + \delta \bullet \sum_{i=1}^{MC} SD(A_i, B_i)^2}$$

where $B_{\kappa(i)}$ denotes the B-constituent closest to A_i , using Euclidean distance, and $tb_{\kappa(j)}$ denotes the B-tuple closest to ta_j , using Euclidean distance between tuples.

If $A \text{ NAT} > \text{NBT}$, $r_i(tb_j) = 0$ for $\text{NBT} < j \leq \text{NAT}$, and similarly if $\text{NBT} > \text{NAT}$.

If $NA > NB$, then $d(A_i, B_i) = SC(A_i)$, for $NB < i \leq NA$, and similarly if $NB > NA$.

If either A or B have no Description, $SD(A, B) = 0$.

δ is an experimentally-determined discount factor reflecting the relative importance of the distance between constituents of A and B. (As with ϵ , preliminary work indicates a value of approximately 0.7 for δ .)

$PD(A_i, B_i)$ measures similarity of properties of each pair of constituents;
 $\sum (r_i(ta_j) - r_i(tb_j))^2$ measures how much the constituents of A and B differ on relationship r_i ; and double sum

$\sum_{j=1}^{MT} \sum_{i=1}^L (r_i(ta_j) - r_i(tb_{\kappa(j)}))^2$ measures the total difference in structure between A and

B, as articulated by the specified relationships r_i , $1 \leq i \leq L$.

$d(A, B) = 0$ if A and B are the same except for differing only in names of constituents and relationships. The distance increases as their properties of A and B diverge, as the number of their constituents diverges, the properties of the constituents diverge, the structure of A and B diverge, and as the substructures of A and B diverge.

Any mathematical definition intended to capture an intuitive idea, such as shape similarity, must be validated experimentally. Work to verify the measure, and identify which of the class of algorithms and parameter settings is optimal, is in progress.

Example: Structural similarity of two families

Family A consists of a mother, father, and two children. The mother and father are married, and love each other. Both parents love both children; the children love each other. However, the children also compete with each other for success in school. Family A has an income of \$70,000, and is Catholic.

Family B consists of a mother, father, and three children. The mother and father are married. Both parents love all the children. The two younger children love each other,

but both resent the third and compete with her for the each parent's affection. (In the standard fashion, the spousal love relationship is distinguished from that of the parent-child love relationship and the sibling love relationship.) The eldest child also has a significant responsibility in caring for the younger children. Family B has an income of \$85,000 and is Presbyterian

We suppose that the member attributes of interest in this case are assumed to be age and health of the family members.

P:

	Age	Health
M_A	40	0.8
F_A	42	0.7
AC_1	12	1.0
AC_2	10	1.0
M_B	35	0.9
F_B	36	0.8
BC_1	8	1.0
BC_2	6	1.0
BC_3	14	1.0

R:

	Rom.. Love	Par. love	Sib. love	Aca. comp.	Resent	Affec. comp.	Care- taker
(M_A, F_A)	1.0						
(F_A, M_A)	1.0						
(M_A, AC_1)		1.0					
(M_A, AC_2)		1.0					
(F_A, AC_1)		1.0					
(F_A, AC_2)		1.0					
(AC_1, AC_2)			1.0	1.0			
(AC_2, AC_1)			1.0	1.0			
(M_B, F_B)	1.0						
(F_B, M_B)	1.0						
(M_B, BC_1)		1.0					
(M_B, BC_2)		1.0					
(M_B, BC_3)		1.0					
(F_B, BC_1)		1.0					
(F_B, BC_2)		1.0					
(F_B, BC_3)		1.0					
(BC_1, BC_2)			1.0				
(BC_2, BC_1)			1.0				
(BC_1, BC_3)					1.0		
(BC_2, BC_3)					1.0		
(BC_1, BC_3)						1.0	
(BC_2, BC_3)						1.0	
(BC_3, BC_1)							1.0
(BC_3, BC_2)							1.0

The normalized values of the properties of the families A and B are (8.2, 10, 10) and (10, 10, 10), so $PD(A, B) = \sqrt{1.8^2 + (10-10)^2 + (10-10)^2} = 1.8$.

$$(NA-NB)^2 = (5-3)^2 = 4.$$

The normalized property matrix P, with rows re-ordered so that the pairs $A_i, B_{\kappa(i)}$ are adjacent, is

	Age	Health
M_A	9.5	8
F_B	8.6	8
F_A	10.0	7
M_B	8.3	9
AC_1	2.9	10
BC_3	3.3	10
AC_2	2.4	10
BC_1	1.9	10
BC_2	1.4	10

and the constituent property distance $\sum_{i=1}^5 PD(A_i, B_{\kappa(i)})^2 = 0.81+2.89+0.16+0.25+1.96$
 $= 6.07$

The normalized relationship matrix R, with rows re-ordered so that the pairs ta_j and the nearest tuple $tb_{\kappa(j)}$ are adjacent, is

	Rom.. Love	Par. love	Sib. love	Aca. comp.	Resent	Affec. comp.	Care- taker
1 (M _A , F _A)	10	0	0	0	0	0	0
2 (M _B , F _B)	10	0	0	0	0	0	0
3 (F _A , M _A)	10	0	0	0	0	0	0
4 (F _B , M _B)	10	0	0	0	0	0	0
5 (M _A , AC ₁)	0	10	0	0	0	0	0
6 (M _B , BC ₁)	0	10	0	0	0	0	0
7 (M _A , AC ₂)	0	10	0	0	0	0	0
8 (M _B , BC ₂)	0	10	0	0	0	0	0
9 (F _A , AC ₁)	0	10	0	0	0	0	0
10 (M _B , BC ₃)	0	10	0	0	0	0	0
11 (F _A , AC ₂)	0	10	0	0	0	0	0
12 (F _B , BC ₁)	0	10	0	0	0	0	0
13 (AC ₁ , AC ₂)	0	0	10	10	0	0	0
14 (BC ₁ , BC ₂)	0	0	10	0	0	0	0
15 (AC ₂ , AC ₁)	0	0	10	10	0	0	0
16 (BC ₂ , BC ₁)	0	0	10	0	0	0	0
17 (F _B , BC ₂)	0	10	0	0	0	0	0
18 (BC ₁ , BC ₃)	0	0	0	0	10	0	0
19 (F _B , BC ₃)	0	10	0	0	0	0	0
20 (BC ₂ , BC ₃)	0	0	0	0	10	0	0
21 (BC ₁ , BC ₃)	0	0	0	0	0	10	0
22 (BC ₂ , BC ₃)	0	0	0	0	0	10	0
23 (BC ₃ , BC ₁)	0	0	0	0	0	0	10
24 (BC ₃ , BC ₂)	0	0	0	0	0	0	10

(We have numbered the rows, for ease of reference.) The first 6 pairs of tuples have identical value; the next 2 pairs are identical in all but one column (relationship); and the bottom 8 tuples have no matching A- tuple, so the sum

$$MT L \sum_{j=1} \sum_{i=1} (r_i(ta_j) - r_i(tb_{k(j)}))^2 = \sqrt{6 \cdot 0^2 + 2 \cdot 10^2 + 8 \cdot 10^2} = 31.6$$

In this example, we are considering entities whose only immediate constituents are individual persons. Customarily one considers persons to be indivisible, and so for the purposes of this example $SD(A_i, B_i) = 0$. (However, if one has a case in which it is appropriate to assign a measure of personality difference, such as differing scores on a range of psychological tests to the individual persons, that difference would be appropriate as a measure of $SD(A_i, B_i)$, and A_i and B_i are persons.)

$$\text{Therefore } SD(A, B) = \sqrt{4 + 6.07 + 31.6} \text{ and } d(A, B) = \sqrt{1.8^2 + 4 + 6.07 + 31.6} = 6.7$$

Consider now the distance between A and B', a family identical to B except that B'C₁ and B'C₂ did not resent and compete for affection with B'C₃. Rows 18, 20, 21, and 22 would be missing, so the sum

$$\sum_{j=1}^{MT} \sum_{i=1}^L (r_i(ta_j) - r_i(tb_{\kappa(j)}))^2 = \sqrt{6 \cdot 0^2 + 2 \cdot 10^2 + 4 \cdot 10^2} = 24.5,$$

$$SD(A, B') = \sqrt{4 + 6.07 + 24.5}, \text{ and } d(A, B') = \sqrt{1.8^2 + 4 + 6.07 + 24.5} = 6.1$$

Rate of change

Since we can now calculate the complexity of any social system, with a measure that incorporates size, difference in properties, difference in structure, and difference in structure of constituents, at all levels of detail, we can now define the rate of change of a society, in the obvious way. Let S₁ denote society S at time t₁, and S₂ denote S at time t₂.

Then the **rate of social change** of S is $\frac{d(S_1, S_2)}{t_2 - t_1}$.

Similarly, the **rate of growth of complexity** of S given by $\frac{SC(S_2) - SC(S_1)}{t_2 - t_1}$.

References

"MENTOR: Replicating the Functioning of an Organization", in *Advances in Descriptive Psychology*, Vol. III, pp. 243-270, K. E. Davis, ed., JAI Press, Greenwich, Connecticut, 1983.

"LDS/UCC: Intelligent Control of the Loan Documentation Process" (with T. Schmid, H.P. Zeiger, & A.O. Putman), *Proceedings of the Second International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems*, University of Tennessee Space Institute, Tullahoma, Tennessee, June, 1989, ACM Press, 1989, pp. 573-591.

Minsky M (1975), "A Framework for Representing Knowledge", in Winston P, ed., *The Psychology of Computer Vision*. New York: McGraw-Hill, NY.

P. G. Ossorio, "*What Actually Happens*", University of South Carolina Press, Columbia, SC, 1978. Republished in by Descriptive Psychology Press (www.descriptivepsychologypress.com), Ann Arbor, MI, 48104.

P. G. Ossorio, *State of Affairs Systems: Theory and Technique for Automatic Fact Analysis*, RADC-TR-71-102, Rome Air Development Center, 1971.

A. O. Putman, "Communities," in *Advances in Descriptive Psychology*, V. I, K. E. Davis ed., JAI Press, Greenwich, CT, 1981.