Structure

H. Joel Jeffrey, Ph.D.
Northern Illinois University

Abstract

The concept of structure, and the related ones of structural complexity and similarity, are ubiquitous in the sciences, arts, and literature. While they are used routinely and to good effect to gain insight into a very wide range of phenomena, they have never been rigorously defined. Beginning with a unification of Ossorio's Process, Object, Event, and State of Affairs Units into a single formal Aspect Specification, this article presents a mathematical definition of structure and structural similarity applicable to any aspect of the world—object, process, event, or state of affairs—and a mathematical quantification of structural similarity equally widely applicable. Intentional and deliberate action and communities, core concepts of Descriptive Psychology, are formalized with Aspect Specifications, and Aspect Specifications of actual objects and processes are given. Examples illustrating the calculation of the structural similarity of disparate kinds of things in the world, ranging from human families to intra-cellular organelles, are given.

Unification of the Descriptive Units

In his seminal work addressing scientific and conceptual issues in describing the real world, Ossorio (1971/1975/1978/2005) presents a formal system of four categories of what there are in the world and the logical relationships between them. The four categories are object, process, event, and state of affairs. As is always the case with fundamental concepts, the four are defined in terms of each other via the single logical transformation "is the same as" in a series of re-write rules comprising the State of Affairs System (SAS). The concepts are then used to develop descriptive formats—the Object, Process, Event, and State of Affairs Units (OU, BPU, EU, and SAU, respectively)—of sufficient expressive power that they may be used to specify any part or aspect of the world. The Object Unit, for example, specifies exactly what information must be given to completely specify an object, at any level of detail: a pencil, a car, a computer, a human body. Each Unit is a parametric formulation of the ways in which that kind of thing—object, process, event, or state of affairs—can vary.

One of the most significant facts about the SAS and the descriptive Units is that they embody no assumptions about what is most basic or fundamental to the world or, more generally, to a world. In particular, the objects, processes, events, and states of affairs need not be physical, as is the case with the customary description of the world found in the physical sciences: fundamental particles comprising atoms, which comprise molecules, etc. The descriptive Units can be used and have, in fact, been used to describe various aspects of chess, banking, organizational management, marketing, and biology.

In this article we use the State of Affairs System (particularly the descriptive Units) to develop a new formulation of an important and extremely widely-used concept, that of structure. Based on this formulation, we develop novel mathematical formulations of two related concepts, structural complexity and structural similarity, which, for the first time, allow precise definition and quantification of the concepts. Because they are based on the formal concepts of object, process, and state of affairs, rather than on any more traditional physicalist or reductionist formulation, the mathematical formulations are directly applicable to definition and quantification of structure, complexity, and similarity in the entire range of phenomena where those concepts are used, ranging literally from physics to poetry and literature. To illustrate this applicability, a range of examples, from basic chemistry to human families, are presented.

To develop these formulations, we must first address a certain difficulty with the descriptive Units as they stand. One of the basic logical facts about the four reality concepts of object, processes, events, and states of affairs is their inter-convertibility, i.e., the same thing can be described as an object, process, event or state of affairs (Ossorio 2005). World War II, for example, can be described as one complex object, event, process, or state of affairs. Accordingly, Ossorio states that the Object, Process, and Event Units may be converted into one another by first converting them into State of Affairs Units (Ossorio, 2005, p. 61). However, examining the respective Units, it is not at all clear how this could be done. The Units have different forms and different parts. The Process Unit, for example, contains Stages and Options, while the Object Unit contains Constituents that are Objects. Because it is sufficiently unclear how one might convert Object descriptions to Process descriptions, the convertibility itself is unclear. This is more than a practical or esthetic issue, for if the formats are not interconvertible then either the logical convertibility codified in the SAS is incorrect, or the formats are incorrect.

We present a formalization of the Object, Process, Event, and State of Affairs Units that unifies the descriptive formats into a single format and completely clarifies the logical equivalence of the Units and their interconvertibility. (Since the Units are already formal, this formalization might most appropriately be termed a re-formalization.) Devising a new formalism to describe a range of already-formalized objects is common practice in mathematics, where it is done for two purposes: to highlight formal similarities and to provide a basis for new insights. A classic example is group theory, in which many different objects and operations on groups are described with a uniform formalism that clarifies similarities of structure irrespective of differences of the particular objects. The formulation presented here has similar purposes: it provides a means of specifying a world (or any part of one) in a uniform format that highlights the formal structure of the "thing" regardless of whether it is an object, process, event, or state of affairs, thereby enabling precise comparison and analysis of various aspects of the world when those operations would otherwise be difficult or impossible. Intriguingly, the two things need not be of the same kind at all. We can, for example, rigorously quantify the similarity between a biological organism and a human society; between a machine and an organization; or, more abstrusely, between a building (an object) and flapping flight (a process). Analogies and metaphors of this sort are extremely common, and this formulation makes it possible use them as a rigorous, quantitative tool.

We then use the formalization to develop a new mathematical formulation of the concepts of structure, structural complexity, and similarity of structure. Because we are beginning with Ossorio's unique conceptual analysis, the formulation provides scientific capabilities not previously available: mathematically rigorous, quantifiable formulations of the similarity and complexity of any parts or aspects of the world.

Ossorio (1971/1975/1978/2005) discusses at some length the SAS as a formal system in the sense that concepts of object, process, event, state of affairs, and relationship are formal, meaning they are defined solely by their logical relationships with each other, much as "point", "line", and "plane" are defined in the discipline of plane geometry. SAS Rule 1, for example, states, "A state of affairs is a totality of related objects and/or processes and/or events and/or states of affairs." The phrases "object", "process", "event", and "state of affairs" are formal identifiers of concepts defined by this and the other SAS Rules, not references to something outside the SAS, empirical generalizations, etc. As elsewhere in Descriptive Psychology, Ossorio deliberately uses the customary terms "object", "process", "event", and "state of affairs" in part to avoid the appearance of inventing something unlike our existing concepts. As well-chosen as this strategy was for the purpose of articulating

Descriptive Psychology and making clear that it is an articulation of concepts already shared by persons, it has a significant drawback when used for other purposes, such as presenting the system to those who do not have a background in Descriptive Psychology. Specifically, because the terms "object", "process", etc. are ordinary English, the SAS is virtually always misunderstood as a set of arbitrary definitions or empirical generalizations rather than as a formal system. In light of these considerations, the formal unification of the descriptive Units is presented in a more traditional mathematical notation.

Examining the BPU, OU, and SAU, we find that each contains four kinds of specification: 1) parts or constituents; 2) relationships between the constituents (attributes being 1-place relationships); 3) contingencies that specify which constituents may occur under various conditions, including occurrence of other constituents in this or any state of affairs, object, or process; and 4) eligibilities, i.e., specification of which actual individuals are eligible for each constituent. Each constituent, relationship, and individual is specified by a formal name, that is, a name that serves merely to distinguish it from others rather than to define it in some way. For example, as we will see in more detail below, one relationship that is part of a paradigm case (Western) family is that the husband and wife love each other, a distinction that can be represented by Love(H, W) and Love(W, H), the standard device in mathematical logic. Love(H, W) is a formal identifier, not a definition; we are not taking on the task of "defining" love mathematically. It could equally be identified by the name "L", "R121", or any number of other formal names.

One of the central concepts implicit in the SAS, and explicit in the descriptive Units, is that specification is done at a chosen level of detail; further detail is specified via further Object, Process, Event, or State of Affairs Units. A description is complete when it fully specifies the object, process, etc., at that level of detail; that is, it specifies all the constituents, contingencies, relationships, and eligibilities necessary for the "thing" to be *that* thing and not some other. Completeness is not related to having lower-level details. A description may be complete and correct but have little in the way of detail specifying the structure of any constituent. For example, we might have complete specification of the practice of "motivating a subordinate to improve their performance" that includes, as a stage, the practice "assess the subordinate's intrinsic motivations" but have no further specification of that assessment practice. Similarly, we might have a specification of an automobile engine with constituents of engine block, valve system, air system, ignition system, cooling system, exhaust system, lubrication system, and power delivery system; and with the various relationships between these constituents; but no further specification of, e.g., the valve system; therefore without any mention of a valve or camshaft, which would be present in a specification of the valve system itself. In short, to say that a description is complete is not to imply that that no further detail can be specified.

In summary, a specification of an object, process, event, or state of affairs via the respective descriptive Unit consists of a specification of the constituents of the object, process, event, or state of affairs; relationships between constituents; contingency rules governing occurrence of constituents; and eligibilities; at a particular level of detail in every case; by formal name.

Aspect Specifications

In developing a single formalism that unifies the formal descriptive Units for objects, processes, events, and states of affairs, it is convenient to have a "cover term" for the four kinds of things, to ease the exposition. In our culture (and, to our knowledge, most cultures), there is a strong tradition of considering the world to be an object and parts of it to be sub-objects. This is, as discussed extensively in (Ossorio, 2005), a logical error, and terms such as "thing" or "entity", carrying the connotation of "object" that they do, are almost unavoidably misleading. A world, and our world, is a single state of affairs that incorporates all other states of affairs, objects, processes, and events (Ossorio, 2005, p. 29). We therefore adopt the term "aspect" as our cover term for any "thing" in the world, whether object, process, event, or state of affairs; and refer to the unification to be developed as Aspect Specification.

To ease the transition to a presentation that is necessarily rather mathematical in style, we introduce a simple example of describing something in the world: an ordinary, paradigm case, kitchen chair. A paradigm case chair consists of a seat, a back, and four legs. However, it is not sufficient to have the correct parts; they must be arranged in certain ways to be a chair. The seat must be attached to the legs, as must the back. The back and legs must be approximately perpendicular to the seat. (We will see below how "approximately perpendicular" is handled formally.) The legs and back must be on opposite sides of the seat. In addition, the parts must have certain properties, such as strength and rigidity. There are also properties of the chair itself, such as having sufficient strength to support an ordinary human: an object like a chair, with parts having the necessary attributes, but arranged in such a way that it collapsed under the weight of a pencil would not ordinarily be called a chair and could not be treated as a chair. Finally, there are other conditions regarding the parts of the chair that must be met, such as the requirements that the seat, back, and legs not be alive.

It should be noted that we are not attempting to "define" the concept of chair, in the traditional mathematical sense, but rather to describe what would ordinarily be called "one ordinary kind of chair"—a paradigm case of chair, one about which it would be said, "If ever there were a chair, this is one."

So far we have described *a* chair, but not an actual thing in the world, an actual chair. To this point, three kinds of things have been included in the description: a list of necessary parts or constituents, a list of relationships that must hold between the parts, and a list of limits or constraints on properties of the constituents and of the chair itself. To describe a particular chair, we identify the specific physical objects that are the legs, seat, and back; that is, that fill the roles of leg, seat, and back. To specify that, we must identify the actual things and say which ones can fill which roles, i.e., are eligible to be each constituent. For example, if we switch two of the legs of a chair, we consider the object to be the same chair.

In summary, to give a description of the chair, one must specify its parts, their relationships, constraints on the parts, and the eligibilities of the actual individual things to be each part. Re-stated in mathematical form, these four parts are the core of the unification of the descriptive Units. Figure 1 presents this mathematical form, which we term *Aspect Specification.*

In more detail, the Aspect Specification consists of an ordered triple (N, T, D), where:

- N is the (formal) name of the aspect including, optionally, a list of alternate names. These names may be any identifying locution, including but not limited to words, phrases, entire sentences, paragraphs, or numerical or symbolic codes.
- T is an element of the set {O, P, V, S}, representing classification of the aspect as Object, Process, Event, or State of Affairs.
- D is a set of *paradigms*, the major varieties or descriptions of the aspect of the world. We often have multiple descriptions of some object, process, etc., and there are often multiple varieties of what is recognizably the same thing. In addition, it is often desirable to specify alternate descriptions due to the state of knowledge of the phenomenon: conjectures, possible alternative mechanisms, etc. The paradigms are the distinct descriptions of the aspect. Each paradigm of D is an ordered 4-tuple (C, CR, G, E), where:
- C = $\{(C_{i,} T_i)\}$, in which C_i are the constituents and T_i is each constituent's classification, an element of the set {O, P, V, S}, representing "object", , "process", "event", or "state of affairs." Each constituent is specified by a formal name, i.e., an identifier that distinguishes the C_i. As discussed in Ossorio (1971/1975/1978/2005), the names are any identifying locution: mathematical symbols, words, phrases, or sentences from a human language, etc.

The Aspect Specification (AS) is an ordered triple (N, T, D), in which:

- N is the formal name of the aspect and, optionally, a list of alternate (formal) names.
- T is the classification of the aspect as classification as Object, Process, Event, or State of Affairs
- D is a set of *paradigms*, the major varieties or descriptions of this aspect of the world. Each paradigm consists of:
  - o $C = \{(C_i, T_i)\}$, in which C_i are the constituents and T_i is each constituent's classification as Object, Process, Event, or State of Affairs, each C_i specified by formal name.
  - o $CR = \{CR_i\}$ is the set of n-ary relationships that must hold between the C_i, specified by formal name.
  - o $G = \{G_i\}$ is the set of contingencies governing the occurrence of a C_j, each G_i being an ordered triple (C_j, GR_i, OC_m) specifying the constituent, the other constituent (of some Aspect) on which the occurrence of C_j is contingent, and the relationship between the two constituents.
  - o $E = \{(C_i, \{(I_{ij}, ER_{ij}, OC_{ij})\}\}$. For each constituent C_i there is some set of individuals that may serve as that C_i, each governed by a rule in terms of a relationship and some other constituent:
    - ▪ I_{ij} is the set of actual individuals, specified by formal name;
    - ▪ ER_{ij} is name of the set of relationships governing the eligibility of each individual I_j for constituent C_i
    - ▪ OC_{ij} is the constituent (of some Paradigm of some Aspect Specification) upon which the eligibility depends.

Figure 1. The Aspect Specification

- $CR = \{CR_j\}$ 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, and not only those definable mathematically. In the chair example, we noted that the legs must be approximately perpendicular to the seat; the phrase "approximately perpendicular" identifies a relationship, one that is difficult to define mathematically but is routinely recognized and acted on by persons. As with constituents, relationships are specified by formal name: R33, mother-of, etc., the representational device used in mathematical logic. The relationship between the legs and seat of a chair can be named with the phrase "approximately perpendicular", or with a style of name that is commonplace in computer programming, approximatelyPerpendicular. Equations specifying quantitative relationships, including differential equations, are formal relationship names. As is customary in mathematics, a property or attribute is a one-place relationship.

- $G = \{G_i\}$ is the set of contingencies or conditions on the occurrence of a Named constituent. Contingencies may be quite complex conceptually, for the occurrence of a constituent may be contingent on the presence of a named constituent of a Paradigm of any other aspect—object, process, event, or state of affairs. For example, it is not uncommon for the occurrence of a stage of a process to be contingent on the occurrence of a stage of an entirely separate, otherwise unrelated, process. Formally, though, specifying the contingency requires only specifying the constituent whose occurrence is contingent, the constituent (of some Aspect) upon which the occurrence is contingent, and the relationship between the two, in each case by formal name. Thus, each G_i is an ordered triple (C_j, GR_k, OC_m). (The constituent OC_m may itself have many constituents and relationships involving them.)

The relationships CR_j characterize the "arrangement" of the Constituents—physical, temporal, logical, behavioral, or any sort, that is, the configuration of Constituents that must be the case for this to be a case of aspect A. By contrast, contingencies are specifications of which Named constituents may occur, depending on the presence of some other constituent of some aspect, and are the means for specifying further restrictions on what can occur and still be a case of aspect N. (Ossorio, 2005, p. 63 and p. 43). Their function in the AS is thus to narrow the range of allowable configurations of constituents, not to specify the configurations themselves.

- The constituents and their relationships specify the structure of the aspect. Additionally, as discussed in (Ossorio, 2005) and as we saw with the chair, one must specify which actual "things" (processes, objects, events, and states of affairs) may or must fill the roles named by the constituents. This eligibility information includes the constituent C_i, the actual individuals that serve as, or take the part of, C_i, and any rule governing the eligibility of an individual to be C_i. This rule, as with contingencies, is specified by naming the relationship and the constituent (of this or another Aspect) on which the eligibility depends. Thus the eligibilities are denoted by a set $E = \{(C_i, \{(I_{ij}, ER_{ij}, OC_{ij})\}\}$. in which, for each C_i,

◊ I_{ij} is the set of actual individuals eligible for constituent C_{ij}

◊ ER_{ij} is (the name of) the relationship governing whether I_j can be C_i

◊ OC_{ij} is the constituent (of some Paradigm of some Aspect Specification) upon which the eligibility depends.

As noted, the AS is not a new conceptual formulation, but a restatement in a different formalism of the descriptive Units given in (Ossorio, 2005). To that end, we show the form the Aspect Specification takes when used to specify an object, process, or state of affairs.

Specifying aspects that are processes

A process is a change form one state of affairs to another with at least one intermediate state of affairs (Ossorio, 2005, p. 38). The states of affairs commonly, but not necessarily, involve objects and their relationships. Processes may occur in many versions, i.e., combinations of the stages.

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

1. Two constituents, specifying the before-state and after-state.
2. A subset identifying stages, i.e., constituents C_j in which $T_j = P$. Some stages may be accomplished via two or more alternatives; these alternatives are included in this subset.
3. A subset identifying the elements, i.e., $T_j = O$ or S

4. A subset identifying the versions of the process. Each of these version constituents is a state of affairs, i.e., $T_k = S$, and its constituents are the stages that comprise the version.

The relationships between stages specify those that happen sequentially, in parallel, overlapping, or in any other temporal relationships.

Specifying aspects that are objects

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

Objects provide perhaps the clearest illustration of the concept of multiple descriptions of something. For example, in biology, a cell has a part called the ribosome. A part of the ribosome called the large ribosomal subunit is very commonly described as having a roughly spherical main body and three lobes (i.e., with three constituents); but equally commonly, it described as being comprised of two rRNA chains (5s, 23s) and a number of proteins.

Specifying aspects that are states of affairs

Ossorio (1971/1975/1978/2005) notes that the state of affairs is the most general kind of thing in the world; in fact, a world is, formally, a state of affairs. It is therefore not surprising that, with states of affairs, we come full circle: the general Aspect Specification is the specification for a state of affairs.

Specifying aspects that are events

As discussed in (Ossorio, 2005), an event is a direct change from one state of affairs to another, so the event description is the simplest of the four kinds of description, consisting simply of the names of two states of affairs descriptions. Thus, the Aspect Specification for an event consists of the triple (N, E, D), where D consists of the pair (A_1, A_2), in which A_1 and A_2 are each are of the form (N, S, D).

Examples of Aspect Specifications

In this section we give examples of ASs. The examples are chosen from disparate realms to illustrate the previously-mentioned range of applicability, including to areas not typically considered amenable to formal representation—such as the structure of a family; and to prepare the ground for showing we can define similarity and complexity measures that can be used to compare the structure of things that are otherwise entirely unlike—such as a family and an automobile engine or the economy of a society and production line.

A practice in an organization

Jeffrey and Putman (1981) give the following Basic Process Unit (BPU) (Ossorio, 2005, p. 38) description of one of the practices of a software development organization(Figures 2A and 2B).

Name: Responsible persons in the Laboratory find and fix a problem in a No. 4 Generic

Stages:

1. Responsible persons in the Laboratory find out about a problem
   - Option 1-1: A person at Indian Hill discovers a problem and reports it
   - Option 1-2: A person at Indian Hill discovers a problem and has the responsible programmer file a Failure Report
   - Option 1-3: A person at Indian Hill discovers a problem and has the FR coordinator tell the responsible programmer about it
2. People who keep track of problems track the course of the problem
3. The responsible programmer decides the response to the problem
4. The responsible programmer implements the chosen response to the problem
   - Option 4-1: The responsible programmer produces the fix for the problem
   - Option 4-2: The responsible programmer files a Not-Applicable Correction Report
   - Option 4-3: The responsible programmer files a Not-Implemented Correction Report
   - Option 4-4: The responsible programmer files a Cancel Correction Report
5. People in a support group install the fix for the problem in a No. 4 Generic.

Versions:

1-1, 2, 3, 4-1, 5; 1-2, 2, 3, 4-1, 5; 1-3, 2, 3, 4-1, 5; 1-1, 2, 3, 4-2; 1-1, 2, 3, 4-3; 1-1, 2, 3, 4-4; 1-2, 2, 3, 4-2; 1-2, 2, 3, 4-3; 1-2, 2, 3, 4-4; 1-3, 2, 3, 4-2; 1-3, 2, 3, 4-3; 1-3, 2, 3, 4-4

Figure 2A. BPU of a software development practice
Process aspects

Elements

1. person at Indian Hill
2. problem
3. fix for the problem
4. response to the problem
5. NA CR
6. NI CR
7. CC CR
8. No. 4 Generic
9. FR coordinator
10. responsible programmer
11. people who keep track

Individuals

1. anyone at Indian Hill
2. member of 4ESAC
3. 4E4
4. 4E5
5. 4E6
6. persons in a support group
7. the FR for the problem
8. the CR for the problem
9. incorrect behavior by the ESS 4 machine
10. people in System Test
11. people in a development group
12. people in the Field Support Group

Eligibilities

1. E8: I3, I4, I5
2. E1: I1
3. E11: I10, I11, I12

Figure 2B. BPU of a software development practice Elements, individuals, and eligibilities

The Aspect Specification of this process description is:

- Constituents:
  - ◊ C_1: "Problem in No. 4 Generic exists"; $T_1 = S$
  - ◊ C_2: "Problem in No. 4 Generic is resolved"; $T_2 = S$
    The BPU does not include specification of the before and after states of affairs; these two Constituents are in addition to constituents named in the BPU.
  - ◊ $C_3 \ldots C_{14}$: the Processes named in the Stage-Option list; each of $T_3 \ldots T_{14} = P$.
  - ◊ $C_{15} \ldots C_{25}$: the Objects identified in the Element list; each of $T_{15} \ldots T_{25} = O$.
  - ◊ $C_{26 \ldots} C_{37}$: the States of affairs in the Versions list (each Version being a set of processes and therefore a state of affairs); each of the $T_{26} \ldots T_{37} = S$.
- Eligibilities:
  - ◊ (E8, I3, Ø)
  - ◊ (E8, I4, Ø)

- ◊ (E8, I5, Ø)
- ◊ (E1, I1, Ø)
- ◊ (E11, . I10, Ø)
- ◊ (E11, I11, Ø)
- ◊ (E11, I12, Ø)

(where Ø denotes “no rule” or, equivalently, “always eligible”).

Comparing the Aspect Specification to the BPU form of this description, we can see that the AS may provide little advantage over the BPU in the way of readability or accessibility. That, however, is not its purpose. Its purpose is to provide a unification of the descriptive Units so we can address the concept of structure.

Intentional and Deliberate Action

One of the foundational formulations of Descriptive Psychology is the parametric formulation of Intentional Action:

IA = <I, W, K, Kh, P, A, PC, S>

(Ossorio, 2006). Each of these parameters is specified by formal name, and each is a state of affairs or process, as follows:

W (want) is a state of affairs.

K (know) is a set of state of affairs descriptions identifying the distinctions being acted on.

Kh (know-how) is a set of state of affairs descriptions identifying the skills necessary for this action. A skill is characterized by the achievements it makes possible, and so each skill is identified by a set of names of states of affairs.

P (performance) is the procedural aspect of the behavior.

A (achievement) is a state of affairs specifying the actual outcome of the behavior.

PC (person characteristics) is a set of state of affairs, each specified by name, identifying the personal characteristics of which this behavior is an expression.

S (significance) identifies the larger intentional action this is an aspect of. Since any behavior is specified by name and IA parameters, this is a configuration of processes and states of affairs, which is formally a state of affairs.

Intentional Action descriptions are cleanly handled by Aspect Specifications. Each item of knowledge, i.e., each distinction in K, is a constituent, as is each Kh item and each PC item. An Intentional Action AS has constituents $\{C_i\}$ for W, each K item, each Kh item, P, each PC item, A, and S; with the corresponding $\{T_i\}$.

Intentional Action is the most general case of behavior. As discussed in (Ossorio, 2006), by setting the appropriate parameters of the intentional action description to null, one can describe the behavior of humans, animals, machines, or more "basic" things such as particles. What distinguishes human beings, the paradigm case persons with which we are all familiar, is Deliberate Action, the case in which the person knows what they are doing and chooses to do it. Formally, this means that W and K are each of the form $(IA_k, \{IA_1, IA_2, \ldots, IA_n\})$ (Ossorio, 2006). Since each IA_j may be specified via an AS, each set $\{IA_1, IA_2, \ldots, IA_n\}$ is an AS, as is the pair $(IA_k, \{IA_1, IA_2, \ldots, IA_n\})$; therefore Deliberate Action may also be formally articulated via Aspect Specifications.

Communities

Communities are a core concept of Descriptive Psychology, because communities have the central place in the life of persons that they do. This article is addressing and formalizing the concept of structure in its full range of applicability, including what is commonly referred to as "social structure", i.e, organized, cohesive groups of interacting individuals in which the individuals are persons and the "interactions" are persons engaging in human behavior:

families, teams, task forces, companies, governmental bodies, nations, supra-national organizations, entire cultures, and so forth. Accordingly, we show here how to apply Aspect Specifications to formally specify communities.

The concept of community, as formulated by Putman (1981), is that a community is a configuration specified by

C = <M, S, Ct, L, P, W>

where

M	Members
Pr	Practices
Cp	Choice Principles
St	Statuses
Ct	Concepts
La	Language
W	World

S, Ct, and P each identify a set of Constituents, formally specified by name (and, as elsewhere, by Description when more detail is needed). W, the community's world, is a single Aspect, the state of affairs incorporating all other Aspects of that community. Language includes the set of all verbal behaviors in a community, and each verbal behavior is specified by the parameters $V = \langle C, L, \{B_i\}\rangle$ (Ossorio, 2006). For any behavior V, $\{B_i\}$ is an AS whose constituents B_i denote the behaviors, i.e., intentional actions, that are cases of acting on C; L, the locution, is a process; and C is a state of affairs description, therefore, a constituent of type S. The entire set L of verbal behaviors V is thus a Constituent of type S, identified by formal name, as with all constituents and relationships.

Structure

The Concept of Structure

We are now prepared to address the central topic of this article: structure and structural similarity.

The concept of structure is ubiquitous in both everyday and scientific life, so much so that its use goes essentially unremarked. References to structure are found in physics, chemistry, biology, mathematics, astronomy, sociology, cognitive psychology, business, finance, economics, art, literature, and poetry, to name only a few. However, ubiquity notwithstanding, there is no rigorous definition of structure, i.e., a formal definition that captures the concept as used. This lack significantly limits its use by scientists and others interested in precise formulation of their subject matter. We now use Aspect Specifications to give such a definition, thereby making it possible to use the concepts of structure and structural similarity formally and quantitatively.

While there is no accepted formal definition of structure, examination of uses of the concept shows that it is intimately related to the concept of relationships. For example, an extremely widely used Internet resource states, "Structure is a fundamental and sometimes intangible notion covering the recognition, observation, nature, and stability of patterns and relationships of entities…A structure defines what a system is made of. It is a configuration of items. It is a collection of inter-related components or services," (Wikipedia, 2009b).

In mathematics, an area to which one might reasonably look to find a definition of structure, we find two disciplines—*universal algebra* and *model theory*—in which structure consists of an underlying set and the relations defined on that set, i.e., between elements of it. The mathematical discipline of *category theory* studies what are called "structure-preserving" functions between mathematical objects, i.e., functions that preserve the relationships between objects.

In biochemistry, discussions of the primary, secondary, tertiary, and quaternary structure of proteins are commonplace (Voet & Voet, 2005). In every case, the discussions articulate relationships between constituents of the protein. Every discussion of the structure of cells and their parts discusses components of the structure and their relationships (Alberts *et al*, 2002). In physics and engineering, one mathematically analyzes how structures behave in various conditions, i.e., how the relationships between components of the structure change under various loads.

In the social sciences, social structure—of organizations, societies, families, professions, and so forth—is typically defined in terms of the statuses within the various Communities. The structure of statuses is another way to talk about relationships: manager-subordinate, physician-patient, teacher-student, husband-wife, mother-daughter, etc. Thus, to talk about family structure is to talk about the relationships between family members. Social structure is also used to refer to the relationships between larger social entities in a society: families, interest groups, religions, ethnicities, gangs, etc., i.e., the entire range of communities that are part of larger communities. Economic structure refers to the relationships between economic processes, objects, and events. Universities teach courses in the structure of the novel, in which the relationships between aspects of the story are the central subject matter. We can see just such a discussion, albeit in literary form, in the following:

> A novel is like a symphony in that its closing movement echoes and resounds with all that has gone before…Toward the close of a novel unexpected connections begin to surface; hidden causes become plain; life becomes, however briefly and unstably, organized; the universe reveals itself, if only for the moment, as inexorably moral; the outcome of the various characters' actions is at last manifest; and we see the responsibility of free will (Gardner, 1991).

The novel, in other words, shows the relationships between the characters' actions and events in the depicted in the novel.

In light of all of the foregoing examples, it seems clear that our concept of something's structure is that of its constituents and the relationships between them.

Using Aspect Specifications to Define Structure

In an Aspect Specification, each Paradigm of the Description of an aspect of the world, or of a world, consists of a specification of a set of constituents, their relationships, contingencies, and a set of eligibility rules. At first glance, the set $C = \{(C_i, T_i)\}$ and $R = \{R_j\}$ appear to capture exactly the concept of structure as consisting of components and relationships. However, a bit of care is needed, for the aspect—object, process, event, or state of affairs—may have multiple descriptions, therefore, multiple structures. To talk about something's structure is to talk about its structure *under a particular description*. With this elaboration, we can define the structure of any aspect of a world as follows:

> The *structure* of aspect A, as specified by Paradigm P, is the ordered pair of sets {C, CR} of constituents and inter-constituent relationships in P.

As shown in Fig. 1, there are three types of relational statements in an Aspect Specification: CR, GR (relationships in contingencies), and ER (relationships in eligibility rules). As we have defined it here, structure includes only the inter-constituent relationships CR, because this appears to capture the concept of structure as it is used. While it is clearly possible to expand the definition to include the contingency relationships GR, this does not appear to be the paradigm case concept of structure. Eligibility rules, because they govern which individuals may serve as the various constituents of an aspect, are analogous to the operational aspects of a process, i.e., what must be given to specify an *actual instance* of the aspect rather than the *structure* of the aspect per se.

There are, in principle, an unlimited number of possible Descriptions of a world or some aspect of it, but in many cases one Description is so commonly used that is considered the "normal" case. This is true with many ordinary objects and processes in everyday life, but it is also true in scientific and technical communities. In such cases, we commonly find the normal Description is referred to as simply, "the structure of X", eliding the phrase "under description D."

When describing the structure of communities and organizations, and when comparing their structure, this phenomenon becomes significantly more important. It is often the case that an organization's Practices, for example, may be described in many ways, i.e., via different Descriptions, but the Description considered accurate and complete by the members of the community is the one customarily called "correct." In short, while many Descriptions of the same Aspect are logically possible, the Descriptions do not all have the same status.

Measuring structural complexity and similarity

In the sciences and in ordinary life, we speak of and use two concepts related to structure: complexity and similarity. We now use the formal definition of structure to derive mathematical definitions of these concepts, thereby providing rigorous and quantifiable measures of them. The goal is to develop a definition of structural similarity that allows us to calculate the similarity of any two Aspects of the world.

Complexity

A preliminary step in the derivation of the structural similarity measure, which makes the derivation somewhat easier, is to develop a measure of structural complexity. We define the *basic structural complexity* of an Aspect A, with N constituents A_1, …, A_N and relationships CR_1, …, CR_K relationships, recursively as:

$$BSC(A) = \sqrt{N^2 + K^2 + \varepsilon \cdot \sum_{i=1}^{N} BSC(A_i)^2} \tag{1}$$

ε is an experimentally-determined multiplier modulating the impact of the complexity of constituents, sub-constituents, etc. An experiment of this kind would involve assembling a panel of qualified judges of complexity of a set of aspects of interest, querying them to determine their assessments of the complexity of the aspects, and then determining the value of ε that yields the best match with the experimental data.

Formula (1) is designed to directly measure the following aspects of the concept of structural complexity:

1. We call something more complex if it has more parts—i.e., constituents.
2. We call that thing more complex if it has more relationships between the constituents.
3. We call something complex when its parts have greater complexity.

One candidate for inclusion in Formula (1) is conspicuously absent: complexity of relationships between constituents. It seems, on inspection, that we use the concept of complexity of a relationship when we call something complex, and it would therefore be appropriate to include it as part of structural complexity. On further examination, however, the situation is more problematical. We can see this by examining two kinds of relationships: those involved in human systems and in physical systems. Since human relationships are defined in terms of the eligibilities for practices, one way to define relational complexity is by the number of Practices in which the relationship appears in an eligibility rule. It is not clear, however, how well this definition corresponds with our concept of the complexity of a relationship.

Consider, for example, two relationships between teacher and student. TS_1 is characterized by the ordinary classroom practices of lecturing, giving homework, doing homework, giving exams, doing exams, and giving feedback on homework. TS_2 is characterized by the same practices plus the additional practice of one-on-one tutoring of students. It is easy to say the complexity of TS_2 is greater than that of TS_1, as a matter of mathematical definition, but it is by no means clear that we would say that TS_2 is more complex than TS_1, other than in this purely definitional sense.

A different example, and a different kind of difficulty, may be found in the physical world. Engineers frequently study what happens to a physical body under various conditions, such as the deformation of a structure in response to a force. This defines a relationship between two quantities, specifiable via an Aspect Specification in which the constituents are the states of affairs "force applied" and "conformation of structure." This relationship is analyzed mathematically, and graphs depicting the relationship between force and deformation routinely constructed. Engineers commonly describe some of these relationships as complex, particularly when the equations governing the relationship of the quantities are complex. There is, however, no accepted mathematical definition of the complexity of such relationships, and no accepted way to mathematically define the complexity of formulas involving complex mathematics.

In short, defining a measure of complexity of relationships applicable across the entire range of phenomena of the real world—which is the range of applicability of Aspect Specifications—is beyond the scope of this paper. For our purpose here, which is addressing structure and structural similarity, we do not need such a measure. We shall see below that the role of BSC in the development is only to provide a means of consistently ordering constituents of an aspect for purposes of a structural similarity calculation. A full development of the concept of complexity appears to be an interesting topic for further research.

We have defined BSC solely in terms of constituent relationships CR_i, excluding contingency relationships GR_i and eligibility relationships ER_i, for the reason noted at the beginning of this section: it is not clear the paradigm case of structural complexity includes those relationships. Should it be practically useful or valuable to include them, the extension of the definition of BSC to incorporate these other relationships is straightforward: the value "K" is replaced by the number of all included relationships.

While there is currently no accepted general formulation of relationship complexity, it may often be the case that there is an *a priori* formulation of it in a particular domain, a formulation defining, for any relationship R in that domain, its complexity $\mathcal{C}$(R). In this case, we define the *extended structural complexity ESC* of an aspect A as:

$$ESC(A) = \sqrt{N^2 + \sum_{i=1}^{K} \mathcal{C}(R_i)^2 + \varepsilon \cdot \sum_{i=1}^{N} BSC(A_i)^2} \qquad (1b)$$

Note that BSC(A) = ESC(A) if $\mathcal{C}(R_i) = 1$, for each R_i.

Complexity has been studied and defined in a number of ways by many authors in a variety of disciplines (Wikipedia 2009a), and this definition is only one of many possible. Of those that appear to address structural complexity (as contrasted with, for example, difficulty of computation), all use the concepts of number of constituents and the number of their relationships in some form. Kolmogorov complexity, for example, defines the complexity of a string of bits as the length of the shortest binary program that can compute the string. Krohn-Rhodes complexity defines the complexity of certain mathematical objects (called "semigroups") in terms of the number of other mathematical objects ("groups"), related in a certain way (the "wreath product") needed to re-describe them. In the field of computer software, a number of attempts have been made to define the complexity of a program in terms of the number of possible ways the code can be executed. In recent years, an entire field of study, known as Complex Systems, has arisen. A complex system is one that has a large number of components related in such a way as to produce nonlinear system behavior. To our knowledge, no previous definition of structural complexity has been devised to directly measure the aspects of number of constituents, number of relationships, and (recursively) the complexity of constituents.

The Pythagorean formula in (1) and (1b), known mathematically as the Euclidean distance, is a traditional and widely-used formula for computing distance. Recalling basic geometry, the differences in x-coordinates and y-coordinates vary independently; thus, this measure is a standard method for combining independently-varying quantities into a single one.

Similarly, the number of constituents and relationships in a Description and the complexity of each constituent are all independent quantities. Nevertheless, any distance measure may be used, both here and in the formulae developed below; nothing depends on the use of this particular measure.

Similarity

In formally defining a similarity measure between any two Aspects, we want to take into account the following intuitions:

- The measure should be responsive to differences in the number of constituents of the respective aspects.
- The measure should be responsive to differences in the aspects themselves. A 2-gallon pail and an 8-ounce drinking glass would be considered different, even though they had identical shape and constituents.
- The measure should be responsive to differences in the attributes of the constituents of the aspects.
- Since relationships between constituents are the heart of the articulation of the concept of structure, the measure should reflect differences in relationships between constituents. The differences may be that the two aspects have different inter-constituent relationships or that they have the same relationships to different degrees.

- When the structure of the constituents of the aspects is known, the similarity between the aspects should reflect the similarity of the structure of their respective constituents.

Our goal is to define a mathematical measure of similarity, but we will do this by developing a formula for "dissimilarity." This lets us use the mathematical concept of distance: a small distance means a small difference in similarity, and the mathematics are a bit easier. We define the structural distance between two aspects in terms of the difference of:

1. the number of constituents of each aspect,
2. the properties of the aspects themselves,
3. the properties of the constituents,
4. differences in relationships between the aspects' constituents, and
5. differences in structure of the constituents, as follows.

Let A and B denote any two aspects—object, process, event, or state of affairs, not necessarily of the same kind—for which we have structural specifications, i.e., Descriptions including at least the constituents and their relationships. Denote the constituents of A and B by A_1, …, A_{NA} and B_1, …, B_{NB}, respectively, and the properties of interest by P_1, …, P_M. (It is not required that all A-constituents and B-constituents have all properties.) Denote the constituent relationships (CR_i, in Fig. 1) between A-constituents by AR_1, …, $AR_{K,}$ and those between B-constituents by BR_1, …, BR_L.

Calculations of the similarity between A and B are affected by the order of the constituents. Consider, for example, two organizations A and B. A has a large and complex marketing department and a small, simple shipping department; while B has a large and complex shipping department and a 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.

In some cases, the Specifications of A and B are in terms of constituents commonly recognized as comparable. If we are calculating the structural similarity of two automobile engines, two human faces, the bodies of two animals of the same species, two versions of a production process, two political communities with recognized sub-communities, etc., the correspondence between the ordinarily-used constituents of A and of B is unambiguous. In the example above, we would normally compare corresponding marketing and shipping departments. In other cases, however, there is no such implied correspondence.

Consider, for example, calculating the similarity of organizations A and B, but with the addition that A has a large accounting department and small IT department; while B has a large department devoted to on-line sales, and a small consulting division. In more extreme cases, we may need to compare aspects with no clearly comparable constituents, such as an automobile engine and a turbofan jet engine; the liver and the pancreas; or putting on a play and building a microchip.

To address this issue, we adopt the following procedure: when there is an already-accepted correspondence between constituents of the two aspects to be compared, those pairs are used in the calculation; when there is no such a priori correspondence between constituents of A and B, the complexity Formula (1) is used to order the constituents of each aspect in decreasing order of complexity, thus providing a consistent basis for the similarity calculation. In the example of organizations A and B, we would compare corresponding shipping departments and marketing departments, because that is

	P_1	…	P_M
A_1			
…			
A_{NA}			

Figure 3. The Constituent Property Matrix PA

the accepted correspondence, and order the remaining constituents according to Formula (1).

We represent the properties of A-constituents in a property matrix PA, as shown in Figure 3:

- PA has M columns, one for each property of interest.
- The matrix entries are the values of each constituent on each property P_i.
- If a constituent does not have property P_i, that matrix entry is blank.

The properties of B-constituents are represented similarly, in the matrix PB.

Re-order the rows of PA and PB to reflect (1) customary correspondence, and (2) decreasing order of BSC of the remaining A-constituents and B-constituents, as discussed above.

The values in PA and PB may represent quite different properties, with numerical values in entirely different ranges, so in order to meaningfully compare numerical values representing disparate properties, the values must be normalized. Accordingly,

- If any column has a value < 0, re-scale the values of the column by adding the absolute value of the minimum value of the column to each value in it. This makes the minimum value of each column 0.
- Letting PA_i denote column i of PA and $pmax_i$ denote the maximum value of column i of PA *and* PB, normalize the values of PA to the range 1 to 10, by setting the new $PA_i(A_j)$ to 10 * $(PA_i(A_j) + 1) / (pmax_i+1)$). (The value of 10 is an empirically-determined value, chosen to emphasize the relative importance of property and relationship differences compared to simple number of constituents.)
- Set each empty entry of PA to 0.

The values of the property matrix PA are now between 0 and 10, 0 indicating the component does not have the property of that column, and (by construction) 1 being the minimum actual property value.

Similarly normalize the values of PB.

We can now define the *property distance* and *constituent property distance* between A- and B-constituents, A_i and B_j, by the Euclidean distance between the corresponding rows of PA and PB:

$$PD(A_i B_i) = \sqrt{\sum_{i=1}^{M} \left(PA_k\left(A_i\right) - PB_k\left(B_i\right)\right)^2} \quad (2)$$

and the constituent property distance between A and B is

$$CPD(A,B) = \sqrt{\sum_{i=1}^{N} PD\left(A_i, B_i\right)^2} \quad (3)$$

It is commonplace to represent properties of aspects—objects, processes, etc.—in form like that of the constituent property matrix but with just two rows, one for each aspect. Assuming we have properties Q_1, …Q_Z of A and B, we define the *property distance* between A and B as:

$$PD(A,B) = \sqrt{\sum_{i=1}^{Z} \left(Q_i\left(A\right) - Q_i\left(B\right)\right)^2} \quad (4)$$

Some properties cannot be compared, i.e., $PA_k(A_i)$—$PB_k(B_i)$ or $Q_i(A)$—$Q_i(B)$ is meaningless, unless there is an accepted *a priori* ordinal representation of the difference $Q_i(A)$—$Q_i(B)$. This is a well-known phenomenon in statistics, where demographic data such as religion, ethnicity, country of origin, etc. are common examples. Such a situation may arise in other domains as well, such as cases in which the physical location of a state of affairs is a property of interest. In such a case, the difference is either ignored in the PD calculation, or the accepted *a priori* definition is used. (We shall see an example of this below, in calculating the structural distance between two families.)

We use a similar matrix technique to calculate differences in relationships between A- or B-constituents. The values of the relationships CR_j are given by ordered n-tuples. Each relationship has a specific value. For example, in human hemoglobin in the R-state, the angle between the two constituents customarily named $\alpha_1\beta_1$ and $\alpha_2\beta_2$ is 15°. Thus, the relationship has the formal name "angle", and angle($\alpha_1\beta_1$, $\alpha_2\beta_2$) = 15.

Denoting the number of A-tuples by NAT, and the number of B-

	R_1	...	R_K	R_{K+1}	...	R_{K+L}
A-tuple$_1$						
...						
A-tuple$_{NAT}$						

Figure 4. The Relationship Matrix RA

tuples by NBT, we define RA as shown in Figure 4:

- RA has K+L columns, one for each relationship between A- or B-constituents
- Each row of RA represents one n-tuple of A-constituents, so there are NAT rows. Denote these n-tuples ta_1, ta_2, …, ta_{NAT}
- The matrix entries are the values of the relationships on the n-tuples. For example, the entry for the matrix at the row ($\alpha_1\beta_1$, $\alpha_2\beta_2$), column "angle", is 15.
- If an n-tuple does not have relationship R_k, the corresponding entry of the matrix is blank.

Similarly, represent the relationships between B-constituents as the matrix RB.

As with PA, the values of RA must be normalized in order to be able to make meaningful calculations:

- If any column has a value < 0, re-scale the values of the column by adding the absolute value of the minimum value of the column to each value in it.

- Letting RA_i denote column i of RA, and $Rmax_i$ be the maximum value of column i of RA *and* RB, normalize the values of RA to the range 1 to 10, by setting

$$RA_i(ta_j) = 10 \times (RA_i(ta_j) + 1) / (Rmax_i+1).$$

  (As with P, the value of 10 is an empirically-determined value, chosen to emphasize the relative importance of property and relationship differences compared to simple number of constituents.)

- Set each empty entry of RA to 0.

  Similarly normalize the values of RB.

As with properties of constituents, it is necessary to have a consistent scheme for calculating the Euclidean distance between rows of RA and RB. Just as the constituents of two aspects may be commonly recognized as comparable, in some cases the rows of RA and RB represent tuples that would commonly be compared. For instance, in the family structure example below, mother, father and siblings have various relationships. We would ordinarily compare the mother-father relationships and the sibling relationships, rather than calculating the difference between the mother-father relationships in one family with the sibling relationships in the other. Again as with properties, in other cases there may be no such implied correspondence. We would see this, for example, if one family has an older estranged half-sibling and a nanny, while the other has a pet dog and a live-in elderly mother.

We therefore re-order the rows of RB as follows. When there is an *a priori* correspondence between rows (tuples) of RA and RB, set row 1 of RB to the row customarily comparable to row 1 of RA, row 2 of RB to that customarily comparable to row 2 of RA, and so forth, until we reach a row of RA which has no customarily-corresponding row in RB, or all rows of RA are exhausted. For the remaining rows of RA, set the next row of RB to the row of RB closest, by Euclidean distance, to the first remaining row of RA, the next row of RB to the row next closest to the next row of RA, and so forth, until all rows of RA have been exhausted. Append any remaining rows of RB to the re-ordered RB matrix, in the existing order. (The order of these remaining rows does not matter, as will be seen in the discussion following Formula (6) below.)

We can now define the **total distance between two Aspects A and B** in terms of the property distance and the structural distance:

$$TD(A,B) = \sqrt{PD(A,B)^2 + SD(A,B)^2} \tag{5}$$

The *structural distance* SD(A, B) is defined recursively, using RA and RB, 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, we define the structural distance SD in terms of the Euclidean distance between tuples of RA and RB:

$$SD(A,B) = \sqrt{\begin{array}{l}(NA - NB)^2 + CPD(A,B) + \\ \sum_{j=1}^{MT} \sum_{i=1}^{K+L} \left(RA_i(ta_j) - RB_i(tb_j)\right)^2 + \\ \delta \cdot \sum_{i=1}^{MC} SD(A_i, B_i)^2\end{array}} \tag{6}$$

If NA > NB, then PD(A_i, B_i) = PD(A_i, 0) for i > NA, and similarly if NB > MC.

If NAT > NBT, i.e, there are more relationship tuples in RA than in RB, then the Euclidean distance for the "extra" RA-tuples (rows) is found by treating RB as having extra rows filled with zeroes, i.e., by considering $RB_i(tb_j)$ = 0 for j > NBT, and similarly if NBT > NAT.

If NA > NB, then for i > NB, there is no B_j corresponding to A_i, so SD(A_i, B_i) = SD(A_i, 0)
where 0 indicates that the values of the property and relationship matrices for B_i are all zeroes, and similarly if NB > NA.

If either A or B have no Description, SD(A, B) = 0, corresponding to the intuition that if we know nothing of the structure of A or B, we can say nothing about their structural difference.

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.)

Intuitively,

- $PD(A_i, B_i)$ measures similarity of properties of each pair of constituents.
- The sum $\sum_{i=1}^{K+L}\left(RA_i\left(ta_j\right)-RB_i\left(tb_j\right)\right)^2$ measures how much the constituents of A and B differ on the entire set of relationships $R_1, \ldots, R_{K+L}$;
- The sum $\sum_{j=1}^{MT}\sum_{i=1}^{K+L}\left(RA_i\left(ta_j\right)-RB_i\left(tb_j\right)\right)^2$ measures the total difference in structures A and B, as specified by the relationships R_i between A- and B-constituents.

If A and B are the same, except having different names of constituents and relationships (mathematically, are isomorphic), TD(A, B) = 0. As the properties of A and B, the number of their constituents, the properties of the constituents, the structure of A and B, and the substructures of A and B diverge, TD(A, B) increases.

It was noted in the discussion of the Basic Structural Complexity measure that in some applications it may be desirable to incorporate a measure of the complexity of relationships, $\mathcal{C}(R_i)$. A somewhat similar situation is the case with the similarity measure SD. Formula (6) can be considered a “basic” or “fundamental” measure, in that it measures differences between relationships of constituents without taking into account any differences in relative importance of relationships. This basic measure may not be what is needed in all applications. For example, in some social structure similarity calculations, such as similarity of families, it may be of value to emphasize certain relationships over others, resulting in a measure that is "weighted" by importance of certain relationships. This may be done by weighting each relationship with a value w_i as follows:

$$WSD(A,B) = \sqrt{\begin{array}{l}(NA-NB)^2 + CPD(A,B) + \\ \sum_{j=1}^{MT}\sum_{i=1}^{K+L} w_i \times \left(RA_i(ta_j) - RB_i(tb_j)\right)^2 + \\ \delta \cdot \sum_{i=1}^{MC} WSD(A_i,B_i)^2\end{array}} \qquad (6b)$$

Examples of Structural Similarity

We noted at the outset of this article that an important fact about the descriptive Units, and therefore of the Aspect Specification unification of them, is their unlimited range of applicability. The (Name, Description) methodology, in which each aspect, constituent, and relationship is identified by purely formal name, and further described via further Aspect Specifications, provides a technical resource applicable to the entire range of what there is in the world: communities, complex individual behaviors, mechanical systems, biological objects and processes (including brains and neurological processes), psychodynamics, molecular structures, etc., ad infinitum. In the examples below we have purposely chosen disparate kinds of aspects, from families to molecules, to illustrate this applicability.

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, and the children love each other. However, the children also compete with each other for success in school. The mother is age 40 and in reasonable health; the father is age 42 and also in reasonable health, though slightly less so than the mother. 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 and love each other. Both parents love all the children. The two younger children love each other, but both resent the eldest 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. For the purposes of this example, we omit the other normal relationship of the children loving the parents, which only expands the size of the relationship matrices without adding clarity.) The eldest child also has a significant responsibility in caring for the younger children. The mother is age 35 and in excellent health; the father is age 36 and in good health. Family B has an income of $85,000 and is Presbyterian. (We are supposing here that these are the attributes of interest in this case. As discussed above, calculated similarity necessarily depends on the properties and relationships represented in the ASs.) This gives the Property and Relationship matrices shown in Figures 5 through 8:

	Age	Health
M_A	40	0.8
F_A	42	0.7
AC_1	12	1.0
AC_2	10	1.0

	Age	Health
M_B	35	0.9
F_B	36	0.8
BC_1	14	1.0
BC_2	8	1.0
BC_3	6	1.0

Figure 5. PA and PB for families A and B

	Income	Religion
A	$75,000	C
B	$80,000	P

Figure 6. Property Matrix Q for families A and B

	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			

Figure 7. RA for Family A

Calculation of TD(A, B) proceeds as follows:

	Rom.. Love	Par. love	Sib. love	Aca. comp.	Resent	Affec. comp.	Care-taker
(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_2,BC_3)			1.0				
$(BC_3\ BC_2)$			1.0				
$(BC_3.\ BC_1)$					1.0		
(BC_2, BC_1)					1.0		
(BC_2, BC_1)						1.0	
(BC_3, BC_1)						1.0	
(BC_1, BC_2)							1.0
(BC_1,BC_3)							1.0

Figure 8. RB for Family B

- The normalized income values are (75+1)/(80+1) and (80+1)/(80+1), yielding values of 9.4 and 10.0. Stipulating for expository purposes the previously mentioned *a priori* ordinalization of the ethnographic categories “Catholic” and “Protestant” as a difference of 0.3 (on a 0—1 scale), and normalizing, we have

	Age	Health
M_A	9.5	9.0
F_A	10.0	8.5
AC_1	3.0	10.0
AC_2	2.6	10.0

	Age	Health
M_B	8.4	9.5
F_B	8.6	9.0
BC_1	3.5	10.0
BC_2	2.1	10.0
BC_3	1.6	10.0

Figure 9. Normalized PA and PB for families A and B

$$PD(A,B) = \sqrt{(9.4-10.0)^2 + 3^2} = 3.06$$

- $(NA–NB)^2 = (5\text{-}3)^2 = 4$.

Figure 9 shows normalized PA and PB, with rows of PB re-ordered as described above:
yielding a constituent property distance

$$CPD(A,B) = \sqrt{\sum_{i=1}^{5} PD(A_i, B_i)^2}$$

$$= \sqrt{\begin{array}{l}(9.5-8.4)^2 + (9.0-9.5)^2 + \\ (10.0-8.6)^2 + (8.5-9.0)^2 + \\ (30.0-2.1)^2 + (10.0-10.0)^2 + \\ (2.6-1.6)^2 + (10.0-10.0)^2 + \\ (0.0-3.5)^2 + (0.0-10.0)^2\end{array}}$$

$$= 10.8$$

The normalized relationship matrices, with rows of RB re-ordered as described above, are shown in Figures 10 and 11: (Rows are numbered for ease of reference.)

	Rom.. Love	Par. love	Sib. love	Aca. comp.	Resent	Affec. comp.	Care-taker
1. (M_A, F_A)	10.0						
2. (F_A, M_A)	10.0						
3. (M_A, AC_1)		10.0					
4. (M_A, AC_2)		10.0					
5. (F_A, AC_1)		10.0					
6. (F_A, AC_2)		10.0					
7. (AC_1, AC_2)			10.0	10.0			
8. (AC_2, AC_1)			10.0	10.0			

Figure 10. Normalized RA for Family A

	Rom.. Love	Par. love	Sib. love	Aca. comp.	Resent	Affec. comp.	Care-taker
1. (M_B, F_B)	10.0						
2. (F_B, M_B)	10.0						
3. (M_B, BC_1)		10.0					
4. (M_B, BC_2)		10.0					
5. (F_B, BC_1)		10.0					
6. (F_B, BC_2)		10.0					
7. (BC_2, BC_3)			10.0				
8. (BC_3, BC_2)			10.0				
9. (M_B, BC_3)		10.0					
10. (F_B, BC_3)		10.0					
11. (BC_2, BC_1)					10.0		
12. (BC_3, BC_1)					10.0		
13. (BC_2, BC_1)						10.0	
14. (BC_3, BC_1)						10.0	
15. (BC_1, BC_2)							10.0
16. (BC_1, BC_3)							10.0

Figure 11. Normalized RB for Family B

The values in rows 1 through 6 of RA and RB are identical, rows 7 and 8 differ only on the Academic Competition column, and the remaining RA tuples have no matching RB tuple, so the sum

$$\sum_{j=1}^{MT}\sum_{i=1}^{K+L} w_i \times \left(RA_i\left(ta_j\right) - RB_i\left(tb_j\right)\right)^2$$

$$= 6\times 0^2 + 2\times 10^2 + 8\times 10^2 = 1000$$

Thus, the structural distance (dissimilarity) SD between families A and B

$$SD(A,B)=\sqrt{4+10.8^2+1000}=33.48$$

and the total distance between the two,

$$TD(A,B)=\sqrt{3.06^2+33.65^2}=33.79$$

Intuitively, the values of SD and TD are so close (0.9% of the TD) because the difference between the families based only on their income and religion, as shown in Fig. 5, is much less than the difference based on their respective structures.

In this example, we are considering structures 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. The following section addresses how non-0 values of SD(A_i, B_i) can be meaningful.

Consider now the distance between A and B', where B' is identical to B except that B'C_1 and B'C_2 do not resent and compete for affection with B'C_3. Rows 11 through 14 of RB would be missing, so the sum

$$\sum_{j=1}^{MT}\sum_{i=1}^{K+L} w_i \times \left(R_i\left(ta_j\right)-R_i\left(tb_j\right)\right)^2$$

$$=6\times 0^2+2\times 10^2+4\times 10^2=600$$

$$SD(A,B)=\sqrt{4+10.8^2+600}=26.84$$

and

$$TD(A,B)=\sqrt{3.06^2+26.84^2}=27.01$$

Incorporating dynamics and personalities of family members

In the examples above, the members of the families are not further described, and the resulting structural similarity measures reflect only what might be called the "role structure" of the families. Families, however, are communities. The descriptions in the above examples are partial, not incorporating any of the practices in the families, either the mundane, such as making meals or cleaning rooms, or the very significant, such as accreditations and degradations, the practices often referred to as "family dynamics." As we have discussed earlier, practices are P-type constituents in a fuller Aspect Specification of a community, and thus easily incorporated in more extensive and informative similarity calculations.

We noted earlier that the term "aspect specification" was chosen to avoid the connotation that constituents are objects. Members of a family are persons, and here we see an example of the benefit of the less-connotation-laden term: it is easy to see how we can extend the Specifications of the family to include not only family structure and dynamics but personality characteristics of the members. Personality characteristics are, as discussed earlier, specifiable with type-S Aspect Specifications. This means all knowledge of the members' traits, attitudes, styles, abilities, and all other personal characteristics may be formally included in the specifications of the families, and used in the multi-level measurement of the complexity of the families and differences between them.

An example from chemistry

In this illustration, we move to an entirely different kind of aspect, simple molecules, as examples of simple physical structures. We consider the water molecule (H_2O) and the ammonia molecule (NH_3), which contain three and four constituent objects, respectively, that are at particular angles and distances from their central atom (O and N, respectively). We use two common properties of the

	Atomic weight	Electronegativity
O	16	3.44
H	1	2.2
H	1	2.2

	Atomic weight	Electronegativity
N	14	2.04
H	1	2.2
H	1	2.2
H	1	2.2

Figure 12. A and PB for H_20 and NH_3

	D	α
(O, H)	95.84	104.5
(O, H)	95.84	104.5

	D	α
(N, H)	101.7	107.8
(N, H)	101.7	107.8
(N, H)	101.7	107.8

Figure 13. RA and RB for H_20 and NH_3

constituents: atomic weight and electronegativity; and one property of the overall molecules: dipole moment. The dipole moments of water and ammonia are 1.85 and 1.42 respectively; the property and relationship matrices P and R are shown in Figures 12 and 13.

Normalizing P and R as before, Formula (6) gives:

$$SD(H_2O, NH_3) = \sqrt{1^2 + 3.4^2 + 14.2^2} = 14.62$$

and

$$TD(H_2O, NH_3) = \sqrt{1.5^2 + 14.62^2} = 14.69$$

An example from cell biology

Jeffrey (2009) uses Aspect Specifications and Formula (6) to formally specify two cellular structures, eukaryotic and prokaryotic ribosomes, and calculate their similarity. Information represented in the ASs is taken from a figure in a classic molecular biology text (Alberts *et al*, p. 343, Fig. 6-63), entitled “A comparison of the structures of prokaryotic and eukaryotic ribosomes.” The figure

N: eukaryotic ribosome
Relationships:
a. molecular weight(eukaryotic riborosm) = 4,200,000
D:
Paradigm: 1
Sub-objects:
1. SRSU
2. LRSU
Relationships:
a. molecular weight(SRSU) = 1,400,000
b. molecular weight(LRSU) = 2,800,000
c. adjacent(LRSU, SRSU)

N: SRSU
Relationships:
a. molecular weight(LRSU) = 1,400,000
Paradigm: 1
Sub-objects:
1. 18S RNA
2. Protein1
…
34. Protein33

N: LRSU
Relationships:
a. molecular weight(LRSU) = 2,800,000
Paradigm: 1
Sub-objects:
1. 5S RNA
2. 28S RNA
3. 5.8S RNA
4. Protein1
…
52. Protein49

N: 5S RNA
Paradigm: 1
Sub-objects:
1. Nucleotide1
…
120. Nucleotide120

N: 28S RNA
Paradigm: 1
Sub-objects:
1. Nucleotide1
…
4700. Nucleotide4700

N: 5.8S RNA
Paradigm: 1
Sub-objects:
1. Nucleotide1
…
160. Nucleotide160

N: 18S RNA
Paradigm: 1
Sub-objects:
1. Nucleotide1
…
900. Nucleotide1900

Figure 14. Aspect Specifications for the Eukaryotic Ribosome

shows that each ribosome has two components, the Large and Small Ribosomal Subunits. The prokaryotic LRSU has 2 RNA molecules and 34 proteins, and SRSU has one RNA molecule and 21 proteins; the eukaryotic LRSU has 3 RNA molecules and 49 proteins, and the SRSU has 1 rRNA molecule and 33 proteins.

To represent these facts about the eukaryotic ribosome in AS format, we need 7 Aspect Specifications, one for each object or sub-object with named components: the ribosome, LRSU and SRSU, and 5S, 28S, 5.8S, and 18S rRNA. These are shown in Figure 14:

Using Formula (6), including the similarity of corresponding constituents in the two kinds of ribosome, we find that the structural distance between them is:

$$SD(A,B) = \sqrt{0 + 31.1 + \left[0.7\left(1509.8 + 360.1\right)\right]} = 36.61$$

and the total distance is:

$$TD(A,B) = \sqrt{16.38 + 36.61^2} = 36.83$$

This example illustrates the practical impact of the robustness of Aspect Specification in the face of incomplete information. Fig. 13, and therefore the Specifications taken from it, is conspicuously incomplete, both at the levels shown and because it lacks further Specifications of the rRNA components and proteins (about which a good deal is known). The similarity calculation nevertheless can be carried out without difficulty. The calculated similarity can be expected to change as more detail is added, in accordance with common use: the more detail included in a description of two things, the more possibilities there are for greater difference to be

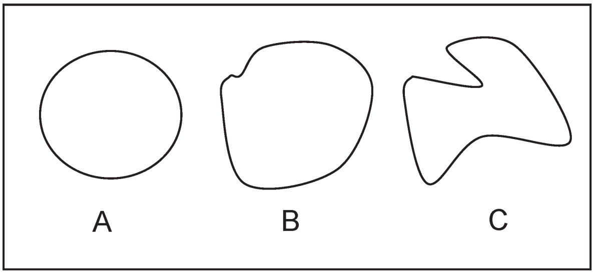

Figure 15. Three objects with different shape

identifiable

This example also illustrates a central theme of this article: the capability of formalizing and quantifying the visual and verbal information found in the figure from Alberts *et al*, a portrayal of structure that is extremely common.

Property distance revisited

Formulas (2) and (3), PD(A, B) and CPD(A, B), define property distance based on the values of other specified properties, using the Euclidean formula. There are instances of property distance, however, in which we have no such other properties to use to calculate the distance. For example, consider the three objects pictured in Figure 15:

Objects A and B are observably more similar in shape than are A and C, but we cannot use the Euclidean formula on properties to calculate these differences.

We therefore generalize Formula (2) to allow for any valid mathematical distance measure d(A, B), as follows: the Property Distance between A and B is defined as

$$PD(A, B) = d(A, B)$$

and

$$CPD(A,B) = \sqrt{\sum_{i=1}^{M} \left(d\left(A_i, B_i\right)\right)^2} \tag{7}$$

In the case of the three objects A, B, and C in Fig. 14, for example, this allows us to use the well-known mathematical measure of shape similarity, the Hausdorff distance (Wikipedia, 2009c), in similarity calculations.

Future Development

The formulations of structure, structural complexity, and structural similarity presented here constitute formalizations of fundamental concepts in extremely wide use in virtually every branch of the sciences and technology. We suggest here only a few of the newly possible lines of development.

Perhaps the most obvious further work is the empirical verification of Formulas (1) and (6), and the above-mentioned systematic investigation of the concept of complexity based on ASs.

Aspect Specifications have been used in several ways, in the form of Process Unit knowledge bases for a number of computer systems (Jeffrey & Putman, 1983), Jeffrey *et* al, 1989), and appear robust both practically and conceptually. The basis for the complexity and similarity formulas is sound, but this by no means ensures that the values derived from them will correlate highly with actual human judgments of these attributes. Verification of the appropriateness of the formulas, and of appropriate values for ε and δ, is clearly needed.

In the field of biology, there are large databases representing the sequence of nucleic acids that make up a gene, or amino acids that make up a protein. These databases are extremely valuable in biological research, largely because algorithms to define and measure sequence similarity have been defined and programmed. The resulting computer systems allow researchers to search large databases for sequences similar to one of interest, to varying degree, and this capability is at the heart of many research efforts in biology. A sequence of constituent nucleic or amino acids is a very special case of an Aspect Description: the Constituents are the component molecules, and the single relationship is that of adjacency. The formulation of structural similarity in Formula (6) is the full generalization of sequence similarity to the entire set of relationships and constituents, at any level of detail, of the DNA or protein, and makes possible computer systems that can search a database for molecules similar in structure to one of interest, based on the relationships between constituents that a researcher finds of interest.

In a variety of fields dealing with organizations, including organizational psychology and consulting, researchers and practitioners often rely on concepts of organizational complexity and similarity, and assessments of them, to analyze organizations and identify problems and potential improvements. Use of terms such as "complexity of organizations" is extremely widespread, and attempts to draw conclusions about organizational effectiveness based on ideas of complexity are commonplace (Anderson 1999, Axelrod & Cohen 2000). It seems likely that the formulations developed here will lead to a number of interesting developments in this field, due to the novel capability of rigorously defining and measuring these virtually ubiquitous concepts.

Organizations have long been compared to mechanisms and organisms. As useful as that analogy has been for developing insights and approaches, it has until now only been possible to use it in that way. We can now move beyond that, to rigorously and precisely articulate and measure the similarity between organizations and organisms. It seems particularly valuable in this regard that Formula (6), the similarity measure, is responsive to differences in structure at all levels of detail.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*, 4th Ed. New York, NY: Garland Science.

Anderson, P. 1999. Complexity Theory and Organization Science. *Organization Science*. *10*(3): 216-232.

Axelrod, R. A., & Cohen, M. D., 2000. *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. New York: The Free Press.

Gardner, J. (1991). *The Art of Fiction*. New York, NY: Knopf Publishing Group. p. 184.

Jeffrey, H. J. (2009). Mathematical Description of Biological States, Structures and Mechanisms. *International Journal on Advances in Life Sciences, International Academy, Research, and Industry Association* (http://www.iaria.org/).

Jeffrey, H. J., Schimd, T., Zeiger, H. P, and Putman, A. O. (1989). LDS/UCC: Intelligent Control of the Loan Documentation Process. *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.

Jeffrey, H. J. and Putman, A. O. (1983). MENTOR: Replicating the Functioning of an Organization. in K. E. Davis (Ed.), *Advances* *in Descriptive Psychology*, (Vol. 3, pp. 243-270). Greenwich, CT: JAI Press.

Ossorio, P. G. (1971/1975/1978/2005). *"What actually happens": The representation of real world phenomena. The collected works of Peter G. Ossorio, Vol. IV.* Ann Arbor, MI: Descriptive Psychology Press. Also published Columbia, SC: University of South Carolina Press, 1975, 1978. (Original work published 1971 as LRI Report No. 10a. Whittier, CA & Boulder, CO: Linguistic Research Institute. Later listed as LRI Report No. 20.)

Ossorio, P. G. (2006). *The Behavior of Persons.* Ann Arbor, MI: Descriptive Psychology Press.

Putman, A. O. (1981). Communities. in K. E. Davis (Ed.), *Advances in Descriptive Psychology*, (Vol. 1, pp. 195-209). Greenwich, CT: JAI Press.

Voet, D. and Voet, J. (2004). *Biochemistry*, 3rd Edition. Hoboken, NJ: Wiley International.

Wikipedia (2009a). http://en.wikipedia.org/wiki/Complexity

Wikipedia (2009b). http://en.wikipedia.org/wiki/Structure

Wikipedia (2009c). http://en.wikipedia.org/wiki/Hausdorff_distance.

Author's Note

I thank Tony Putman and Paul Zeiger for their helpful comments and suggestions.