Gregory H. Leazer – University of California, Los Angeles, USA
Robert Montoya – Indiana University, USA
Jonathan Furner – University of California, Los Angeles, USA

Numerical Classification and Complexity
Developing a Classification of Classifications

Abstract:
The difference between monothetic and polythetic classification is well established in the literature (Sneath, 1962; Sokal & Sneath, 1963; Needham, 1975). Monothetic classification defines a group such that all members share specific common features, and that, at least in regard to their defining characteristics, any member of the group is substitutable for another. Polythetic groups, on the other hand, are "composed of organisms with the highest overall similarity, and this means that no single feature is either essential to group membership or is sufficient to make an organism a member of the group" (Sneath, 1962, p. 291). Numerical classifications, per Sokal and Sneath, are defined as a type of polytheticism. We argue that polythetic and numerical classification are not coterminous, and that all three classifications vary along an axis of complexity. Distinguishing characteristics of complexity include the number and nature of membership criteria, the internal structure of a classification, and the nature of consensus used in determination of a classification.

1.0 Introduction: Monotheticism, Polytheticism, and Complexity

The difference between monothetic and polythetic classification is well established in the literature (Sneath 1962; Sokal and Sneath 1963; Needham 1975). Monothetic classification defines a group such that all members share specific common features, and that, at least in regard to their defining characteristics, any member of the group is substitutable for another. Polythetic groups, on the other hand, are "composed of organisms with the highest overall similarity, and this means that no single feature is either essential to group membership or is sufficient to make an organism a member of the group" (Sneath 1962, 291). Monotheticism is commonly attributed to Aristotle (*Topics*; see, e.g., Hjørland 2017); polytheticism a creation of Wittgenstein (*Philosophical Investigations*). Lakoff also attributes polytheticism to the work done by Rosch in psychology, a body of work she called "prototype theory."

In what ways can a classification be said to be *complex*? On the one hand, classification generally represents a reduction of the universe—a simplified model—by substituting a smaller number of kinds for a larger number of individuals. The substitution of types for tokens also foregrounds important or essential features—*definitional characteristics* in the parlance of monotheticists—so that not merely the number of things perceived is reduced, but the act of perception itself is concentrated on meaning-bearing features, with other features rendered to the background or discarded as inessential or non-salient. Finally, the identification of essential or salient features allows for cognitive economy. Categories (or the prototypes that reside at the center of polythetic groups) foreground a series of quick mental operations including recognition, memory, and making inferences regarding classified individuals or objects. For example, classifying something as *metallic* allows for expedited reasoning regarding the object's look and feel, its hardness and weight, durability, and the possibility that it could be magnetic. Exceptional metals, such as mercury or sodium–potassium alloy (NaK), which are liquid at room temperature, are handled as mental exceptions. Such categories, including *flightless birds*, while exceptions, still represent simplifications of the world insomuch as a general category and its exceptions is a reduction in complexity from mentally listing all the subordinate species of birds, much less the extraordinary social and psychological burden of treating all birds as unspeciated individuals.

While a classification is by its nature is a simplification, it can be simple or complex in relation to another. Monothetical classification, as a generalization, can be said to be simpler than polythetic classification. Once we have consensus regarding membership criteria for a class, the process of assigning individual cases to the class can be relatively simple. For example, osteoporosis is defined as "A value of [bone mass density] 2.5 standard deviations or more below the young adult mean (T-score ≤ –2.5)" (WHO Scientific Group on the Prevention and Management of Osteoporosis 2003, 57), and further provides descriptions of standard diagnostic procedures and reference standards for generating T-scores. Clear definitional criteria provide simple binary membership rules. The internal organization of monothetic classes are also comparatively simple. Once an individual meets membership criteria, they are in the class, and all members of the class are substitutable for each other in reference to those definitional characteristics.

This relative simplicity of monothetic classifications isn't to suggest that all such classifications are easy to construct. In fact, because the membership criteria are of high consequence—disallowing for partial membership at the margins of the class, for example—it can be difficult to specify those criteria so that the resulting class includes what we want to include, and excludes what we have habitually or tacitly excluded. Bennett (1980) and Gould (1981) discuss the construction of the class "zebra" which traditionally refers to several extant and extinct species of the genus *Equus*. Gould asks whether "they form a true evolutionary unit" (6) and finds that the class as popularly constructed may not meet the cladistic requirements derived from evolutionary theory. How do we proceed? Do we admit that we have irreconcilable differences on the construction of the class and arbitrarily choose one method for developing definitional characteristics, or do we gerrymander the class in such a way that it includes what we want to include and rig the rules of definition so we get the right result?

And just because we have definitional characteristics doesn't mean they are easy to apply. The Miller test (United States Supreme Court 1973) provides the following definitional criteria for the class "obscene materials" (24):

a. whether "the average person, applying contemporary community standards," would find that the work, taken as a whole, appeals to the prurient interest …;
b. whether the work depicts or describes, in a patently offensive way, sexual [or excretory] conduct specifically defined by the applicable state law; and
c. whether the work, taken as a whole, lacks serious literary, artistic, political or scientific value.

To be considered obscene, all three criteria must be met. The third criterion, known colloquially as the SLAPS test, is, like many legal definitions, notoriously complex in interpretation and application, requiring an assessment (*inter alia*) of literary value. The U.S. Internal Revenue Code of 1986 (United States 2018) is primarily a classification that assigns people into tax brackets—at 3,842 pages, it can only be characterized as a highly complex monothetic classification.

However, despite their potential complexities, monothetic classifications are generally simpler than polythetic classifications. Like a good theory, monothetic membership criteria are designed to be parsimonious in nature and limited in number, and when well designed, are made with an effort to find clear and meaningful distinguishing markers of membership. Monothetic criteria are designed to yield clear outcomes—either in or out—and therefore the internal structure of a class is monotonic: in regards to membership criteria, each member of a class is substitutable for another.

The standard description of polythetic classification is given in Wittgenstein's notion of family resemblances which describes a nuanced yet relatively simple form of polytheticism based on a limited number of generally well-understood though perhaps hard-to-measure criteria. "We see a complicated network of similarities, overlapping and criss-crossing: similarities in the large and in the small … I can think of no better expression to characterize these similarities than 'family resemblances': for the various resemblances between members of a family—build, features, color of the eyes, gait, temperament, and so on …" (Wittgenstein 1953/2009, §66–67). The notion of family resemblances is perhaps understood as a discrete and small number of criteria whose similarities are best intuited rather than measured. Compared to classical monothetic groupings, polythetic classes are more nuanced and complex. Definitional criteria are potentially larger in number, with an appeal to a gestalt of formal patterns of overlapping criteria. Because not all definitional criteria are universally present and are understood best in relation to each other, the internal structure of a class generally includes membership gradation, with central and marginal members, and fuzzy membership boundaries. The distinction between monothetic and polythetic classifications along an axis of complexity is pretty straightforward: discrete definitional criteria, all of which must be present, versus a shifting array of definitional characteristics, incorporating nuances of similarity and difference.

2.0 The Construction of Numerical Classifications

Sokal and Sneath define *numerical classification* as "the grouping by numerical methods of taxonomic units based on their character states" (1973, xii). For Sokal and Sneath, numerical classification is a type or subset of polytheticism: it is polytheticism with empirical observations, concrete measurements, and statistical assessments of similarity and correlation—and an appeal toward scientific justification, or alternatively, scientific pretense. Montoya (in press) states that numerical classification is the grouping of entities (organisms, documents, data, etc.) using quantifiable measures of evaluation (such as traits, terms, values, etc.). Numerical taxonomy includes the processes of selecting representative entities, weighting entity values, clustering entities based on these values, relating entities into clusters, and including them into a system based on uniform theoretical commitments. Strictly speaking, numerical classification entails that membership criteria be expressed in numerical form, and the monothetic classification defining osteoporosis is precisely of that nature.

We argue that while monothetic, polythetic, and numerical classifications contain simple and complex examples, they are not coterminous, and they can be viewed as existing at different positions along an axis of complexity. The purpose of our paper is to lead to the development of a classification of classifications. One dimension of this classification is to consider the *complexity* of the similarities or resemblances that form the basis of a classification. Eventually we will discuss additional modalities of classification, including Hjørland (1998; 2017) who differentiates classifications by their "methodology" of class formation, or the various kinds of theoretical commitments implicated by various forms of classification. *Queer classification* in particular is a useful concept for understanding the implications of those theoretical commitments, and emphasizes membership criteria that are semiotic (i.e., interpretive and socially constructed) in nature. However, in this paper we will present an initial assessment of complexity, in relation to monothetic, polythetic, and numerical classification, to determine whether complexity is a useful criterion of comparison for classifications generally. Finally, we also believe that numerical classification is perhaps the least understood and least theoretically integrated into the literature of knowledge organization, and so the bulk of our remaining comments will be aimed to address it primarily.

2.1 Assessing the Complexity of Numerical Classification

While we have defined numerical classification above as strictly being numerical in nature, it has a number of additional features as it is presented in its canonical form by Sokal and Sneath (1963). The features used in numerical classification—which is aimed at the identification of species and their evolutionary development—are not merely *numerical* but also *numerous*. For example, Bennett (1980, 273–274) identified 21 features in her analysis of zebras, including, for example:

1. Number of functional digits;
2. Degree of isolation of protocone and hypocone;
3. Presence and size of preorbital facial fossa;
4. Presence and degree of development of secondary infundibular fold on I2;
5. Size and position of inferior canines;

...

21. Presence of opisthotic dolichocephaly.

Common uses of numerical classification incorporate large numbers of observations which form the basis of similarity that can only be assessed statistically via complex computational techniques. The feature sets can be quite large, and generally a feature is included if it is hypothetically relevant to class formation. Because features have uncertain relationship to any class, some features may be non-salient, and even redundant, in the sense that they strongly co-vary with another feature, marking the possibility that both features might be the consequence of some unidentified latent feature. These all indicate complexity in a potential classification, and are in contrast with polythetic classifications where features may be difficult to explain but all have at least tacit relevance to the class in question.

2.2 Consensus

How consensus functions in numerical classifications depends, in part, upon *when* in the process consensus is being implemented as a rubric for structuration and decision-making. If we look to consensus classifications in the biological or biodiversity world, as we see in the Global Biodiversity Information Facility (GBIF), for example, consensus is often used as a mechanism to create "taxonomic backbones" upon which data points for various species are appended (2019). In this case, consensus is used as a mechanism to provide a generically-agreed upon taxonomy that can then serve as an organizational and access mechanism for species data that, at their points of origin, may or may not have been contextualized within taxonomies with the same philosophical or methodological commitments. Consensus backbones essentially present a *universal* structure to avoid the inevitable conflicts between one taxonomic opinion and another. The same kind of universal approach is used in many bibliographic systems as well, though often not in automated ways—the *Dewey Decimal Classification* (*DDC*) system, for example, uses disciplinary subject partitions to organize documents. The result is that with the *DDC*, as in GBIF, some class decisions are counter to some pockets of expert opinion. The reorganization of the *rosid* family of angiosperms is a case in point (Green and Martin 2013). Due to the prevailing popularity and use of phylogenetic approaches, the *DDC* found enough warrant to change the schedule to reflect new scientific approaches. Yet despite the fact that phylogenetic approaches are now *preferred* and *accepted*, this same schedule must be used to organize documents even *not* subscribing to this schema. Consensus as an organizational approach defers to the majority knowing that the scientific world of opinion is plural. Given this reality, entities placed within consensus systems should be understood to have several distinct, and perhaps conflicting, identities: on the one hand, they have their position within the consensus taxonomy, juxtaposed with other entities within an environment that is ostensibly more global in nature; on the other, one also must understand that, at its point of origin, that same entity may or may not have been constructed or contextualized on the same ontological terms. Complexity increases if we cannot wrest these two identities apart from one another.

So, while the *DDC* is certainly consensus-based, a body of editors, as well as particular cited evidence, or warrant, can be pinpointed as the source of this change—and thus is also a source of bias in a classification's construction. In the case of GBIF, however, and other automated synthetic systems, consensus decisions are not so easily visible, nor are the arbiters of this change identifiable. GBIF's backbone taxonomy is "updated regularly through an automated process in which the Catalogue of Life acts as a starting point also providing the complete higher classification above families" (GBIF, 2019). It is precisely this "black box" of automation that makes numerical approaches especially complex, more difficult to understand. If we think of a more embedded automated system such as the Google search engine, we see this phenomenon clearly: there is no way to understand how algorithmic techniques are intervening to propose certain top-level classes for a seemingly-endless corpus of online documents. In these spaces we can ask, Why do these terms (or traits, or phenomena, or entities) mean more than any other? How are entities and traits valued in relationship to one another? And, What were the alternate possibilities by which these results could have been interpreted to an equally-valid state?

Google is also a prime example of how the question of *when* consensus intervenes becomes important. For Google, or any other dynamic searching or retrieval mechanism, classifications are dynamic, which means that the principles for construction, quantities being examined, and the resulting classes, differ each time we query a classification. For example, searching for a complex phrase in Google one day can provide a different set of results than another day. This is because the commitments for classification construction and the body of possible entities are changing. Humans also intervene in these algorithmic structures in ways that we cannot totally understand. Safiya Noble made this readily apparent in her work on race and algorithmic power (Noble 2018). In 2012, Noble published an article in *Bitch* magazine noting the marginalization and racial classification of "black girls" and women on Google's interface. "By August 2012," Noble states, "Panda (an update to Google's search algorithm) had been released, and pornography was no longer the first series of results for 'black girls'; but other girls and women of color, such as Latinas and Asians, were still pornified" (2018, 4). In cases like these, automated consensus mechanisms, and the constant rate at which they are applied, confound our ability to understand them: even *if*—and that is a big *if*—we manage to understand the logic of classification at one moment, the results may have very little bearing a week later.

Returning to the notion of consensus, it is necessary to state an obvious fact: consensus is *not*, despite rhetoric to the opposite, equatable to universal *agreement*—at least not in the case of classification. Any one person can contest decisions made through automated means and, in fact, a critical approach to this work would support and popularize this approach. We can perhaps go so far as to say that consensus may be more-or-less equivalent to *authority*, so far as we, the users, acquiesce in some way to the fact that so-and-so system will be authoritative in one situation or another. GBIF, for example, is an authoritative source for data, but it certainly makes no claims about agreement within the scientific community about the taxonomic perspective it proliferates. Problems arise, again as Noble shows, when the authority of systems becomes authoritative without a sense of critical analysis. Google results should hold no authority on the question and the formation of our racial, ethnic, or cultural identities, and yet, this is precisely how they are being used whether purposefully or not. This is a critical point to understanding, and limiting, the impending wave of classificatory systems resulting from the application of artificial intelligence solutions to big data, for example, in "smart city" projects. The complexity of data and the sophistication and apparent neutrality of algorithms result in decisions that bear the authority of the only-partially understood classificatory regime but whose actual heuristics and resulting classificatory decisions make little sense, and which fail to provide justification or alternative possible constructions. Unquestioned algorithmic complexity is a dangerous social reality and, as such, it makes good sense to delineate what we do and do not understand about these systems of organization.

2.3 Issues of Complexity: What We Know and What We Know We Don't Know

Classification and knowledge organization's long history of scholarship gives a good grounding to understand some of the basic known factors about classification that we understand to be fluid, contended, and arbitrary. The space here is not sufficient to mention them all, but some basic issues can be identified as they relate to numerical classification. When thinking about the quantification of factors in numerical and algorithmic taxonomic methods, we know well that the values we apply to attributes or entities are, if well intended, arbitrary nonetheless. Let us take the example of a phenetic classification of pine trees—a classification based on formal physical characteristics. There are some quantities that we might find important: that they are evergreen, the texture and structure of their bark, dimensions and characteristics of cones, needle count and position, height, etc. That these qualities are used to classify a pine tree is arbitrary in that we could have identified any number of what might be considered non-essential characteristics: flexibility in the wind, utility as fire-wood, etc. Likewise, when we think of the numerical classification of documents, a system may use terms, co-term prevalence, document source, authorship, keywords, etc. In documentary analysis this aboutness is essential to description but also evasive in terms of method identification.

And then there are factors for classification that may be difficult to identify and measure. It wasn't until genetic testing could identify sequences of importance (the COI or COX1 "barcoding" gene, for example) that phylogenetic approaches opened the door for revolutionary taxonomic methods in the biodiversity sciences, for example. We could finally "measure" organism classes in a way that was "universal" and replicable. In the bibliographical world, the notion of relevance (Wilson 1973) has always been identified as central to information retrieval and selection, and yet truly quantifying relevance in a way that meets the searching criteria for infinite moments of need still evades us. Relevance is the primary goal of search engines given that searches are explicitly intended to satisfy some situational need. And so we know that there are some qualities that are obvious, some quantities that are difficult to define, and in both cases, what we choose to seek out is wholly arbitrary based on our assumptions about the world, our ontological commitments, and our contextual purposes for organizing.

On top of the qualities we use to class entities, we must also insert our own hermeneutic skills to interpret their meanings as they relate to one another. Relationships are neither given, nor obvious, and will always depend on the context in which they should function. "To specify a relationship, we may first designate all the parties bound by the relationship (hereafter referred to as the participants in the relationship) and then specify the nature of any relationship that binds them together" (Green 2008). This means that relationships made by a scientist using phylogenetic methods, for example, will be based on specific and arbitrary quantities and articulated in equally arbitrary thresholds for a given set of taxa (even if the decisions are evidence based and properly "scientific"). But these constructed relationships are not natural relationships: they are imposed interpretive frames. In the end, why we make any given decisions can be based on clear guidelines or can be based on tacit or unconscious factors. In phylogenetics, clear mechanisms to distinguish one distinct species from another can be identified, along with the thresholds used to assess taxa. When Francis Galton was using composite photography to classify types of criminals in the last quarter of the 19th-century, however, it is clear that racial factors were taken into consideration. The history of the classification of race and humans is riddled with these conscious and unconscious biases (for example, see Smith 2015).

Some unconscious biases impact classifications that are less easily identifiable. Ontological commitments, for example, are sometimes difficult to archaeologically unearth in certain biological taxonomies without the producer on hand to explain certain decisions. Numerical classifications add a layer of complexity onto this that is significantly more complicated: the fact that statistical models are both mathematically complex and difficult to reverse engineer to understand class partitions at any given point in time. Once again, looking at Noble's work (2018), an essential problem with automated organization is that it becomes very difficult to identify the location(s) of error when assessing a given set of results. Error in this sense is multi-valenced; it is locational (as in, there is ostensibly a code location and directive to locate), temporal (when, in fact, did this decision occur in a long scale of decision locations?), and multivariable (what variables or quantities were being referenced at that particular point and time?).

3.0 Function

Our paper has attempted to appraise three kinds of classifications in terms of their complexity. Monothetic, polythetic and numerical classifications are judged to be increasingly complex, by virtue of the number and nature of their definitional characteristics, their internal structure, the nature of consensus in their formation, and the comprehensibility of their resulting classes.

However, as we have also noted, there are simple and complex examples within each type of classification. While we believe that monothetic classifications are generally simpler that their polythetic and numerical counterparts, such a conclusion may be shaped in the way we have generally understood each type of classification. By choosing Sokal and Sneath, for example, to represent the canonical form of numerical classification, we may have unwittingly opted into a more complex version of that classification. We would not typically use the taxonomy of species as the primary and nearly exclusive use of any one type of classification.

Additionally, we are certain that we have not located all the various modalities of classificatory complexity. Definitional characteristics and the nature of consensus have figured prominently in our previous work on classifications, but there are certainly other salient facets to complexity that we have not yet explored. For example, we know that the automated and numerical classifications that predominate in common web-based applications are also closed and proprietary (e.g., Pasquale 2015), making them inherently more complicated to assess. This paper should be viewed as an initial foray into assessing the complexity of classification.

Additionally, we have been operating under an assumption that the three kinds of classifications presented here are substantially different in kind. There may yet be a theory that, for example, presents them each, in sequence, as a generalization of a previous model. That is, perhaps, that monothetic classification is a more specific kind of polythetic classification, with more precise definitional characteristics. That a given classification might be more complex may not, strictly speaking, mean that it's actually different.

Finally, we need to consider that complexity ultimately may not be the right dimension for explaining differences amongst classifications. Another characterization, one that partially replicates the criteria used here or even one entirely novel, might be more effective at differentiating and understanding classifications. Complexity is simply our first, best stab at trying to understand classifications. Numerical classifications, in their automated and recent guises, are relatively entries in the history of classification, and accommodating them into our understanding of classifications generally is an unfolding process. Over time, the novelty of numerical classification—and the significance of its differences with antecedent models—may fade, and numerical classification may eventually be viewed as the same as its polythetic cousins. But right now, as automated approaches emerge in novel and not entirely welcome ways, with uncertain social and political consequences, we endeavor to understand what is new, and what is the same, with numerical and automated approaches to classification. Numerical classifications certainly feel inventive, riskier, and more complex. Their deployment, arriving as they do without complete user comprehension as to the nature of their operation, represents a new period in the history of classification, and their complexity masks uncertainty in the consequences of their use.

References

Bennett, Debra. 1980. "Stripes Do Not a Zebra Make, Part I: A Cladistic Analysis of Equus." *Systematic Biology* 29, no. 3: 272–287.

Global Biodiversity Information Facility. 2019. *GBIF Backbone Taxonomy*. Copenhagen, DK: GBIF Secretariat. https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c.

Green, Rebecca. 2008. "Relationships in Knowledge Organization." Knowledge Organization 35, no. 2/3: 150–159.

Green, Rebecca and Giles Martin. 2013. "A Rosid Is a Rosid Is a Rosid . . . or Not." *Advances in Classification Research Online* 23, no. 1: 9–16. https://doi.org/10.7152/acro.v23i1.14228.

Gould, Stephen Jay. 1981. "What, If Anything, Is a Zebra?" *Natural History* 90, no. 7: 6–12.

Hjørland, Birger. 1998. "The Classification of Psychology: A Case Study in the Classification of a Knowledge Field." *Knowledge Organization* 25: 162–201.

Hjørland, Birger. 2017. "Classification." *Knowledge Organization* 44: 97–128.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Montoya, Robert. In press. "Numerical Classification." *Knowledge Organization*.

Needham, Rodney. 1975. "Polythetic Classification: Convergence and Consequences." *Man* 10, no. 3: 349–369.

Noble, Safiya Umoja. 2012. "Missed Connections: What Search Engines Say about Women." *Bitch*, no. 54 (Spring): 36–41.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by Eleanor Rosch and Barbara B. Lloyd. Hillsdale, NJ: Lawrence Erlbaum Associates, 27–48.

Smith, Justin E.H. 2015. *Nature, Human Nature, and Human Difference: Race in Early Modern Philosophy*. Princeton, NJ: Princeton University Press.

Sneath, Peter H.A. 1962. "The Construction of Taxonomic Groups." In *Microbial Classification: Twelfth Symposium of the Society for General Microbiology*, edited by Geoffrey C. Ainsworth and Peter H. A. Sneath. Cambridge: Cambridge University Press, 289–332.

Sokal, Robert R., and Peter H.A. Sneath. 1963. *Principles of Numerical Taxonomy*. San Francisco, CA: W. H. Freeman.

United States. 2018. *U.S. Code: Title 26, Internal Revenue Code*. https://www.govinfo.gov/content/pkg/USCODE-2018-title26/pdf/USCODE-2018-title26.pdf.

United States Supreme Court. 1973. *Miller v. California. 413 U.S. 15 (1973)*. https://cdn.loc.gov/service/ll/usrep/usrep413/usrep413015/usrep413015.pdf.

WHO Scientific Group on the Prevention and Management of Osteoporosis. 2003. *Prevention and Management of Osteoporosis*: Report of a WHO Scientific Group. Geneva, Switzerland: World Health Organization. https://apps.who.int/iris/handle/10665/42841.

Wilson, Patrick. 1973. "Situational Relevance." *Information Storage and Retrieval* 9, no. 8: 457–471.

Wittgenstein, Ludwig. 1953/2009. *Philosophical Investigations*, 4th ed., trans. G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte. Malden, MA: Wiley-Blackwell.