Changing Perspectives on Classification as a Knowledge-Representation Process †

Barbara H. Kwaśnik

Syracuse University – School of Information Studies, Hinds Hall,
Syracuse New York 13244-1100, United States, <bkwasnik@syr.edu>

Barbara H. Kwaśnik taught in the areas of knowledge organization, theory of classification, and information science. Her main research interests are formal aspects of classification structure, personal information management, browsing, and how classifications intersect with everyday human endeavor. Previous research, with professor Kevin Crowston, included investigating whether genre information can help in searching. Most recently she is working with Brian Dobreski on the issues of personhood in knowledge representation. She retired in 2018 from the School of Information Studies at Syracuse University, Syracuse NY.

Kwaśnik, Barbara H. 2019. "Changing Perspectives on Classification as a Knowledge-Representation Process." *Knowledge Organization* 46(8): 656-667. 37 references. DOI:10.5771/0943-7444-2019-8-656.

Abstract: No matter how immutable a classification may seem, it is, after all, an artifact of the human imagination and functions in a particular place and time. The author describes her personal inquiry into classification as a knowledge-representation process. She traces her changing perspectives on how classifications should be viewed and evaluated by posing the following questions: 1) How does the classification process enable or constrain knowing about something or discovering something we did not already know?; 2) In what ways might we develop classifications that enhance our ability to discover meaningful information in the information stores that form a part of our scholarly as well as our everyday lives?; and 3) How might classifications mask or distort knowledge, and how might they serve to disenfranchise people and ideas? These questions are considered through a discussion of classification structures, personal classification, the link of classification to theory, everyday working classifications, translation of classifications, cognitive aspects, browsing, genres, warrant, and the difficulties of navigating complex ontological commitments. The through thread is the importance of context, because classifications can only be seen with respect to the human endeavors that generate them.

Received: 13 November 2019; Accepted: 16 November 2019

Keywords: classification, classifications, information, knowledge

† In writing about my research trajectory, I was repeatedly brought up by the awareness of how many people were involved. Here's my opportunity to thank all of them: not only the groundbreakers who were, and continue to be, my intellectual inspiration, but also my professors and dissertation committee at Rutgers University, headed by James Anderson, as well as the colleagues on a day-to-day basis, the friends who bore my annoying habit of viewing just about any issue as a classification problem. I have been fortunate that my research and teaching at the School of Information Studies at Syracuse University have been closely intertwined, which is an increasingly rare privilege. The engagement of hundreds of students who brought so much to my classes was the source of my best examples and my deepest reflections. The International Society for Knowledge Organization and the Special Interest Group for Classification Research in ASIS&T were my home where I was welcomed as a young scholar and encouraged to develop my own trajectory. The influences and opportunities for a life of the mind were constant and enveloping. My most sincere thanks, though, go to my doctoral student co-authors, who participated with me in exploring interesting questions and helping me translate what we learned into the work summarized in this essay. My work on genres with Kevin Crowston was funded by an NSF grant: Crowston, Kevin (PI). How Can Document-Genre Metadata Improve Information Access for Large Digital Collections? NSF IIS Grant 04–14482, Jan. 2005 – Jan. 2007.

1.0 Introduction

We are incapable of "not" classifying. There are classifications everywhere, both formal and *ad hoc*; both enduring and ephemeral. After decades of studying and teaching I now understand better how classification both in and outside the library is not foremost about being tidy, but rather, about having a tool for seeing the world and understanding it. Our efforts at creating and using classifications serve to enable more efficient communication; they show patterns; they help visualize an overall view by showing clusters and areas of density and gaps. There are decisions made about "first cuts," scope, definition, and relationships among the parts, all of this rendering a representation of some area of our lives and what we know about it.

My formal interest in classification goes back at least four decades, when as a beginner cataloger I realized that behind the library schemes used to organize resources physically on shelves was a world of intellectual richness and complexity that barely tipped the surface. The ques-

tions I asked then persist: 1) How does the classification process enable or constrain knowing about something or discovering something we did not already know?; and, 2) In what ways might we develop classifications that enhance our ability to discover meaningful information in the information stores that form a part of our scholarly as well as our everyday lives? (Kwaśnik 1999). To that I have added, over the years the thorny questions of how classifications may mask or distort knowledge, and how they might serve to disenfranchise people and ideas. In the following sections I describe the various threads of my inquiry into classification as a knowledge-representation process and my changing perspective on the process and the outcomes. It is not a straight line, even though I must present it that way. Rather, the various aspects inform each other and lead to some insights, but also to fresh questions that as of now I am not able to answer.

2.0 Knowledge and classification

I draw a distinction between merely observing, perceiving, or even describing things and truly knowing them. To know implies a process of integration of facts about objects and the context in which the objects and processes exist (Kwaśnik 1999, 23). It is not enough to know about the things but, rather, it is the relationship of one thing to another that creates the deeper understanding. We are familiar with how knowledge discovery and creation in the sciences may follow traditional processes, such as exploration, observation, description, analysis, and synthesis, as well as testing of phenomena and facts. It may also follow more interpretive paths, which are, nevertheless, based on evidence that is collected and appraised following consensual norms. All of this is conducted within the communication framework of a research community with its accepted methodology and set of techniques (Kwaśnik 1999, 23). Bronowski (1978) describes how even in the empirical sciences, though, the process is not entirely rational but is often sparked and then fueled by insight, hunches, and leaps of faith. Moreover, research is always conducted within a political and cultural reality (Olson 1998).

In my early thinking in the 1980s, I was drawn to the classification work in the sciences, not because I thought it was worthier of being prioritized, but because it is linked to underlying theories or conceptual frameworks. This link to theory seemed a very important one to me since it rendered the classification as somehow more substantive, more enduring. I have since modified this view, but at the beginning of my career I assumed that the more theoretical a classification was, the better it was at representing knowledge. I agreed with Kaplan (1963) who said that theories and models are a "symbolic dimension of experience as opposed to the apprehension of brute fact" (294). In the same way, I thought a classification that is thrown together without any conceptual glue to hold it together is typically wobbly as a knowledge-representation structure. Overall, I believed that, indeed, a classification itself could be construed as a theory—a structure of entities and their specified relationships (Kwaśnik 1994). Whether it was good or flawed depended on the value of the conceptual scaffolding.

Following on this analogy of classification to theory, I observed that the process of classification can be used in a formative way during the preliminary stages of inquiry as a heuristic tool in discovery, analysis, and in theorizing (Davies 1989). Once concepts gel and the relationships among them become understood, a classification can be used as a rich representation and is thus useful to communication and in generating a fresh cycle of exploration, comparison, and theorizing. Kaplan (1967) states that "theory is not the aggregate of the new laws but their connectedness, as a bridge consists of girders only in that the girders are joined together in a particular way" (297). I believed that a classification works in much the same way, connecting concepts in a useful structure. If successful, it is, like a theory, descriptive, explanatory, heuristic, fruitful, and perhaps also elegant, parsimonious, and robust (Kwaśnik 1994).

This led me to believe that classification is somehow a fundamental aspect of nature, sort of like the Fibonacci Sequence, and that the "lawful" integration of theory and classification yielded the most robust schemes. There are examples to suggest this, the periodic table of the elements being one. Here we have a classification that has endured through different theoretical explanations and continues to be useful to this day. The scheme, which started in the nineteenth century as an observation of the regular change in atomic weight among elements, eventually yielded a pattern that led to the discovery of new elements, and through the lens of increasing theoretic understanding an explanation of why this occurred. From my perspective, it was a comforting thought that while the principles of a classification could evolve, the underlying concept remained solid.

I have come to believe that this might be a dangerous belief in the sense that not everything sorts itself out as beautifully as the elements in the periodic table seem to do. In fact, deeper analysis reveals beyond a doubt how any classification, no matter how immutable it may seem, is after all an artifact of the human imagination and functions in a particular place and time. For many phenomena there are many ways to interpret, visualize and explain them, even my revered periodic table. The value of any one approach depends on many factors, including the context in which the classification is being invoked. Putting up any one classification as ideal implies that classification schemes that do not measure up are somehow inferior rather than simply different, or that they do not have a useful purpose. While many traditional classifications have strength and merit, I have

come to realize the exclusive embrace of elite classifications and classificatory structures may build in subtle assumptions and biases, making it more difficult to admit diverse perspectives.

3.0. Classification structures

When I first investigated classification as a research topic the question I posed was, "How do we systematically evaluate a classification?" I wanted to step back and impartially analyze what made a classification tick—any classification. What are the features that provide strength in terms of knowledge representation, and what features misguide or interfere with it? I was additionally motivated by my need to develop techniques for teaching about classification in a way that gave students a toolkit they could use to describe and evaluate any classification that came their way in a thoughtful and careful way, especially the legacy tools they would be using in their work. An important part of this profile was a description of the structural properties of the classification, how the parts worked together in terms of relationships. There are many kinds of classificatory structures, but for this review I will discuss the two that are probably most familiar: hierarchies and trees.

3.1 Hierarchies

I start with hierarchies, because they are perhaps the most recognizable (and perhaps the most misunderstood). People call all kinds of things hierarchies, but here I refer to the logical structure we have inherited from Aristotle (1963). His view was that, after careful observation, one could learn how things could be aggregated and differentiated "naturally." It assumed the division and aggregation of classes would be valid, because one had to arrive at the "essential qualities" of what was being classified. Yes, we now question that such structures require us to choose one ideal representation over possible others, but nevertheless, hierarchies are often sought out as the structure of choice for their many strengths described below.

Essentially, a hierarchy is a structure with a top class that defines the scope of the classification. The top class includes all its subclasses and sub-subclasses. A true hierarchy has only one type of relationship, the "species/differentia" relationship, also known as the "generic" relationship, or informally the "is-a" relationship. The strict requirement for inclusion ensures that what is true for the top class is true for all the subclasses. This property is called "inheritance," that is, attributes are inherited by a subclass from its superclass. "Transitivity" is an important outcome of this carefully controlled structure, because one can assume that all classes are members of not only their immediate superclass but of every superclass above that one.

Hierarchies have predetermined, predictable, and systematic rules for association and distinction. There are "necessary" and "sufficient" conditions for when something may belong to one class and be distinguished from another class. The rules attempt to use the most "essential" type of information for distinguishing one class from another, and thus, entities differ from sibling entities in a predictable way. Often the rules are based on some theory that is the foundation for the classification. Finally, a hierarchy invokes the rule of "mutual exclusivity," which means that each entity can belong to one and only one class.

To take one example, in western medicine, concepts lend themselves to hierarchical arrangement. Here's an abbreviated snippet from the National Library of Medicine's *Medical Subject Headings* (*MeSH*) (https://meshb.nlm.nih.gov/record/ui?ui=D013494).

Nervous System Diseases [C10]
  Central Nervous System Diseases [C10.228]
  Brain Diseases [C10.228.140]
  Basal Ganglia Diseases [C10.228.140.079]
    Basal Ganglia Cerebrovascular Disease [C10.228.140.079.127]
    Chorea Gravidarum [C10.228.140.079.294]
    ....
    Parkinsonian Disorders [C10.228.140.079.862]
    Supranuclear Palsy, Progressive [C10.228.140.079.882]
    Tourette Syndrome [C10.228.140.079.898]

In this example, searching for the rare disease progressive supranuclear palsy, we learn that it is part of several linked hierarchies. One is shown here, under nervous diseases. The information flows in many directions: from the top terms down, from the "sister" terms, and also from the related terms (such as movement disorders) in other parts of the schedules. It helps the searcher identify the landscape. Knowing that PSP is located near parkinsonian disorders and tourette syndrome helps define the nature of the disease, and indeed, helps explain why it is difficult to diagnose in its early stages.

It is obvious why such careful constructions are appealing in knowledge representation. If valid in their underlying assumptions and definitions they offer complete and comprehensive information. The affordance of inheritance provides for an economy of notation, and perhaps most important, a hierarchy offers the ability to make inferences from incomplete information. For instance, one can infer that a female kitten is, like all cats, a mammal, and by her essential mammalian features could be expected to eventually bear live young and breast feed them, even if she is not at present doing so. Put another way, the hierarchy succinctly carries a great deal of information that can

be used to represent the domain, to explore, and to provide conceptual fodder for further discovery. Above all, a hierarchy is built on logical principles, so to many people it seems trustworthy. The parts must fit together; it should be comprehensive and not have loose ends. Building and maintaining hierarchies requires a strong *à priori* conceptual framework as well as consensus to guide the development of the rules. That is why hierarchies are often deductively created, rather than built from the bottom up.

Not all knowledge domains lend themselves to a strictly controlled hierarchical representation, though. Hierarchies are problematic for a number of reasons. First and foremost, we acknowledge that it is often very difficult to identify the ideal "essential" partition points in any domain of knowledge. Many phenomena can be seen from several perspectives, depending on the context and the goal of the classification. There may be multiple and diverse criteria, and there may be a question of which criterion to invoke first. For example, the traditional classification of animals divides them up into ever more specific taxonomic ranks, from kingdom to species. This order precludes the consideration of the differences between animals in the wild and animals in captivity. Moreover, it is awkward to use this classification to represent ecological systems gracefully, since there are many other factors to consider besides the animal's morphological "essence."

Finally, a true hierarchy requires deep knowledge and consensus about the domain in order to determine the rules for defining classes, partitioning, and aggregation. If there is no conceptual framework guiding these choices, the classification can seem incoherent or contradictory. Thus, in new and emerging fields where knowledge is incomplete, it is sometimes unwise to commit to a hierarchical classification. Even when the field is mature, though, but rapidly changing due to new incoming information, maintaining a hierarchy can be dicey. For instance, we are familiar with the muddy classification of heavenly bodies, such as planets.

The principle of transitivity and inheritance requires that all the entities in a hierarchy be at the same level of conceptual granularity, thus a hierarchy does not accommodate differences of scale for the same phenomenon. For example, a beach might be construed as a kind of land mass as you might see it from space, or it might be a kind of habitat, or it might be viewed as an aggregate of materials, such as sand. A hierarchy encompassing all the views in one structure would be impractical and confusing in terms of making inferences and comparisons.

3.2 Trees

Trees are another familiar type of classificatory structure. A tree divides and subdivides its classes just as in a hierarchy, but the relationship among the classes is not necessarily generic. There are many types of relationships possible, such as part/whole, a kinship tree, or an organizational chart. In a part/whole scheme such as:

> North America → United States → NY State → Onondaga County → City of Syracuse

you can see that Syracuse is a part of Onondaga County, which is part of NY State, and so on, but Syracuse is not a "kind of" county, nor is the county a "kind of" state. This structure limits the transitivity of information, because what is true of North America does not carry down to the city level in terms of shared "essential" features. Similarly, if an engine of a car contains pistons, spark plugs, and valves, you do not have a great deal of information about the relationship between pistons and spark plugs except to know they are both part of a car engine. They can, in fact, be totally different entities. What unifies them in the scheme is their position in the engine.

In a traditional kinship tree, you may describe the flow of who begat whom, but a daughter is not a kind of mother, nor is a mother a kind of grandfather. Instead, the relationships are determined by blood and legal affiliation. In an organizational chart there is a clear purpose, and that is to show "who reports to whom" or perhaps "who is managed by whom." This is not to say that kinship trees or organizational charts do not yield a great deal of information, but they are not as rich and inclusive as the classic hierarchical structure in terms of showing the unity of the whole. Instead, one or two critical relationships are highlighted, which makes them easy to comprehend and analyze along those relationships.

In summary, both hierarchies and trees are useful systematic knowledge-representation structures with different properties and strengths and different constraints. In both you must know about the domain to pre-determine the first cut and the rules for membership in classes. Both kinds of classification structures have many challenges in being applied, though, because the realities of application do not always map well onto the requirements of such structures.

3.3 Facetted classification

My growing awareness that classificatory thinking was both culturally and psychologically influenced led me to further explorations. Among these was facetted classification, which is not a different representation structure, but rather an approach that allows the classifier to view the world as dynamic, and indeed, provisional, in how it is construed. The approach is credited by many to S.R. Ranganathan, an Indian scholar, who posited that any entity could be viewed from a number of fundamental perspec-

tives or facets. He suggested that these are: personality, matter, energy, space, and time (Ranganathan, 1967). While the discussion of the nature of these facets varies in his work, the principle has caught on and endured. The notion that any entity can be analyzed into aspects, each representing some feature or quality, freed up designers to create schemes that were multidimensional. Note this is not the same as breaking down an entity into components, but rather viewing the same entity from different perspectives—same object different views. One of the clearest examples that explicitly built on Ranganathan's principles is the *Art and Architecture Thesaurus,* a compilation of vocabulary for the indexing and retrieval of objects and literature on material culture, which in its diversity defies easy description and classification into any one classificatory structure. The *AAT* allows the creation of a string to represent a topic or object using the core categories of period/style, place, process, material, and object. For example:

19th Century Japanese raku ceramic vase
Arts & Crafts American oak desk

In doing so it is then possible to search by any one of the components (e.g., all things "arts & crafts," or all "vases"), or in combination. It also allows for the graceful addition of new objects and topics, so long as they can be analyzed using the five categories.

This approach has extended well beyond formal collections and is very popular in shopping sites as well as visualized analyses of all kinds. Not all use Ranganathan's principles, but the result is essentially the same. Each facet can be developed following its own logic and structure, and then synthesized into expressive strings. Obviously, the advantages are you do not need exhaustive knowledge so long as you can identify important fundamental "aspects." It is a hospitable and flexible approach without the need to have a strong, immutable theory for the scheme overall. At the same time, it can accommodate a variety of theoretical structures and models in the facet components, and, most important, it can sustain a variety of perspectives. Thus, a flower can be considered as food, as a feature of specific habitats, as a commodity, and so on.

While facetted schemes have pragmatic appeal, there are some things to consider. First, is the difficulty in identifying the core facets. They should be robust and so, while complete knowledge of the domain is not necessary, enough must be understood to accommodate all new entities. For example, you might view the traditional classification of instruments as a facetted scheme, incorporating material, process of creating sound, origin, and so on, but the scheme hits a speed bump when you want to also include electronic instruments. Second, while it is freeing not to have a required binding conceptual framework, this also means that the scheme remains essentially descriptive. There is no guidance for how to read or interpret the relationship "among" the facets. Finally, it may be difficult to visualize all in one grand picture. A scheme might include a timeline, a hierarchy and a tree, but no built-in clue on how these should be presented. Even so, chances are if you look around at modern classification, it will likely as not be a facetted scheme (Kwaśnik 1999, 39-43). In much of my research I've employed the notion of facetted classification, especially in the challenges presented by multidimensional situations.

4.0 Practical classifications

The study of formal classification led me to an appreciation of their formidable power, but also piqued my interest in simple, practical classifications that perhaps deviated from the "ideal." Many of these exist to help with tasks such as shopping, diagnosis, or description without being necessarily bound by a particular theoretical framework. That is, they are not without an underlying rationale, but they do not purport to be "true" or enduring. They are simply there to organize some phenomenon in a useful way. After having scoffed at such schemes I became fond of them, because they demonstrate our human ability to be in touch with the power of a good classification that uses visualizations and simple metrics to help navigate through more complex information. The point of the following examples is that in many ways they provide for enhancement of description and searching where sometimes the formal classifications are lacking or overly complicated. These classifications demonstrate that iterative, flexible design, combined with other search features can be very effective and certainly easier to maintain.

4.1 Keys

My trusty old Peterson's field guide to wildflowers (1974) is an example of such a pragmatic classification that makes identification of wild plants more accessible. It is a key, which is a type of classification that chooses one or two obvious features to lead into the more formal classification. In this case, the plants are organized by petal color, a feature not "essential" in the Aristotelean sense but recognizable. Then, by icons and clear descriptions, the identification can proceed further. In other words, the key is just that, a key. The design of such a key might be frustrating to the botanist because it might seem superficial, but to a novice it is a testament to the communicative power of such tools.

4.2 eBay.com

Commercial websites use classifications in contexts where the content may be in constant flux, the user population is unknown, or if it is known, we can assume it is diverse; and where it is desirable that the classification be very simple and straightforward so that all levels of users can learn it. For instance, eBay.com maintains a classification of millions of objects. It fails miserably when analyzed using the formal criteria of coherence, but it is a classification of current objects—everything on eBay exists and is for sale—it is not meant to endure forever. Despite its rather sloppy design, it is surprisingly robust and hospitable. When you consider that it reflects the terminology used by several million people for an amazingly wide and constantly shifting array of items, it is quite impressive. In terms of accessibility, there are very few terms in the main categories that are difficult to understand. One of the strong points is the meshing of the classification with many other access strategies. If the classification fails, there are other avenues to pursue (Kwaśnik and Liu 2000).

4.3 Amazon.com

Along the same lines, when I studied the amazon.com book division years ago (Kwaśnik, 2002), even back then the affordances of multiple access points ensured that "something" would be found no matter what the user entered. The classification achieved a multi-perspective view allowing for a facetted approach, and if one approach did not work another was readily available. I concluded (284) that "in general, amazon.com's scheme can be viewed as more pragmatic and enumerative than as based on a model of knowledge." It is a classification meant to encourage buying and uses as many routes to the goal as possible, including a simple but redundant vocabulary without much attention to structural integrity but able to provide a rich network of subjects.

4.4 Scientific and naïve classifications

For a final example, we explored the idea of "teaming up" scientific and naïve classifications. We compared two separate but related classification schemes in the area of medical information to better understand how they might be used together and inform one another. We contrasted *MeSH* with the consumer health website, WebMD.com. Using the term "autism" we compared the strengths and limitations from the perspective of vocabulary, syntax and classificatory structure, context, and warrant. We conclude that in terms of vocabulary and concepts, *MeSH* may benefit from WebMD's approach to ongoing updates and currency as well as the contextualization of terms. At the same time, WebMD.com may benefit from some form of vocabulary control for richer expansion of terms and archival retrieval (Kwaśnik and Flaherty 2010).

5.0 Challenging the Aristotelian paradigm

As mentioned, two foundational but possibly conflicting assumptions took shape in my early pursuit of studying classification: the first was my firm belief that formal classificatory structures, such as hierarchies and trees, help advance understanding because of their ability to represent not only the elements of a domain but also the relationships among them, thus yielding knowledge structures that not only describe but also explain. At the same time, I realized that people in their enactment of classification brought their own contextual understanding to them and the two did not always map well.

5.1 Accounting for context in the classification of personal documents

In my dissertation, *The Influence of Context on Classificatory Behavior* (Kwaśnik 1989b, 1991), I wanted to see how people create and then utilize classification, that is, I wanted to learn about personal information management in everyday life. I interviewed university professors in their offices and recorded their documents and the organization of these documents on shelves, in drawers, on the computer, in their briefcases, taped to the door, and in various piles and files, always in their own words. The findings, in a nutshell, were they did not organize things as they are organized in formal library collections by subject and form. In fact, the contextual factors played a critical role—factors such as the purpose of the document or its currency. Documents with the same content or subject could be classified differently depending on how they would be used (Kwaśnik 1989a). The findings suggested that while formal classifications exist and prove very useful, the establishment of universal schemes is more problematic in situations in which context plays a part. So, formal logical representation works when the domain is well understood and there is consensus on the underlying conceptual structure, but what about all the rest?

5.2. Influences of cognitive anthropology

Along the way to finding a conceptual framework for my dissertation I was introduced to the field of cognitive anthropology, and most intriguing to me, category choosing. In a hierarchy, each member of a class is an equally good representative of that class, since each member must possess all the requisite attributes. In a pure hierarchy an entity cannot be "sort of" in a class. The boundaries and rules for inclusion and exclusion are defined and predetermined.

An entity cannot belong to more than one class following the principle of mutual exclusivity. These traditional assumptions came increasingly into question, because anthropological and cognitive evidence from studies of humans did not support them. The formal properties do not necessarily map accurately to our human cognitive processes, that is, we do not all store our concepts in hierarchies. We know that humans have fuzzier notions of what constitutes a boundary on a class of things. These boundaries may change with circumstances and the experiences of the classifier.

5.2.1 Prototype theory and the principle of family resemblances

I was influenced in particular by researchers such as Eleanor Rosch whose prototype theory posits that not all members of a class may be perceived as equally good representatives of that class. Not all birds fall naturally into the class of "birds;" some seem more birdlike and others less so. She posited that for every class, some objects become prototypical and form a best example of that class (Rosch 1973). Another notion that influenced me was the idea of family resemblances (Rosch and Mervis 1975), where they argue that the principle of family resemblance can be construed as "a logical alternative to criterial attributes." They were arguing against (603) "a tenacious tradition of thought in philosophy and psychology which assumes that items can bear a categorical relationship to each other only by means of the possession of common criterial attributes." Their study presented empirical confirmation that formal criteria are neither a logical nor psychological necessity. This means that for any given class, certain attributes define members of that class, but not all members must possess all the attributes nor exhibit them as strongly (thus defying the principle of necessary conditions). Imagine in your family there are recognizable family traits, but they are not distributed equally among everyone. The idea of family resemblances raised some interesting questions, such as which of these attributes is defining? Are all combinations defining?

5.2.2 George Kelly and personal construct theory

Another important contribution to understanding the cognitive aspects of classification was George Kelly's Personal Construct Theory (1955, 1970). Kelly posited that everyone construes the world in a different and individual way. His original work included an intriguing appendix: The Repertory Grid. This tool eventually was used outside its original intention and became popular for making people's individual implicit constructs explicit. Building on the finding from my dissertation that variations in naming were large, we used the Repgrid technique to explore the naming of office documents. The study yielded a fine-tuned descriptive analysis of consensus, conflict, and correspondence among people for common documents, demonstrating that in fact perfect correspondence in naming between individuals is not the norm (Kwaśnik and Jörgensen 1992).

6.0 Extending the borders

In the time I have been studying classification, we have seen a shift to unification and standardization of bibliographic systems, not just in the United States but also globally. This means that traditional classifications, originally designed in a particular country or for a particular collection are now being stretched, in Michèle Hudon's words (1997), to cover cultural and linguistic artifacts and concepts quite different from those originally intended. This had special significance for me, because given my understanding of a classification's structure and impact, I knew that extending them was not simply a matter of one-to-one translation.

6.1 Translating classifications

In a study comparing the *Dewey Decimal Classification* and the Korean Decimal Classification, two bibliographic schemes from different cultures, we found that obvious differences could be accommodated (Kwaśnik and Chun 2004). For instance, the *DDC* emphasized Christianity, while the KDC allowed more room for Buddhism. The KDC offered greater expressiveness for terms such as "calligraphy." The differences that were more profound, however, were those that construed subjects very differently. For example, "war" is treated as a social process in the *DDC*, and is placed near diplomacy, whereas in the KDC it is classified as a social problem and is near suicide (197). Such a difference in conceptual mapping makes culturally sensitive translation challenging.

Translating the vocabulary of a classification has the typical issues of translation in general. In a study of people's use of even the very most basic kinship terms such as "mother," there are many problems. Among these are finding corresponding terminology and being able to reflect the relationship between terms in the target language correctly. It is surprising how many denotations and connotations the term "mother" has, even in English. We found in the process of translation there may be structural shifts; some terms may have broader definitions and some narrower. There may be differences in how similar terms are construed, and there may be additional criteria of distinction (such as birth order in the case of kinship systems). We suggested that not only terms themselves but also inter-term relationships need to be preserved in cross-cul-

tural, cross-lingual classification translations so that both the source and the target schemes are truly reflective (Kwaśnik and Rubin 2003).

7.0 The importance of context

It was evident to me that a key ingredient in making classifications more responsive and resonant was to find some way of incorporating context into the process. A professor organizes office documents with an eye to the potential uses. A person browsing a collection brings to it personal insights or needs and uses those to help navigate the space. The situation in which classificatory decisions are made plays an important part. Yet, it is quite difficult to reconcile rigid classification schemes with infinitely individual ones. Starting with the findings of my dissertation, the dilemma of creating and using classifications that are accommodating of many perspectives always seems to boil down to one important factor: context. Context defines the scope and the vocabulary. It decides on the elements themselves and which classificatory relationships are pertinent. The following examples show two streams of research in which I tried to find ways of identifying and then representing contextual factors.

7.1 Context and discovery: browsing

One of the features of a classification, any classification, is that it creates affordances for exploration. A classified environment can be searched or browsed for something even if we only suspect or expect it will be there but do not know for sure. Browsing is a method of information seeking that allows the user to explore and navigate without having to specify a query. As such, it is a good way of dealing with an unfamiliar environment or with multiple options or choices. In this way, browsable systems can be invaluable to users crossing over into a new and unfamiliar domain. Browsing reduces cognitive load, because it is generally easier to recognize something once it is viewable rather than to recall a term for it. As well, a key feature of browsing is the ability to hold several parallel paths at once without having to commit to just one.

I wanted to investigate what people do when they browse. The term had been variously defined as searching, scanning, navigating, skimming, sampling, and exploring. It was often described as searching "without a particular purpose" and without a set structure as compared to a database search, for instance. We conducted some informal observations of people browsing in catalogs, online, at a farmer's market, and so on (Kwaśnik and Yoon 1990). The purpose was not simply to record what they did or what "nodes" they visited and how often, but more fundamentally to identify what function they accomplished. Ultimately, we hoped this would provide a set of principles for designing browsable interfaces.

The studies showed that with respect to the structure of the environment, the notion of an unstructured environment is probably not as useful as observing what structures are perceived and how they affect behavior. People will create structures even amidst seeming chaos. Similarly, they will develop a purpose to the process, even if they seemingly started off without one. Comparisons and strategy are developed iteratively. This amazing human capability can be described by the following functions: orientation, place marking, identification, resolution of anomalies, comparisons, and transitions. Each of these functions is performed by constant interaction with the browsing environment, but also with past experiences, future plans, and many other factors the browser brings to the experience. Thus, we can say that browsing is not a passive activity, because there is a formidable amount of sense making involved. As a way of coping with the browsing environment, the browser is constantly devising classifications or views based on a shifting context. Being able to harness these abilities would make interfaces easier and more productive (Kwaśnik 1993).

7.2 Genres

My growing awareness that classification of any kind is a social act led me to the study of genres as they play out in knowledge representation for information seeking and use. A genre identifies something as an integrated cluster of features enacted in a social environment. My colleague, Kevin Crowston, and I conducted a series of studies to see if identifying the genre of a document would improve information access in large digital collections through the identification of document genre as a facet of document and query representation. For instance, knowing something was a computer program might help distinguish it from a musical program, each of these being a different genre. Because most genres are characterized by both form and purpose, identifying the genre provides information as to the document's purpose and its fit to a user's situation, which can be otherwise difficult to assess (Crowston and Kwaśnik 2003, 2005).

First, we needed to define genres for ourselves since this is a very old area of study and crosses many disciplinary lines, from the arts to business. Genres are a way people refer to communicative acts that is understood by them, more or less, but is often difficult to describe in its particulars. Thus, genres are recognized and used, but not so readily described and defined. In our work, we drew on the definition of genre proposed by Yates and Orlikowski (1992, 543), who describe genre as "a distinctive type of communicative action, characterized by a socially recognized communicative purpose and common aspects of form." Note this does not

mean that a genre can be seen purely as a set of document attributes, making the representation of genres a complex and difficult proposition.

Among other things, we wanted to know how people talk about the genre of documents. How do people make use of new, unnamed, and emerging genres? What clues do people use to identify genre when engaged in information-access activities? What facets (basic attributes) of genres do people perceive (Crowston and Kwaśnik 2003)? Our plan was to create a taxonomy of genres by studies of people searching for information in the field. This taxonomy would be used to create a simulated search situation in which we could observe the difference between searching with the aid of genre information and without.

Our assumption going in was that a facetted scheme for genres would be best given their multidimensionality and complexity (Kwaśnik and Crowston 2004). We attempted to harvest clues people used to identify genres, such as "scholarly language" or "reverse chronological dated content" and then reduce them by analysis into possible genre facets. The clues and resulting user-generated scheme were not possible, because the concept of genre was even more slippery than we anticipated. We had difficulty defining genres and developing the scope and expressiveness of the scheme from what our participants told us. They were not able to reliably identify the genre unit or provide unambiguous genre labels. When prompted, they found it difficult to identify genre attributes. There were challenges in distinguishing form and content, as well as challenges in identifying purpose. Finally, the granularity of their tasks differed immensely creating imbalances in the granularity of the terms we could use (Kwaśnik, et al. 2006). As a result, we worked around the lack of a user-generated facetted view of genres and created a researcher-compiled working taxonomy for the purposes of the experiments (Crowston, et al. 2011). In the end, we were not able to demonstrate a substantive change between plain searches and those enriched with genre information, but I believe the full potential of genre representation remains to be explored.

8.0 Classification at the intersection with human endeavor

Sorting Things Out by Geoffrey C. Bowker and Susan Leigh Star (1999), made an enormous impression on me and on the knowledge organization community and beyond. In this work, the authors examined revelatory classifications and standards to show how such classifications silently influence the infrastructure of information, affecting not only policy, but also our daily lives. They were not the first to urge that we question classifications, but their insights were profound, vivid, and compelling. They showed how the system of apartheid, for instance, embodied the pain of South Africa's history, or the classification of tuberculosis affected people's life trajectories.

Classification schemes reflect the knowledge of the domain being classified but also the perspective of the classifier, thus no classification can ever by understood out of context. While we take for granted that classifications do have a social impact, it is not always easy to say precisely how, although we can certainly feel the effects. Potential marginalization, rules for inclusion or exclusion, labeling and naming are all outcomes of classification decisions. Those in power design the classification and then have power over those who are not able to change it. The news is full of examples on a daily basis, from pressure on the Library of Congress to change the term "illegal aliens" to "undocumented immigrants" to who can use the term "champagne." Political resistance often means changing the ruling classification. Many standards are based on classification. Many conflicts have at their core a dispute over basic classifications: when does life begin and when does death occur? What makes a crime a crime? What defines a country? In learning how to evaluate a classification we should always take the critical view. Who devised it? Whose purpose is being served?

8.1 The case of ontological commitment and warrant

Sorting Things Out engendered a critical eye with respect to my analysis and perception of classification, but it was one thing to find the strengths and flaws and another to develop a vocabulary for discussing this systematically and coherently. Fortunately, I was asked to contribute to a festschrift for Claire Beghtol (Kwaśnik 2010) and chose to focus on her pivotal and far-reaching 1986 article "Semantic Validity: Concepts and Warrant in Bibliographic Classification Systems" (Beghtol 1986). In this article she explores the semantic, rather than the syntactic axis of bibliographic classification systems. According to her, the attention of scholars on facetted schemes and classificatory structures had heretofore pulled our attention to the syntactic aspects (e.g., concept division and citation order), with semantics being considered more or less a question of the terms and their relationships and somewhat taken for granted. In this paper she states (110-11) that "the warrant of a classification system can be thought of as the authority a classificationist invokes first to justify and subsequently to verify decisions about what class/concepts should appear in the schedules …. The semantic warrant of a system thus provides the principal authorization for supposing that some class or concept or notational device will be helpful and meaningful to classifiers and ultimately to the users of documents." Warrant emerges from various points of authority: literary warrant, scientific/philosophical warrant, educational warrant, and cultural warrant, each with its own effect in terms of establish-

ing the semantics and then also the syntax of any given classification (119-221).

This was a revolutionary idea in the sense that notions of meaning being fixed have guided the design of many of our systems, because it was assumed that meaning became more stable and consensus firmer as the evidence mounted and the ideas withstood the test of time. Yet, modern approaches assume that meaning is not fixed and is created in use. It is also interesting to consider contemporary phenomena such as wikipedia.org, where the classification and the content are built cooperatively. That is, in principle, both the text and the classification that organizes the texts in such emergent systems are not managed from the top. Nobody questions the fact that such systems must be flexible and dynamic, and yet nobody wants an amorphous mess either. Our challenge is to assess the warrant for any given classification project and judge the classification against it.

One example of such a challenge is the classification of academic departments and programs. In the modern American university, there is often a federated system of schools and colleges, each with its own warrant for how it describes and labels the knowledge in its purview, what Elaine Svenonius called ontological commitments (Svenonius 1997). Chemistry views its own world differently than do the performing arts, and the differences are evident in academic processes. In a study of my own university, I used the collection of hundreds of courses to see how this played out (Kwaśnik 2016). It is clear, for example, that the term "girl education" is construed differently in economics than it is in education. In one the presence of girl education is a factor in economic development, and in the other a subject of interest in its own right. Resolving such territorial disputes in claiming courses is left to the curriculum committees.

A more interesting example, though, is forensic science, an instance of mixed ontological commitments. There are many such examples at universities: archival studies, physical education, and environmental studies among them. The study of forensic science is the use of science to help solve crimes. It calls upon an array of disciplines to support a specified set of professional practices. Courses in the FS curriculum include forensic anthropology, human osteology, forensic entomology, forensic chemical analysis, forensic linguistics, forensic evidence in law, and forensic psychology. Each of these comes with its own ontological commitments and its own body of knowledge. The forensics student's program of study is not based on the supporting and contributing disciplines, however, but rather on a prescribed sequence of professional practice: identification of crime; collection of evidence (autopsy, traces); analysis of evidence; and support of the preparation of a legal case.

My takeaway from studying these cases is that, broadly speaking, when classification is structured to support human endeavors, the purpose is different than when it is structured to support science. Thus, understanding the underlying warrant is all the more important.

9.0 Full round back to hierarchies

In a way, then, my early respect for hierarchies would seem to have been validated up to a point, if, and only if, the circumstances supporting a hierarchical structure were evident. Recently, though, even this qualified view has been somewhat shaken. Hope Olson's mission is to analyze our traditional knowledge representation systems from the point of view of those whose voices are not well reflected. In her article "How We Construct Subjects" (2007), she takes apart the notions behind hierarchies and brings to bear feminist thinking to offer a penetrating critique. She posits that hierarchies are by nature flawed because they require one element to be in the superior position and all other elements subordinate to that. This structure creates skewed assumptions that privilege one set of elements over others. I will use my own example here: imagine a hierarchical classification of astronauts. At the top is the term "astronauts." On the next level down are subclasses of astronauts: "minority astronauts," "women astronauts," and so on. This may seem like a laudable effort at inclusiveness. There is no subclass for "men astronauts," however, because the notion of astronauts being men is the default and is baked into the assumptions. In the chain of transitivity, men hold the defining set of attributes. This is a dilemma, because while one would like a way to represent the special attributes of women astronauts, placing them in the subordinate position means that they are defined by the male criteria first and foremost. From my perspective, there does not seem to be a good way to undo this imbalance in a hierarchy.

Having laid out the limitations both in content and structure, Olson suggests rewriting and restructuring our schemes so that the all-important connections are visible—a web instead of a hierarchy (522). According to her (522), we need "richer and more situated logical models" that allow for the representation of interdependence and connectedness. I am just now beginning to rethink how my favorite classifications could be reframed in this way, or if they even should be. The power of hierarchies and other formal classifications is not easily dismissed, but at the same time the fact that they are so embedded in our culture should be explored. What we see as taken for granted could be hiding subtle and not so subtle biases.

9.0 Summary

Classification is beautifully recursive. What we know guides our classifications, and in turn, our classifications guide what we are able to know. Many questions remain: 1) Who creates the classifications by which we must all live?; 2) Who has the authority to change them?; and, 3) What is an effective way of creating classifications that are inclusive but also effective? We use classifications to better capture what we know; we also use them to embody our values and perspectives. We don't have a choice of whether to classify or not, but we are obliged to pay attention to the consequences of what we do.

References

Aristotle. 1963. *Categories, and De interpretatione*, trans. J.L. Ackrill. Clarendon Aristotle series. Oxford: Clarendon Press.

Beghtol, Clare. 1986. "Semantic Validity: Concepts of Warrant in Bibliographic Classification Systems." *Library Resources & Technical Services* 30: 109-25.

Bowker, Geoffrey C. and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press.

Bronowski, Jacob. 1978. *The Origins of Knowledge and Imagination.* New Haven, CT: Yale University Press.

Crowston, Kevin and Barbara H. Kwaśnik. 2003. "Can Document-Genre Metadata Improve Information Access to Large Digital Collections?" *Library Trends* 52, no. 2: 345-61.

Crowston, Kevin, Barbara H. Kwaśnik, and Joseph Rubleske. 2011. "Problems in the Use-centered Development of a Taxonomy of Web Genres." In *Genres on the Web: Computational Models and Empirical Studies,* ed. Alexander Mehler, Serge Sharoff, and Marina Santini. Dordrecht: Springer, 69-84.

Davies, Roy. 1989. "The Creation of New Knowledge by Information Retrieval and Classification." *Journal of Documentation* 45: 273-301.

Hudon, Michèle. 1997. "Multilingual Thesaurus Construction: Integrating the Views of Different Cultures in One Gateway to Knowledge and Concepts." *Knowledge Organization* 24: 84-91.

Kaplan, Abraham. 1967. *The Conduct of Inquiry: Methodology for the Social Sciences.* Addison-Wesley.

Kelly, George A. 1955. *The Psychology of Personal Constructs*. New York: Norton.

Kelly, George A. 1970. "A Brief Introduction to Personal Construct Theory." In *Perspectives in Personal Construct Theory*, ed. D. Bannister. London: Academic Press, 1-29.

Kwaśnik, Barbara H. 1989a. "How a Personal Document's Intended Use or Purpose Affects Its Classification in an Office." In *SIGIR '89 Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval Cambridge, Massachusetts, USA June 25 – 28, 1989*. New York: ACM, 207-10. doi:10.1145/75334.75356

Kwaśnik, Barbara H. 1989b. "The Influence of Context on Classificatory Behavior." PhD diss., Rutgers.

Kwaśnik, Barbara H. 1991. The Importance of Factors that Are Not Document Attributes in the Organization of Personal Documents." *Journal of Documentation* 47: 389-98.

Kwaśnik, Barbara H. 1993. "A Descriptive Study of the Functional Components of Browsing." In *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction Ellivuori, Finland, August 10-14, 1992*, ed. James A. Larson and Claus Unger. Amsterdam: North-Holland Publishing, 191-203.

Kwaśnik, Barbara H. 1993"The Role of Classification Structures in Reflecting and Building Theory." In *Proceedings of the 3rd ASIS SIG/CR Classification Research Workshop,* ed. Raya Fidel, Barbara H. Kwaśnik and Philip Smith. Advances in Classification Research 3. Medford, NJ: Learned Information, 63-82.

Kwaśnik, Barbara H. 1999. "The Role of Classification in Knowledge Representation and Discovery." *Library Trends* 48, no. 1: 22-47.

Kwaśnik, Barbara H. 2002. "Commercial Websites and the Use of Classification Schemes: The Case of amazon.com." In *Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries: Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain,* ed. Mariá J. López-Huertas; and Francisco J. Muñoz-Fernández. Advances in Knowledge Organization 8. Würzburg: Ergon, 279-85.

Kwaśnik, Barbara H. 2010. "Semantic Warrant: A Pivotal Concept for Our Field." *Knowledge Organization* 37: 106-10.

Kwaśnik, Barbara H. 2011. "Approaches to Providing Context in Knowledge Representation Structures." In *Classification & Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar 19-20 September 2011, the Hague, the Netherlands*, ed. Aida Slavic and Edgardo Civallero. Würzburg: Ergon, 9-24.

Kwaśnik, Barbara H. 2016. "The Web and the Pyramid: Hope Olson's Vision of Connectedness in a World of Hierarchies." *Knowledge Organization* 43: 367-72.

Kwaśnik, Barbara H. and You-Lee Chun. 2004. "Translation of Classifications: Issues and Solutions as Exemplified in the *Korean Decimal Classification*." In *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference, 13-16 July 2004, London, UK*, ed. Ia C. McIlwaine. Advances in Knowledge Organization 9. Würzburg: Ergon, 193-8.

Kwaśnik, Barbara H. and Kevin Crowston. 2004. A Framework for Creating a Facetted Classification for Genres: Addressing Issues of Multidimensionality." In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences: Abstracts and CD-ROM of Full Papers: 5-8 January, 2004, Big Island, Hawaii*, ed. Ralph H. Sprague, Jr. Los Alamitos, CA: IEEE Computer Society Press. doi:10.1109/HICSS.2004.1265268.

Kwaśnik, Barbara H. and Kevin Crowston. 2005. "Genres of Digital Documents: Introduction to the Special Issue." *Information, Technology & People* 18: 76–88.

Kwaśnik, Barbara H. and Mary Grace Flaherty. 2010. "Harmonizing Professional and Non-professional Classifications for Enhanced Knowledge Representation." In *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference, 23-26 February 2010, Rome, Italy*, ed. Claudio Gnoli and Fulvio Mazzocchi. Advances in Knowledge Organization 12. Würzburg: Ergon, 229-35.

Kwaśnik, Barbara H. and Corinne Jörgensen. 1992. "The Exploration by Means of Repertory Grids of Semantic Difference among Names for Office Documents." In *Classification Research for Knowledge Representation and Organization: Proceedings of the 5th International Study Conference on Classification Research,* ed. Nancy J. Williamson and Michèle Hudon. Amsterdam: Elsevier.

Kwaśnik, Barbara H. and Xiaoyong Liu. 2000. "Classification Structures in the Changing Environment of Active Commercial Websites: The Case of eBay.com". In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*, ed. Clare Beghtol, Lynne C. Howarth and Nancy J. Williamson. Advances in Knowledge Organization 7. Würzburg: Ergon, 372-7.

Kwaśnik, Barbara H. and Victoria L. Rubin. 2003. "Stretching Conceptual Structures in Classifications across Languages and Cultures." *Cataloging & Classification Quarterly* 37: 33-47.

Kwaśnik, Barbara H. and Kyung-Hye Yoon. 1990. "A Design Study for the Investigation of Browsing Behavior." In *CAIS/ACSI '90: Proceedings of the Canadian Association for Information Science Annual Conference, Kingston, Ontario, May 24-26, 1990*.

Olson, Hope A. 1998. "Mapping beyond Dewey's Boundaries: Constructing Classificatory Space for Marginalized Knowledge Domains." *Library Trends* 47: 233-254.

Olson, Hope A. 2007. "How We Construct Subjects: A Feminist Point of View." *Library Trends* 56, no. 2: 509-41.

Peterson, Roger Tory and Margaret McKenny. 1974. *A Field Guide to Wildflowers of Northeastern and North-central North America: A Visual Approach Arranged by Color, Form, and Detail.* The Peterson Field Guide Series 17. Boston: Houghton Mifflin.

Rosch Eleanor H. 1973. "Natural Categories." *Cognitive Psychology* 4: 328-50.

Rosch Eleanor H. and Carolyn Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7: 573-605.

Ranganathan, S. R. 1967. *Prologomena to Library Classification.* 3rd ed. Bombay: Asia Publishing House.

Svenonius, Elaine. 1997. "Definitional Approaches in the Design of Classification and Thesauri and Their Implications for Retrieval and Automatic Classification." In *Knowledge Organization for Information Retrieval*, ed. L.C. McIlwaine. FID 716. The Hague: International Federation for Information and Documentation, 12-16.

Yates, Joanne and Wanda J. Orlikowski. 1992. "Genres of Organizational Communication: A Structurational Approach to Studying Communications and Media." *Academy of Management Review* 17: 299-326.