

The Relationship between What We Know and How We Classify: Some Philosophical Bases for Inquiry

Janice Woo

School of Library and Information Studies
University of California, Berkeley
Berkeley, CA 94720
jwoo@info.berkeley.edu
jlwoo@uclink2.berkeley.edu

According to Bliss (1929), "The simplest definition of *class* is: *things related by some likeness*. This may be reduced to two words, *like things*." (Bliss, p.119) Further, he states that "A series or system of classes arranged in some order according to some principle or conception, purpose or interest, is termed a *classification*." (Bliss, p.142) From this definition I would like to explore three areas that are essential to our understanding of what the process of classification entails. Two are to be found in Bliss's "like things." What exactly are these things which we endeavor to classify, and what is our basis for determining similarity? The third, and in this paper the most significant, area revolves around the notions of principle and purpose. If we acknowledge from the outset that classification is dependent on some sort of preconceived or preexisting schema, what import does this have for a) how we construct our classifications, and b) how our classifications subsequently are interpreted or utilized, say in the pursuit of further knowledge.

The method of this exploration will be analogy or juxtaposition. The goal is to gain a greater understanding of how we classify by mapping our practices onto certain concepts proposed in the theory of knowledge. This relationship is warranted by the fact that we must know something about the things we classify and that our classifications are intended to be a means of access to knowledge. The issues surrounding how and what we know have long been the subject of philosophical investigation, and as such, should be able contribute to the foundations for our own research. This position paper will highlight some basic ideas in three particular areas: theories of perception, similarity, and conceptual frameworks. It is not being proposed, however, that any one of these philosophical theories, or combination of, should serve as a model for our own theories or applications. Rather it is the juxtapositions themselves that might open up new or different avenues of inquiry.

To begin let us look at some simple examples that illustrate our concerns.

What is the nature of the things we classify? If we want to classify a photograph of the Washington Monument, should we regard our subject as the physical edifice out on the Mall, as a document representing the monument in some manner, as the idea of the monument and what it stands for, or perhaps only in formal terms as a two-dimensional object? If we were going to verbalize the content in this way, we might come up with descriptions as varied as:

Washington Monument (Washington, D.C.)
Washington Monument (photograph)
Monuments and memorials, American
Color landscape photograph

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Certainly these descriptions are not exhaustive, and there may be many other ways to conceptualize a photograph of a famous landmark. But what should be apparent is our need to be clearly cognizant of how we have decided to construe the entity in question. I will suggest that three avenues of epistemological investigation — realism, representationalism, and phenomenism — may guide us in this process.

What is our basis for determining similarity? Consider these three groupings.

Kitchen chair, milk stool, beanbag chair, tree stump, plastic milk crate

Kitchen chair, coffee table, dresser, torchiere, TV

Kitchen chair, electric piano, easel, cow, lunar landing pod

Hopefully, it is obvious that the kitchen chair has found its way into three different groups because each grouping attends to a different attribute of the chair. Traditionally we have sought to accommodate as many attributes of an entity as is useful or appropriate. Interpreting which are to be deemed as such, of course, relies on one's particular outlook on the matter, an issue that will be taken up in our third area of concern.

At a more fundamental level, however, what is it that makes similarity operative? Is commonality of any particular attribute necessarily an indicator of similarity? Does resemblance provide better results and under what circumstances? What exactly is the basis for knowing when similarity obtains?

How do conceptual frameworks affect classification? In 1985-86 a series of articles and letters appeared in *Library Journal* calling attention to a conflict between library classification and an alternate belief system. The issue was how creationism ought to be classified in the Dewey Decimal system, as science or religion?

Berman [a cataloger]... laments the fact that Creationists want Creationist literature to be classified in the 500s [science] rather than the 200s [religion] ...recent literature on Creationism, however, is strictly of a scientific nature. (*Library Journal* 2/1/86, p.12)

The Evolution Scientist and the Creation Scientist are observing, measuring, and studying the same evidence, the same phenomena, the same amazing intricacies of nature; but they are drawing different conclusions.... The Creation Theory of Origins has a place on the library shelf beside the Evolution Theory of Origins. (*Library Journal* 1/86, p.10)

Berman along with heavy-hitters like the Library of Congress held the view that creationism should be classed as a component of religious doctrine, while the creationists themselves held their view to be a rationally arguable scientific proposition, which warranted classification as science. When such divergence of conceptual frameworks occurs, this being just one example, how should or can

the practice of classification respond? This problem will be examined in the third section of this paper. First, however, we turn to the question of like-things.

THINGS

Theories of Perception

Our first question was: what is the nature of the things that we classify? Rather than pursue this ontologically, we might also gain insight through epistemological theories of perception, which offer up possibilities about what exactly we are perceiving and hence what constitutes our source(s) of knowledge. There are three basic models: direct realism, representationalism or indirect realism, and phenomenalism or anti-realism. Searle (1983) gives a readily understandable diagram for each.¹

REALISM

Perceiver -----> object perceived

REPRESENTATIONALISM

Perceiver ---> sense-datum -----> object

PHENOMENALISM

Perceiver---> sense-datum

The direct or naive version of realism is the most commonsense model: we perceive (via any of the senses) that there is some object out there in the world and we gain our knowledge from this object. In a rough manner of speaking, what you see is what you get. Critics of this view, however, reply that appearances can be deceiving. For example, the earth is not really flat, stars are not pinpoints of light, and ocean water is not blue. Similarly, physics tells us that objects that appear to be solid are actually made up of vibrating atoms. There is also the question of why two people may perceive the same object in different ways.

This sort of complaint is handled by the representational theory. Here our perception of things in the actual world are mediated by another entity, called a sense-datum — or sometimes an idea, percept, or appearance. (Dancy, p.145) The sense-datum, which is usually considered an individual's mental construct, represents or reflects real-world entities. A mirror reflection would be a comparable effect: though we see our own image, we are not actually seeing ourselves; it is an indirect perception. This accounts for how we are able to see a star in tonight's sky that might have burned out hundreds of years ago. The light waves that emanated from the star create the appearance of a small point of light in the retina, which is transmitted to the mind so that we can then interpret it as being a star. This theory seems to be most in keeping with current concepts in

1. Searle's diagrams also represent his notion of visual experience (the arrows). Searle argues for a more sophisticated form of direct realism that is not considered here, but should be in future inquiry.

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

science. For Frege (1892), it also accounted for synonymies like "morning star" and "evening star," two sense-data that refer to the single object, Venus.

If one accepts that a star that no longer exists can still be visible to people on earth, then the phenomenalist view is fairly easy to comprehend. This anti-realist theory claims that all we ever see or experience are appearances, and that these perceptions are in fact all that the world is made of. "All physical objects are mental phenomena that would cease to exist if they were not perceived." (Pojman, p.67) The sensation of pain and hallucinations are good examples of phenomenalist perceptions. A question arises, however, about the existence of seemingly stable and permanent objects. Is my desk only here when I'm perceiving it? Does it cease to exist when I leave the room, and if so, why does it reappear exactly the same on my return? This is usually accounted for with the notion that the possibility of something existing is always there, but that the existence is not realized until perceived. For instance, though you might think or believe your wallet to be in your briefcase, you wouldn't really know unless you were to see or feel it there. Phenomenalism is best at explaining individual or idiosyncratic experiences of the world.

Within the philosophical milieu, Dancy (1985) notes that the representationalist or indirect realism theory was the favored one for the first part of this century but that more contemporary thought has shifted towards a form of direct realism. In some ways, especially with the advent of automation, notions about the things we classify have followed a similar course.

The Nature of Things We Classify

It was noted that direct realism is the commonsense view. Likewise the most straightforward way of classifying seems to be simply to organize things in the world. Botanical and zoological classification are of this mode, and according to Langridge (1992) the primary forms of classification are scientific and philosophic, approaches that look directly at the world. In applying the direct object of perception approach to our example of the photograph of the Washington Monument, we would therefore want to classify the thing before us, i.e. a "color landscape photograph". In library practice we can see this method in the segregation of books by size and in the distinction of material format and genre, but obviously this has not been found to be a generally effective form of organization.

Rather, again according to Langridge, bibliographic classification is a secondary form that "deals with knowledge after it has been embodied in documents." (Langridge, p.23) In other words, the documents we classify are indirect representations of the object or knowledge source in question. For instance, although we may be classifying our sample photograph as "Washington Monument (Washington, D.C.)", we are making no claim that the document in hand actually is the monument; nor if the photo were in a New York Public Library collection, should we construe the document as being somehow in Washington, D.C. To think otherwise, which would be an attempt at a direct interpretation, seems fairly ludicrous. Thus traditional library classification appears to be quite in keeping with the representationalist way of thinking: a document is an entity representing some piece of knowledge in the world, in the same relation as a sense-datum is to the object of perception. It is indeed the *idea* of the Washington Monument that we want to account for, and we receive it via the photograph.

In this sense we might think of a document as anything (be it image, text, object, databits) that we want to classify in terms of its content as opposed to its physical form. However, because the representational theory relies in large part on some kind of resemblance, it is less successful at accommodating verbal text. The words "Washington Monument" treated as *tokens* on a printed page are comparable in form to the photo, yet how they attach to the physical edifice — presumably through their word *types* — is a process that has not achieved a satisfactory explanation. It is not clear, therefore, how a text document represents a real world object, if at all.

The problem of textual referent is bypassed when texts are classified by a direct method, that is as objects themselves. Unlike traditional library classification, automatic classification such as described in van Rijsbergen (1979) does treat documents in a direct primary manner. In its basic form this kind of classification or document clustering is carried out by judiciously collocating (for retrieval purposes) similar words or strings of characters. An IR system simply treats the (token) occurrence "Washington Monument" as "W-A-S-H-I-N-G" etc., and what this series of letters might stand for in the real world does not enter the picture. And unlike when we manually classify, an IR system is unable to interpret and interject whatever *ideas* we may have about what that monument is about. It is entirely possible that someone might classify a photo of the Washington Monument as "my vacation, 1985," which a machine could not do unless that information was physically present. So in this realistic view of documents, we have gained simplicity but at the cost of richer interpretational significance.

Gaining greater insight into more diverse interpretations depends upon access to personal experiences and views, which finds its most accommodating analogy in the phenomenalist theory. Here a material object is constituted solely on the basis of sense-data, in the form of coherent collections of experiential perceptions. Since these experiences might vary from person to person, the outcomes can be quite varied. Thus our sample photograph would not be classified according to the object it refers to, but rather in virtue of each classifier's or viewer's perception of it, e.g. as a tourist site, as an architectural feat, or as a patriotic symbol. This could lead, of course, to a kind of Rashomon effect, or more precisely a state of solipsism. A response to this criticism is that highly individualized sense-data will not correlate properly with the dominant collection of perceptions; inappropriate perceptions will be lacking in enough coherence to be maintained within the total picture, leaving only normal or regular views to obtain. In automatic classification, relevance feedback and heuristics provide some means of incorporating personal input, but here too there is the question of how truly idiosyncratic responses ought to function in relation to the larger group of "normal" or averaged responses.

Discussion

We have seen three theories of perception that hold different views on the source or object of knowledge. All three have advantages and problems in explaining how we are able to determine what it is that we know. The purpose of mapping certain classification practices onto these theories is to help shed light on the objects of our endeavors. That is what is the nature of the things that we classify? From the preceding observations it should be clear that there does not seem to be one single kind of thing, rather it depends upon our goals, purposes, and techniques. The point that I would like to make is that epistemological theories enable us to see what kind of ramifications arise from our particular practices. We can then ask whether an effect is accounted for, whether it

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

contributes to the desired outcome, or if this was indeed the nature of the thing we intended to classify.

A few of these ramifications present themselves in the very cursory examination above, while in the detailed philosophical literature there are many others that could be investigated as well. For instance, the primacy of the actual objects in the world results in an essentializing or universalizing tendency, at least in the basic form of direct realism. If our knowledge comes out of these kinds of objects, then what that object is should be all that is the case. Differing perceptions cannot be explained in terms of that object itself, even though it is supposed to be the only source. (More recent versions of realism, such as Searle's (1983) do attempt, however, to provide an explanation.) We have also seen that for word tokens or even word types, the realist account seems to leave us somewhat bereft of at least a certain kind of interpretational meaning. The representationalist theory is better equipped to handle variant perceptions and situations like synonymy, but seems to fair no better in accommodating intermediate sense-data that do not rely on resemblance or direct reference. Indeed there has not been a completely satisfactory account of what provokes these mediating entities and what relation they are supposed to have with the real world source. Phenomenalism dispenses with the objects-in-the-world problem, but here the question is how to handle the relativism of "unanchored" sense-data.

In doing research on classification we might then ask ourselves:

- Are we classifying physical objects or ideas?
- If a physical object, is it the object itself or a representation of it?
- Do we want to assume that full knowledge about an object is in the object itself, or do we want to accommodate personal interpretations?
- How can we know what the personal interpretations are?
- If we assume that we are dealing with intermediate entities, how do we determine what they refer to and in what relation?
- And perhaps most importantly, what kind of things are words?

SIMILARITY

Goodman's Strictures on Similarity

The second question under consideration is the nature of similarity. The idea here, which will be based on Goodman's (1972) essay, "Seven Strictures on Similarity," is simply to unpack what, on the face of it, seems to be a quite straightforward notion. Goodman's critiques of similarity will be briefly paraphrased and exemplified.

His first claim is that resemblance is not necessarily a condition for representation. Two things that look alike may have less connection than two apparently different things. For example, one might say that a beanbag chair holds a greater resemblance to an inflatable beach ball than to a kitchen chair, at least in terms of shape. But within our grouping of things to sit on, we want to represent the beanbag chair as a chair, and not as a toy projectile. Perhaps more to the point is how we might want to connect objects with words. Suppose there is an object, Bill's chair. Would this be better represented by a doll-house's chair that looks very similar but in miniature or by an inscription that

says "Bill's chair." It would more likely be the latter yet the resemblance is nil. In a multimedia situation, similarity in resemblance would thus be inadequate.

Goodman's second observation makes the same point, only for words as tokens or inscriptions. For example, "chair" looks exactly like "chair" but if one is in English and the other in French, you get two word types, "chair" in English and "flesh" in French. Again this is simply to say that resemblance does not give the whole story.

The next two issues involve purpose and use. Here Goodman is saying that similarity is not an intrinsic phenomenon, but rather is defined by how we decide to construe it. The three groupings presented earlier demonstrate this point clearly: these were grouped according to our own specifications — things to sit on, household furniture, and things with four legs. Through practice and use our various conceptions of similarity become entrenched so that we might think that they were always naturally the case. In the special case of metaphor, "proud as a lion" is not an inherent connection but one that emerged because we have made it so.

The fifth stricture has to do with similarity's predictive powers, or rather lack of. For Goodman any grouping, series, or even description (cf. his "grue" example) is subject to alteration so that we are never in a position to know what eventual outcomes will evolve. If it were 1950, for instance, adding "computer" to our group of household furniture — kitchen chair, coffee table, dresser, torchiere, TV — would have been inappropriate. A computer was not a consumer item and thus would have seemed very dissimilar. Granted a computer is not exactly a piece of furniture, but nevertheless it would not seem out of place in that list today.

Next is the problem of likeness between attributes or particulars and their ability to indicate an overall basis for commonality. Using Goodman's reasoning, a beanbag chair can be sat on and is a piece of furniture; a table is a piece of furniture and has four legs; a cow has four legs and one could sit on it. In virtue of the circular connection of attributes, the supposition was that the beanbag chair, table, and cow would therefore have an overall quality in common that could define them as a class. This appears not to be the case, however, and for Goodman this shows one of the limits for inferring similarity.

Lastly, there is the general notion that having something in common does not go very far in establishing similarity. The argument is that everything has at least one thing in common with something else, and that nothing has everything in common with something else. Therefore commonalities alone do nothing to distinguish what ought to constitute similarity, because on this account everything would be similar and nothing would be similar.

In sum the problem with similarity is that its basis is tenuous. There seems to be no categorical definition of its precise nature. How then can we know when two things are similar and in what way they are similar. For Goodman the concept is entirely relative and thus depends upon frame of reference to have any functional utility. This is in fact how similarity operates in everyday usage but without the benefit of explicit motivation. A more informed use of similarity — say for classification theory — therefore requires specification of the criteria for likeness and commonality and for purpose and intent. At the same time, we would like to be able to find grounds on which to

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

justify these specifications, but for Goodman these turn out to be just the similarities we were attempting to define in the first place.

Questions for Classification

If we say that classes are like-things, what then do we mean by likeness or similarity? Some pertinent questions might be:

- To what extent should we rely on resemblance, especially in cases like "chair" and "chair"?
- Other than resemblance, what representational relations can we use to determine similarity?
- Are there factors other than similarity or likeness that should be used to determine classes or categories?
- How much predictive force can we expect from categories already established? In other words, how stable are they?
- How do we incorporate a dynamic view of similarity?
- How do we ascertain the appropriate or applicable frames of reference, and how do we integrate diverse frames of reference?
- To what extent to frames of reference derive from prior notions of similarity? Are these prior notions justified and how?

CONCEPTUAL FRAMEWORKS

Much of the preceding discussion has seemed to indicate that a key notion for resolving our difficulties lies in the recognition of conceptual schemas or frameworks. For instance, if we can isolate our particular conception of entities, then we are in a better position to know what kind of knowledge relations we are dealing with. If we define our conception of similarity, then it becomes a functional criteria. As Goodman pointed out, however, there is at least one problem with the latter, which is that our notions of similarity have as much power in shaping our frameworks as in reverse. In other words, what can conceptual schemas tell us about anything if we already had to know these things in order to have a schema in the first place.

Justification and Truth

One way of working through this seeming paradox is to unravel a) how we justify the things we know, that is those things that make up our conceptual schema and b) how we are able to know which things in our conceptual schema are true. From the epistemological literature (Dancy 1985 and Pojman 1993), one can discern two distinct yet intertwining lines of pursuit on this matter, which I will characterize as independence-motivated and dependence-motivated methods. The independence view claims that there is some basic foundation for beliefs and truths that does not depend on any particular frame of reference. On the other view, beliefs and truths are dependent, that is arise out of, the entire set of beliefs and truths that make up the conceptual framework of what we know.¹

1. Please note that these notions have been greatly oversimplified for the present discussion in order to serve as guiding postulates. Their treatment in the philosophical literature is very complex and not so neatly delineated as described here. Further research needs to examine the full complement of arguments supporting the various positions.

Foundationalism, which comes out of Cartesianism, posits that there are some basic beliefs and truths that exist as first principles, that is things that we simply know to be the case or that we perceive as self-evident. The greater part of knowledge is comprised of additional non-basic beliefs that are derived from these original ones, but the foundational basis is independent of everything save what is given by itself. A related notion is the correspondence theory in which truth is established by its fitting with the "facts," those things that are just indubitably the case. By accepting that there are some entities, beliefs or truths, that are always the case, the structure of knowledge gains a stable foundation on which to build. Knowledge can be justified if it can be linked back to the essential basic beliefs, and it is true if it is in accordance with the "real facts" about the world. One might question, of course, whether we are ever in a position to accept anything as being either essential or real. Once we do, however, the outcome is a universalizing tendency. All knowledge points back to an immutable foundation of self-evident facts and beliefs. Conceptual frameworks based on any sort of conflicting evidence must therefore be incorrect.

To support the theory of evolution we have the evidence of the fossil record, radiocarbon dating, geological strata, indications of the age of the universe, etc. Our intuitive experience with the natural world confirms that biological species evolve rather than spontaneously appear. The large body of scientific investigation makes us justified in believing that this is the right account of things. The fact that creationists may have constructed an account that is plausible in its own right doesn't make it an acceptable theory because it is not based on the overwhelming evidence. In other words, our knowledge of how the world operates is independent of creationist (or any alternate) beliefs; no matter how strong or credible, they do not change what has been justified as being the case.

In contrast, belief systems are part and parcel of the dependence-motivated method, which takes form as the coherence theory of justification and the coherence theory of truth. Dancy (1985) provides these definitions.

Coherence theory of justification:

This theory holds that a belief is justified to the extent to which the belief-set of which it is a member is coherent. (p.116)

Coherence theory of truth:

This theory holds that a proposition is true if it is a member of a coherent set. (p.112)

According to Annis (1978), what is meant by coherence is having the qualities of "consistency, connectedness, and comprehensiveness." (p.213) To put it another way, knowledge is dependent on or is justified by the coherent relationships that obtain within an entire system of beliefs. In contrast with foundationalism, none of these beliefs has any special justificatory status; rather each and all are justified in terms of all the others, i.e. are completely interdependent. Similarly, truth is not tied to particular facts about the world, but is a function of being consistent with all that is known and held to be true. It is in essence a holistic view of knowledge.

What coherence theory does not do, however, is to claim a mutual dependence for justification and truth. "A belief can be true without being justified and justified without being true." (Dancy, p.116)

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

This is what opens the door for the emergence of divergent conceptual schemas. Because we usually make the presumption that there is only one truth or set of truths, if truth were a necessary component of justification, only one set of justified beliefs could ever be true. Without this restriction, any coherent set of beliefs becomes justified grounds for a knowledge base or conceptual framework. Those who deny the very possibility that there might be more than one schema critique this view with what is known as the plurality objection. According to this argument, correspondence with the truth can only occur for one complete set of coherent beliefs; other schemas, though coherently justified, could not be coherently true and thus not viable. For our present purposes, however, this reliance on truth seems only to beg the question.

Pojman (1993) lays out a scenario in which the creationist theory might in fact be the accurate account, but due our delusions or state of unenlightenment, we have been unable to grasp this truth. The independence-motivated view (which Pojman calls justificationist) would hold that evolution is still the correct theory because everything that we do know to be the case — in our state of unenlightenment — points to the veracity of the scientifically supported theory. In the alternate view (called reliabilism), “so long as a reliable process caused creationists to believe the way they did [true or not], they have knowledge as well as justification.” (Pojman, p.290)

As presented in this discussion, we have two ways of dealing with creationist/evolutionist debate. We can decide that only one of the conceptual frameworks is acceptable, or we can hold that both are acceptable. In the former, knowledge tells us which schema is the operative one; in the latter, the schemas help construct what it is we know. To put it crudely, the data drives the theory, or the theory drives the data, respectively.

Implications for Classification

At the beginning of this paper, we were concerned about the import of conceptual schemas for a) how we construct our classifications, and b) how our classifications subsequently are interpreted or utilized in pursuit of further knowledge. Following Kwasnik (1992) let us say that “classifications are really very much like theories.” (Kwasnik, p.63) In terms of justification and truth we can thus reframe the questions as to a) how data/knowledge motivates classification, and b) how classification motivates data/knowledge.

It seems to me that traditional classification practice has taken the universalist approach. We have sought to organize knowledge according to how we think the world really is. This kind of project is manifest in both the Baconian-influenced Dewey Decimal system and in Ranganathan's five facets, to name just two. And, of course, the very name “Universal Decimal Classification” attests as well. On this line of reasoning then, we should strive for and expect to arrive at one correct classification scheme. In practice this approach is warranted by the fact that, for the most part, we are dealing with secondary knowledge. We are willing assume that the things we classify are justified truths about the world and not illusion or falsities.

To explain the less than successful endeavor to produce a single, universal classification system, we might take the position that we have yet to capture those very basic building blocks on which our body of knowledge rests. Indeed, on the foundationalist plan, the schema is a tree-like structure with knowledge sprouting out from its indubitable roots. But the problem, as exposed by Wittgenstein (1921) and others, is that even if these basic truths or entities are somehow self-

evidently known, we, say as classifiers, have no means to analyze them down to their pure atomic essences. So if we intend to arrive at a scheme of knowledge by this method, we can have little hope of success. On a more practical level, however, we might counter that we will not demand complete certainty of knowledge, that we will use only those elements that seem likely to be the essentials, whether proven or not. This will perhaps give us the 95% universal solution, but the five-percent alternative knowledge (creationists for example) remains yet unaccommodated. Another way to account for the creationist version is to allow that it was, in fact, justified at one point in time (e.g. 1400 AD) but that advances in science now clearly point to the evolutionist theory as the correct scheme. This is Kuhn's (1962) paradigm shift model in which one conceptual schema usurps another. But again this does nothing to accommodate alternative views other than to say that we were once previously misguided in thinking they were correct. If we are wedded to the notion that a good classification scheme is derived on the basis of data that we "know" to be "true," then the outcome must necessarily be the one scheme that reflects that data. Other schemes must be jettisoned or abandoned.

Perhaps at this point we are thinking that the problem is not the basic data itself but that we are not organizing or ordering it properly. By this line of reasoning then we turn to the idea that the theory (how things are ordered) is what motivates our data (knowledge). All we need to do is to arrange the data so that it makes sense, so that it is justified in terms of "consistency, connectedness, and comprehensiveness." (Annis, op.cit.) According to the coherence argument given above, this method will not only accommodate multiple conceptual schemas, but in so doing should actually increase our knowledge because we are not discounting nonconformist theories and views. The question for classification is how exactly are multiple schemas to be reconciled and what kind of knowledge can we expect to ensue. I will suggest just one approach to this matter, Wittgenstein's (1953 and 1979) notion of perspicuous representation.

Wittgenstein's Perspicuous Representation

Wittgenstein's remarks on this topic stem from a critique of James Frazer's *The Golden Bough*. The basic issue revolved around how one ought to interpret so-called primitive beliefs and practices. Frazer's account had explained them as proto-scientific or erroneous scientific belief, that "magic is essentially false physics or, as the case may be, false medicine, technology, etc." (Wittgenstein 1979, p.67) According to Wittgenstein, however, Frazer was mistaken in his assumption that beliefs from one culture can have a direct correspondence with those of another — for instance, when we classify creationist "science" as being a component of religion in the Dewey Decimal 200-range. Such an interpretation is defective because the connections are asymmetrically motivated from within our own frame of reference. Wittgenstein's analogy is that work on one system cannot be accomplished using the tools of another (e.g. metric/English). We should also note that Wittgenstein had already abandoned the idea of an "objective world" and foundational reality in favor of multiple language games and systems of thought. Given then the incommensurability between systems (to whatever definitional extent required), trying to force an explanatory hypothesis from one set of beliefs onto another has the potential to seriously distort any knowledge we might be seeking to gain. Wittgenstein's alternative was the method of perspicuous representation.

According to Rhees (1979), this concept may be translated as "a way of setting out the whole field together by making easy the passage from one part of it to another." (Rhees, p.9) Wittgenstein

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

himself described it in terms of an arrangement of factual content that facilitates understanding by clearly showing connections, similarities, and intermediate cases. (Wittgenstein 1979, p.69 and 1953, §122) It contrasts with Frazer's method in virtue of its descriptive aim. "There must not be anything hypothetical in our considerations. We must do away with all *explanation*, and description alone must take its place." (Wittgenstein 1953, §122)

If we were to heed Wittgenstein's advice, how might this affect our approach to classification? First we would attempt to make our classification schemes more descriptive. But what does this mean? Although it may not be so very different from what we presently do, the emphasis would be less on properly situating some piece of information within a pre-existing schema and more towards presenting the data in an expressive arrangement. The criterion for arrangement would be something like coherence, an order in which all elements obtain some kind of holistic logic. In most cases the classification would need to be dynamic or what we sometimes call synthetic. Through development of multiple classification schemes, we would not be compelled to explain or locate an alternate belief in terms of a disparate scheme. Each belief system — e.g. creationist, evolutionary, or "2001: a Space Odyssey" — would maintain its own descriptive arrangement. Of course, this probably wouldn't work if those external to a particular conceptual framework attempted to classify its information, for we might still end up with creationism in the Dewey 200's.

The picture thus far is a relativist morass of competing and incommensurable classification schemes. Wittgenstein for one, however, was not willing to leave affairs in such an enigmatic condition. For him the key was connecting links.

The perspicuous representation brings about the understanding which consists precisely in the fact that we "see the connections." Hence the importance of finding *connecting links*.

But an hypothetical connecting link should in this case do nothing but direct attention to the similarity, the relatedness, of the *facts*. (Wittgenstein 1979, p.69)

The import for classification is that these links are not constructed, rather they are observed. A good arrangement makes the connecting links evident to those who want to make sense of diverse forms of knowledge. The job of the classifier would be to arrange information so that associations are facilitated, a task that would probably involve both formal and value related considerations. The job of identifying the actual links and using them falls to the user. The scenario seems somewhat vague, but on the other hand, it does not seem particularly out of keeping with the goals of automatic classification and IR. We want to set up finely organized databases so that the user can manipulate the data to suit his or her own purposes.

I have suggested a possible way in which classification could reconcile multiple conceptual frameworks. There was also the question about what kind of knowledge we could expect to ensue from our classification projects. Out of the universalist program, which was predicated on a particular view of knowledge, one can only assume that that very same sort knowledge is all that can be derived; in other words, no radically new knowledge. The juxtaposition of multiple schemas, on the other hand, clearly opens up possibilities for cross-pollination of diverse thoughts and ideas, providing means to a more heterogeneous and pluralistic body of knowledge.

Questions for Classification

- Are we or have we ever been committed to the realization of a universal classification scheme?
- Does it behoove us to assume the things we classify are true and accurate pieces of information, i.e. should we accept them at face value?
- If not, should we give greater credence to information that holds greater conformity with prevailing ideology?
- Should we and can we provide classification schemes for divergent frameworks? How would we do it?
- What exactly are Wittgenstein's connecting links and how might we deal with them?
- What are the optimum roles for classifier and user in the classification process?
- Is the data/theory relation necessarily asymmetrical? What would be a symmetrical model?
- Is knowledge production a primary goal of classification or a secondary outcome?
- Just how much relativity can we support without losing sight of the very purpose of organizing information?

SUMMARY: CLASSIFICATION AND PHILOSOPHICAL PURSUIT

The preceding discussions of perception theories, similarity, and conceptual schemas have been presented with the intent to open up avenues of inquiry for classification research. Key elements in classification, like-things and ordering principles, have been brought into juxtaposition with some philosophical notions. In response to the philosophical issues raised, a series of questions for classification has been posed.

One of the philosophically-based methods proposed was Wittgenstein's perspicuous representation, "a way of setting out the whole field together making easy the passage from one part of it to another." (Rhees 1979, p.9) This position paper has attempted to set out two fields of inquiry, philosophy and classification, so that the "connecting links" — so important for Wittgenstein — can be brought to bear.

BIBLIOGRAPHY

- Annis, David. "A contextual theory of epistemic justification." *American Philosophical Quarterly* 15 (1978) 213-19.
- Bliss, Henry Evelyn. *The organization of knowledge and the system of the sciences*. Henry Holt and Company, 1929.
- Dancy, Jonathan. *An introduction to contemporary epistemology*. Basil Blackwell, 1985.
- Frege, Gottlieb. "On sense and reference" (1892). Translated and reprinted in *Translations from the philosophical writings of Gottlieb Frege*, ed. by P.T. Geach and M. Black. Oxford University Press, 1960.
- Goodman, Nelson. *Problems and projects*. Bobbs-Merrill, 1972.
- Kuhn, Thomas S. *The structure of scientific revolutions*, 2nd ed. enl. University of Chicago Press, 1962.

PROCEEDINGS OF THE 5th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

- Kwasnik, Barbara H. "The role of classification structures in reflecting and building theory." *Advances in Classification Research* 3 (1994) 63-81.
- Langridge, D.W. *Classification: its kinds, elements, systems and applications*. Bowker-Saur, 1992.
- Pojman, Louis P. *The theory of knowledge: Classical and contemporary readings*. Wadsworth, 1993.
- Rhees, Rush, ed. *Remarks on Frazer's Golden Bough*. Humanities Press, 1979.
- Searle, John R. *Intentionality: An essay in the philosophy of mind*. Cambridge University Press, 1983.
- Van Rijsbergen, C.J. *Information retrieval*, 2nd ed. Butterworths, 1979.
- Wittgenstein, Ludwig. *Philosophical investigations: The English text of the third edition*. Macmillan, 1953.
- Wittgenstein, Ludwig. "Remarks on Frazer's Golden Bough." *Wittgenstein: Sources and perspectives*, ed. by C.G. Luckhardt. Cornell University Press, 1979.
- Wittgenstein, Ludwig. *Tractatus logico-philosophicus* (1921). Routledge & Kegan Paul and Humanities Press, 1961.