160

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

Classifying for Social Diversity

Rick Szostak

Department of Economics, University of Alberta, Edmonton, Alberta, T6G 2H4, Canada
<rszostak@ualberta.ca>

Rick Szostak is Professor of Economics at the University of Alberta. His research has long focused on facilitating interdisciplinary research and teaching. He is president of the Association for Interdisciplinary Studies. Knowledge organization has become the dominant element in that research agenda, and he has authored two books and several articles that develop classifications of things studied, theories and methods applied, types of data, ethical perspectives, research practices, and relationships. He has contributed to the Integrative Levels Classification and is developing the Basic Concepts Classification. He has argued in several places for the value of a classification grounded not in disciplines but in the things we study and the relationships among these. He is working with others on a book about interdisciplinary knowledge organization. See http://www.economics.ualberta.ca/en/FacultyandStaff/SzostakRick.aspx

Abstract: This paper argues that a new approach to classification best supports and respects social diversity. We should want a classification that facilitates communication both within groups and across groups. We should also want no group to be privileged within the classification. These goals are best accomplished through a truly universal classification, grounded in basic concepts, that classifies works in terms of authorial perspective. Strategies for classifying perspective are discussed. The paper then addresses issues of classification structure. It follows a feminist approach to classification, and shows how a web-of-relations approach can be instantiated in a classification. Finally the paper turns to classificatory process. The key argument here is that much (perhaps all) of the concern regarding the possibility that classes can be subdivided into subclasses in multiple ways, each favored by different groups or individuals, simply vanishes within a web-of-relations approach. The reason is that most of these supposed ways of subdividing classes are in fact ways of subdividing different relationships among classes.

1.0 Introduction

Classification necessarily involves structure, and thus there is a perhaps inevitable tension between classification and the recognition, appreciation, and support of social diversity. Since not all societal groups—gender, ethnic, sexual orientation, or socio-economic, among others—have been equally represented in the development of major classifications, a related concern arises that existing classifications privilege certain ways of looking at the world while obscuring others.

This paper will argue that it is both possible and desirable to develop a universal classification (for documents, and perhaps also for ideas) that reflects and supports social diversity. "Universal" should be taken here to mean integrated: a classification that follows the same approach and uses the same terminology throughout, rather than merely juxtaposing domain-specific classifications that follow different principles and employ different terminology, and thus do not encourage exploration across domains. It would be unfortunate to simply assume that such a project was not feasible. Indeed, it could be that some of the advocacy of the *exclusive* performance of domain analysis in the literature reflects an unstated view that each societal group is best served by its own classification. As I noted in Szostak (2013a), an argument that domain analysis is all that we can do should be carefully distinguished from an argument that it is all that we should want. They are, I would suggest, misguided for quite different reasons.

This paper will first explore the purpose of classification from the perspective of social diversity: what exactly should we want a classification to do in order to respect and support diversity? Once we have set goals for classi-

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

161

fication, we can then proceed to examine questions of structure: how should a classification be organized in order to achieve these goals? Finally we can turn to process: Is it feasible to develop such a classification?

2.0 Purpose

While there have been many critiques of existing practice in the field, there has been much less identification of what our goals should be with respect to social diversity, and then how these might best be achieved. We might start with some stylized facts with respect to social diversity:

- People are inevitably members of multiple groups, based on ethnicity, occupation, gender, sexual orientation, and a variety of other factors.
- Group boundaries are often fuzzy, though this varies by group type.
- In contemporary societies it is common to interact outside of one's groups as much or more than inside.
- It is generally difficult to identify the dominant perspective of a particular social group. [It is much easier to identify the dominant perspective of an academic discipline.] For example, there are many kinds of feminism; some women reject them all; and some men embrace feminist perspectives.

We could then ask a set of questions. This list is not exhaustive, but can guide our inquiry here:

- Do we want to emphasize access to literature written by (or ideas generated by) members of one's groups, other groups, or both?
- Do we want (if possible) to allow identification of authorial group membership, authorial perspective, or neither?
- Do we want to capture how a work might be perceived by different groups or from different perspectives, or do we want to capture the perspective/group associated with a work so that users can better judge its potential relevance, or neither?
- Do we want to encourage equivalent treatment of groups in our classifications, limit group references, or neither?
- Since groups might be thought of as (or like) domains, is diversity best served by exclusive reliance on domain analysis?

Though these stylized facts and questions are exploratory—other stylized facts and questions may be apposite—it is notable that the answers to each question will point toward an approach to classification quite different from that generally pursued. In other words, while the myriad and longstanding efforts to support social justice by reforming existing classifications are highly meritorious, they are by their nature limited in effect. If we really wish to have classifications that reflect our shared goals with respect to social diversity a more revolutionary classificatory project is called for.

We can begin with the first of the five questions above. If it was thought that each societal group should have unique and privileged access to its own literature, then domain analysis would be the obvious way to go. Each group could classify its own literature in terms of concepts with which non-group members would be unfamiliar—or better yet apply unique meanings to terms which others might mistakenly believe to be familiar. Outsiders would then find it extremely difficult to access the literature. Only those within the group, and possessing a clear apprehension of the true meaning of the concepts employed in the classification (grounded in turn in how the concepts are employed in the group's literature) would be able to readily navigate the literature. Classification would support group solidarity.

Of course domain analysis will only support group solidarity if the group in fact shares a set of definitions. It might be hoped that the project of defining a literature in terms of conceptual definitions loosely shared within a group might encourage greater consensus among group members. But in any case our concern for now is with the goals of a classification rather than feasibility.

The cost of such an approach would be (and is in practice) in terms of understanding across groups. Group members would have to master other classifications if they wished to read in the literatures generated by other groups. And they would have to publish in venues classified in other ways if they wished to speak beyond their group. Information scientists could generate translation devices, perhaps, but even these would be costly to master for each group one wished to engage.

Understanding across groups would be best facilitated by a truly universal classification, one that applies the same approach and terminology across domains. Then members of any group would have equal access to all literatures. The cost—again leaving aside issues of practicality for the moment—would be that group members would lack any special access to the literature of their own group: works authored by group members might be hard to distinguish from works on similar topics by others. Especially if the group were small members might legitimately feel that a universal classification militates against a sense of common cause and identity.

The challenge for information science is that the answer to our opening question in this section is almost certainly "Both." Only the most xenophobic would wish to

162

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

cut their group off completely from interaction with others. This is especially so given that individuals in the contemporary world increasingly have several cross-cutting group memberships. And even the most universalist in outlook can appreciate the value of people being able to readily communicate to others with certain shared characteristics. The lesson for information science is clear but rarely stated: we should strive to facilitate communication both across groups and within groups. And this result holds for any type of group: disciplinary, gender, ethnic, occupational/class, religious, sexual orientation, and so on.

Is it possible to pursue both goals simultaneously? I have argued elsewhere (Szostak 2010) that domain analysis and the pursuit of a truly universal classification can and should be complementary approaches (see also Kleineberg 2013). The key argument is that complex concepts—those that are understood differently across groups—can be broken into basic concepts that lend themselves to broadly similar understandings across groups (Szostak 2011). Domain analysis has a critical role to play in ensuring that each domain's terminology is accurately translated into basic concepts. If different groups and individuals will share broadly similar understandings of basic concepts then the biases that Mai (2010) argues must characterize any classification will be far less severe in a classification organized around basic concepts than in one organized around contested complex concepts. Basic concepts generally refer either to the things we perceive or study in the world or the relationships that we perceive among these. Works and ideas are then classified in terms of combinations of these basic concepts. Notably, a classification grounded in such basic concepts will not only aid users in finding works written by members of any group but will aid them in understanding those works (by translating complex concepts into basic concepts) (Szostak 2013c). But these arguments—though showing that domain analysis is far from incompatible with universal classification and that a classification can support understanding across groups—do not on their own provide an answer to our present question. There is a piece missing: can we signal within such a universal classification the group membership of authors? This was the second question listed above.

This second question raises in turn both philosophical and practical questions. Philosophically, authors may often not wish to be seen as speaking as a member of a particular group. And so we need to appreciate that attempts to classify works in terms of group membership have a potential downside. Authors may be striving to generate universal understanding. More pragmatically they may worry that group identification will blunt their ability to reach out to others. And they may have personal reasons for not wishing to advertise group membership. Allowing authors themselves to decide whether their works should be given any particular group identification provides an imperfect solution, and only for the living.

There is perhaps an even greater practical objection to such a project. As noted above, it is often difficult to identify the particular perspective associated with a particular group. Yet users are likely much more interested in finding works that express a particular perspective (say, a feminist perspective) than works that are written by members of a particular group. For both philosophical and practical reasons, then, it would be better to classify works in terms of authorial perspective rather than group membership.

Gnoli (2012) notes that information scientists have been talking about classifying works by authorial perspective for over a century. None that I am aware of has outlined a detailed classification of perspectives. Langridge (1989, 45-7) notes that the failure to classify by what he calls "viewpoint" often interferes with subject classification: a book on the Christian approach to education may be misclassified as on education about Christianity. He suggests that viewpoints such as "Christian" and "Marxist" deserve recognition and speculates on the value of classifying in terms of "philosophical viewpoints, such as rationalism, empiricism, pragmatism, realism, idealism, humanism, and agnosticism." He notes that some degree of subjectivity might be unavoidable in such an approach. He also appreciates that most/all viewpoints are also potentially subjects too and thus can be captured through a synthetic approach. One clear implication of Langridge's analysis is that group membership will often not be the best signal of perspective. Describing a work as "gender studies," or as applying "feminist theory," will send a more valuable signal than merely noting the gender of the author.

Some elements of authorial perspective will be captured by reference to the theories and methods that the author applies. Marxism, indeed, might be handled in this fashion. I have long urged the classification of works in terms of the theories and methods applied, and this was a key component of the León Manifesto (2007). Yet there are further elements of authorial perspective that merit recognition. Kleineberg (2013) has urged us to think in terms of the 'what' 'how' and 'why' of a work: the methods and theories employed would largely but not entirely reflect 'how' rather than 'why.' Further reflection regarding 'why' is thus called for.

A variety of dimensions might be useful in capturing the motives and beliefs (Kleineberg's 'why') of an author: inter/disciplinary; theoretical, methodological, rhetorical, epistemological, ideological, aesthetic, ethical. Feinberg (2011) speaks in particular of logical argument (manipu-

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

163

lation of evidence), ethos (incorporation of audience beliefs and values to establish trust), and genre adaptation (adjustment of formal elements). Feinberg suggested that classificationists in particular should confess their perspective, but we might well propose that authors also could express this. Clavier and Paganelli (2012) argue that we should classify works by stance: criticism, agreement, consensus, and so on. A key challenge for the information scientist is that there are imperfect correlations across these dimensions: not all who apply feminist theory are female, and feminist theory is applied outside of gender studies. A work classified along only one dimension will be missed by users searching along another.

In sum it is likely possible to enhance communication both across groups and within groups through a universal classification that classifies works in terms of authorial perspective. But our ability to achieve both depends on our developing a useful classification of authorial perspective. It may turn out that we cannot agree on a set of dimensions (or of a set of possibilities along those dimensions) that best capture authorial perspective. But we cannot know this until we try. It would seem likely that we should be able to classify at least some elements of perspective. If successful such an approach would allow a user to evaluate the relevance of a work by (among other things) reviewing the authorial perspective of the work. This would be useful whether the user was looking for works expressing a perspective similar to their own or was looking for alternative perspectives.

As noted in our third question above an alternative approach would be to try to signal how users from different perspectives might perceive a particular work. A variety of practical barriers might stand in the way of such an alternative. While we might rely on authors to self-declare their own perspective, we could hardly expect them to reliably estimate how others would perceive their work. Indeed it is hard to imagine how anyone could reliably do so. And it is not clear that different users with a particular perspective will always see a particular work from a different perspective in the same light. Last but not least, since perspective can be indicated along several distinct dimensions, evaluating each work in terms of each possible combination along these various dimensions would be unmanageable, even were it possible. For a variety of reasons, then, it seems best to indicate the perspective that the author believes they pursue, and let the user judge relevance.

The fourth question above reflects the fact that existing classifications often privilege certain groups. Male nurses may be given special treatment, while female nurses are assumed to be the norm. Information scientists have long sought to eliminate such biases from our classifications. What, though, if members of some

groups would wish that group references were eschewed entirely? They might, for example, wish that no effort was made to distinguish female from male nurses. Such a wish would necessarily confront the longstanding principle of literary warrant: if works are written about the special challenges facing male nurses, we must classify these even if we might wish that such special challenges did not exist. We can, though, essay to treat male and female nurses equivalently in a classification so that the classification itself does not reinforce cultural attitudes that certain combinations are anomalous.

Our fifth question has been addressed along the way, but deserves special note. It might be easy to assume that domain analysis—with its special examination of the terminology of a particular group—is ideally suited to the recognition and support of social diversity. While domain analysis does indeed have an important role to play in ensuring that the terminology employed within any group is reflected in a classification, communication across groups will be difficult unless this is supplemented by the pursuit of some sort of truly universal classification.

As a segue to our next section, it is useful to engage here with arguments made by Mai (2011). He argues that contemporary approaches to classification (grounded in ontology) reflect a modernist view that imagines in a realist fashion that the things we study exist separately from those who study them. He instead recommends an epistemological approach to classification that appreciates subjectivity. Though I am epistemologically more confident than Mai that consensus is possible due to our ability to fairly accurately apprehend reality, I can nevertheless appreciate that Mai provides a further justification for classifying works by perspective: this will help to identify some of the biases that an author brings to the work (it will not cope, though, with the perceptual and cognitive biases that all humans share; see Szostak 2004 for a classification of all scholarly biases). But Mai is not sure what a classification grounded in subjectivity would look like (nor is Hjørland 2012). He might thus be skeptical of our ability to classify perspectives in a manner that would respect all perspectives. But this is an empirical question that is best evaluated as we strive to develop a classification of perspectives.

More generally, Mai doubts that there can be consensus on the classes within any classification. Though I am again much more optimistic than Mai, it will prove useful to try to meet this concern as much as possible. That is, if different people or groups will (even just sometimes) disagree over the nature of classes, then we should strive to minimize the scope for disagreement. This we will do in each of the next two sections by simply limiting the degree of hierarchical organization, and focusing on the

164

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

classification of basic concepts. We will in section 4 outline a strategy for reducing/eliminating the problem of multiple ways of subdividing a class into subclasses. Mai also urges transparency: it should be clear how a classification was developed, so that the user can evaluate whether/what biases drove its development. This principle will also guide us.

The thrust of Mai's argument is that we may need different classifications in order to reflect the biases of different groups or individuals. As we have seen above, such an approach risks limiting communication across groups, and thus reinforcing group biases. We will strive in what follows to limit bias as much as possible within a universal classification. This is our best and perhaps only hope of encouraging communication both within groups and across groups.

3.0 Structure

We have identified above some characteristics to be sought in a classification. We should seek a universal classification informed by domain analysis. We should, if possible, classify works by authorial perspective. We should develop a classification that by its nature does not privilege one group over another. But what sort of structure will best achieve these goals? We start this section by responding to a feminist critique of classificatory practice, and then showing that—not surprisingly—the classification that responds to this critique serves the various goals outlined above.

Olson (2007) suggested that hierarchy is more reflective of a masculine perspective, and that a classification that blended hierarchy with a web-of-relations approach would be more gender-neutral. Women, she argues, are more likely to see the world in terms of a web of relations. But Olson also argues that all underprivileged social groups would likely also benefit from a less hierarchical approach to classification.

3.1 Instantiating a web-of-relations approach

Is such a classification possible? This critical question has not been addressed in the detail that it deserves. It can only be addressed by outlining in some detail the sort of classification that would instantiate a web-of-relations approach. Such a classification is indeed possible. But it requires a synthetic approach that spans disciplinary or group boundaries and is grounded in basic concepts that lend themselves to broadly shared understandings across groups and individuals. We address these three characteristics—universal, basic concepts, and synthetic—in order.

Though Olson does not explicitly urge a universal classification, any web-of-relations approach to classifica-tion will be frustrating for the user if their desire to follow their curiosity from one topic to another is regularly interfered with by tears in the web. A user interested in why dogs sometimes bite mail carriers does not want to have to figure out where dogs and mail carriers are classified, but wishes to move seamlessly from one to the other. Though domain-specific webs are surely possible, the inter-connected nature of the world in which we live will ensure that users will frequently wish to move from one such web to another. They might be provided with translation devices—perhaps some sort of pidgin or Creole—between any two domains, but learning new translation devices for each new domain confronted will of necessity be time-consuming and annoying. A web-of-relations approach will work far better with one big web than multiple little ones.

The faceted classifications of the past—such as Colon or Bliss—were grounded in disciplines and thus necessarily treated combinations within disciplinary classes differently from combinations that spanned disciplinary boundaries (Langridge 1989 urged a move away from disciplines, but then stressed the primacy of ten forms of knowledge). Such classifications would thus arbitrarily constrain a web-of-relations approach. The more recent Integrative Levels Classification (see www.iskoi.org/ilc) and Basic Concepts Classification (Szostak 2013b) are grounded in the things we study rather than disciplines or groups. Their approach is similar in many ways to that long pursued by the Classification Research Group; such an approach is much more feasible with digitization. Both of these classifications also seek to stress basic concepts. Classifications grounded in complex concepts do not lend themselves to a web-of-relations approach: the nodes on such a web need to be simple and unique whereas the elements of complex concepts necessarily overlap. The BCC will be stressed here for its particular approach to synthesis may best instantiate a web-of-relations approach (the BCC approach is faceted but the logic of its approach means that facet indicators are not necessary (see Szostak 2013b for a discussion of how each facet in Bliss and ILC is captured, and also for a discussion of how works could be classified and searched for with BCC); this should facilitate the user's movement from one topic to a related topic.)

Szostak (2011) argued that the complex concepts that are understood differently across groups or indeed individuals can generally be broken into basic concepts that lend themselves to a much greater degree of shared understanding. That paper was grounded in the philosophical literature, but noted that information scientists do not need the degree of precision long sought unsuccessfully by philosophers. Philosophers have been frustrated at their inability to precisely define concepts such as 'free-

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

165

dom.' Information scientists need not the same precision in order to guide users to relevant works. Moreover philosophers have naturally focused on the most difficult concepts. The theory of conceptual atomism tells us that we will achieve the greatest degree of shared understanding regarding things and relationships that we perceive in the world. The BCC (and ILC) focuses wherever possible on classification in terms of such basic concepts. An example may help here. 'Globalization' is a concept with multiple meanings. But it comprises a variety of concerns, including the effect of (expanded)(trade) on (employment) or (wages), or (watching)(American)(movies) on (French)(cultural attitudes). A far greater degree of shared understanding can be expected regarding the terms in parentheses than can be achieved with respect to 'globalization' itself. Still further clarity is provided by placing these basic concepts within clear logical hierarchies: we can, for example, provide a fairly exhaustive list of what are considered to be 'cultural attitudes.' And we can whenever possible classify works or ideas with respect to particular cultural attitudes. What about a concept such as 'time': surely some societies see time as cyclical and others as unidirectional? Yet time passes for each, and can be linked synthetically to cycles. Differences in meaning across groups can generally if not always be accommodated through this type of (synthetic) clarification.

The key to the web-of-relations approach lies in then classifying works synthetically in terms of combinations of basic concepts. One work might be classified in terms of how a phenomenon A influences a phenomenon B in a particular manner Z. Such a work will be found easily by anyone interested in how A might affect B in manner Z, regardless of the user's group membership. It can also be found easily by anyone studying how B influences C who then becomes curious about how to encourage changes in B. And someone interested in how F influences G in manner Z might become interested in other cases of influence of type Z. So this sort of compound classification utilizing basic concepts in fact instantiates a web-of-relationships at the level both of works and of the key arguments expressed in works. A user can thus follow, if they wish, a complex set of issues from one work to another. As Olson notes, present classifications facilitate browsing only within a hierarchy; the proposed structure also facilitates browsing across hierarchies.

It might be worried that a different user would think the work addressed in the preceding paragraph was instead about how C affected D in manner X. There is of course some ambiguity and subjectivity in classifying works in any classificatory system (as Mai 2011 suggests). Langridge (1989, 9), at least, would argue that there is a clear answer to the question of what a work is about, though it may in practice be difficult to agree on what

this is for certain works. The point to stress here is that such ambiguity is lessened by classification in terms of basic concepts. If a particular work really can be viewed as about A, B, and Z by one user, and about C, D, and X by another, then no classification will satisfy everyone. Attempts to place the work within *Library of Congress Classification* or *Dewey Decimal Classification* will be at least as problematic. But whereas authors cannot be expected to master LCC or DDC, they can reasonably be expected to say what causal relationships are key to their own work. The classifier can judge how many of these are dealt with in enough detail in the work as to merit classificatory treatment.

Hierarchy is still necessary in such a classification, but to a much lesser extent. Types of influence can be captured through combinations of some 100 basic types of influence that can be organized in just two levels of hierarchy (occasionally three; see Szostak 2012). The things we perceive can, at least in the human sciences, be captured in very compact schedules (Szostak 2011, 2013b). Natural science—biology especially, but to a lesser extent chemistry—requires much more detailed hierarchies of species and chemical compounds.

3.2 Additional concerns

Olson is concerned that hierarchical approaches tend to privilege "being a Y" over "not being a Y" in general, and "being male" over "not being male" in particular. A classification that did not distinguish males from females in any class except gender itself could obviate this concern. In place of present practice, in which male nurses and female engineers are treated as some sort of anomaly, the classification here would use linked notation: (nurse)(male) and (nurse)(female) would be classificatorily equivalent—as indeed would be (nurse)(transgendered).

Olson worries that hierarchy privileges deduction over induction. I have long argued that the best approach to classification blends induction and deduction. [Olson and I use induction broadly such that it would include Charles Sanders Pierce's concept of abduction (see Fann 1970); it thus involves any attempt to formulate hypotheses from non-conclusive information.] And indeed this is one key reason for urging us to blend universal and domain-analytical approaches (see Szostak 2010), for a universal classification demands some logical structure whereas domain analysis is inherently inductive. I will urge a very deductive approach to hierarchy below. Allowing elements in any hierarchy to be freely linked with elements in any other hierarchy provides immense scope for an inductive appreciation of any connection drawn in any literature, as long as we ensure that any thing or relationship discovered in the literature is represented in some hierarchy (Szostak 2013b).

166

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

Breaking complex concepts into basic concepts mimics the approach taken in developing a "parts of" type of hierarchy. The approach to classification urged here encourages classification in terms only of the basic concepts themselves. Importantly, a particular basic concept can be combined with others to generate multiple complex concepts. Yet each basic concept is itself a subclass of one and only one generally compact hierarchy.

Olson also worries that the logical philosophy that underpins hierarchy privileges reason over emotion and intuition, and assumes away bias. I concur that emotion and intuition are important parts of the process of discovery (Szostak 2002), and have attempted to classify the types of bias that characterize scholarship (Szostak 2004, chapter 5). Classifying works in terms of authorial perspective will, as noted above, provide some insight into potential biases; it may also tell us something about the particular role of emotion and intuition in a work. Allowing free combination will—as in the male nurse example above—provide a powerful antidote to bias. And it will be argued in the next section that an emphasis on combinations reduces and may even eliminate the biases that creep into hierarchies themselves.

Though Olson did not describe in detail what her recommended classification would look like, she did appreciate that it would rely heavily on a synthetic approach. She noted that even when a synthetic approach is pursued within contemporary classifications some combinations are privileged over others. It is thus critical that it be possible to freely combine any set of concepts. Only the ILC and BCC allow all concepts to be freely combined in precisely the same fashion.

Of particular note, Olson appreciates that existing classifications [The ILC and BCC were not within her purview here] handle paradigmatic relationships best. Yet since paradigmatic relationships, such as those between elements in a logical hierarchy, are enduring, it is generally not necessary to devote much energy to identifying such relationships. It is syntagmatic relationships, where the connection is not essential, as in embroidery of Christmas ornaments, which we will often wish to express [search for], but these are handled poorly. Boolean searches will yield many hits that do not capture the desired relationship. Again, the solution involves allowing us to freely connect any set of concepts both in classifying a work and in searching.

The sort of classification outlined here, which addresses each of the concerns raised by Olson, not surprisingly serves also the goals identified in the second section of this paper. It is universal: as noted above the web-of relations approach will frustrate the user if there are boundaries between domains. It facilitates exploration and understanding across groups, by relying on combina-

tions of basic concepts that are broadly understood in similar ways by members of diverse groups. It utilizes domain analysis to ensure that the complex concepts of any group are translated into basic concepts. It thus severely limits the scope for classificatory bias. It is amenable to classification by authorial perspective (or group membership) because it allows any concepts to be combined. It treats all groups equivalently of necessity, since in all cases references to any group occur through a synthetic linkage to lists of genders, occupations, ethnicities, and so on.

Olson is critical not just of classification but of standard practice in constructing thesauri. Hierarchical relationships are captured fairly well by the terminology of BT (broader term) and NT (narrower term)—though such terminology does not distinguish "type of" hierarchy from "part of" hierarchies or other types of hierarchical division. We could aspire to identify different types of BT and NT. A host of different relationships are lumped together as RT (related term). But this need not be: we could aspire to recognizing several key types of relationship. Olson notes that the thesaurus construction standard, ANSI/NISO Z39.19, provides for a limited set of allowed RT relationships: process/agent, process/counteragent, action/property, action/product, action/target, cause/effect, concept or object/property, concept or object/origins, concept or object/measurements, raw material/product, and discipline or field/object or practitioner; and also antonyms (plus a few arcane exceptions) The standard allows these to be explicitly indicated on a local basis. But why not insist that these and others are always designated?

3.3 Advantages beyond social diversity

The preceding paragraphs have argued that a certain sort of synthetic approach to classification instantiates a web-of-relations approach and thus serves each of the diversity goals outlined above. But of course social diversity is not the only consideration that a classificationist must take into account. A classification that supports diversity but is shunned by users would be of little use. A classification that is useful but somewhat less supportive of diversity might then better achieve social justice. Any classification must be judged in terms of a broad array of both philosophical and practical considerations (see Szostak 2015). It is thus worth noting that the sort of classification (and thesaurus) urged here has many further advantages. I have argued (2011) that breaking complex concepts into basic concepts is the ambiguity-minimizing strategy in classification. Several other advantages can be briefly noted (see Szostak 2013b):

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

167

– Because most works can be classified as links between phenomena, we are able to achieve very precise classifications with limited and expressive notation.
– Users are thus better able to find precisely what they want, whether they wish to search in one discipline or across all.
– By distinguishing different sorts of relationship (especially causation/influence), we enable searches by verb-like terms as well. Friedman and Smiraglia (2013) find that most concept maps employed in knowledge organization have nouns as nodes and verbs as arcs. But our classifications do not reflect this strategy.
– While other classification systems provide specific instructions in multiple places for coding by time or place or people, this system has a universal coding for such elements. This renders both classification and searching easier.
– Note that the use of linked notation serves to place works [but not individual concepts] within multiple hierarchies (and of relations as well as things).
– It should be possible to translate all search or entry terms employed in other classifications into basic concepts. The system may provide a solution to the fact that online databases employ a bewildering array of classification systems.
– Note that in addition we create the possibility of (fairly) automatically coding for new works or for existing works that are at present poorly classified.
– Such a classification may also serve as a bridge between other classifications. Yi and Chan (2010) explore the possibility of rendering *LCSH* interoperable with other systems. They criticize *LCSH* both for inconsistent application of hierarchy and for unclear semantics and syntax.

Moving forward it would be useful to show how this classification could be applied to a variety of works. The various advantages outlined above are best appreciated and evaluated in application. Such an approach should look beyond books and articles, and engage also with museum artifacts and archival documents and websites (Szostak 2014 suggests applicability to the semantic web). It could well be that a universal and synthetic approach grounded in basic concepts will have appeal both to (especially digital) libraries and to other information repositories that have tended to shun complex library classifications.

Lambe (2007) also appreciates that hierarchy is not the only way to classify (he mentions matrices, system maps, and facets), and is often not the best. He appreciates that our goal is to show how things are related. He suggests that users by looking at a classification (he stresses taxonomies) should gain a sense of how things connect. Classifications should also serve as artificial memory aids:

helping us to remember things by relating them to others. It deserves to be stressed that a classification that relies on combinations across a very manageable set of schedules is both much easier to master and to understand. Most users approach subject headings within existing classification systems with no understanding of how these are generated or related to each other. Indeed hierarchical tree structures have to be identified by researchers rather than being provided by the classificationist (see Julien et al. 2013). The suggested classification is purposely transparent.

Lambe addresses in detail the fact that many concepts appear within multiple hierarchies in existing classifications. He is very critical of this practice, arguing that a hierarchical approach becomes too complicated if concepts appear in many places (see also Soergel 1985, 254-6). It is thus better to capture this sort of situation in other ways than through hierarchy. This is precisely what we have done. Individual concepts appear in only one hierarchy, but complex combinations of these link all hierarchies. And thus our approach serves diversity while also generating a less problematic classification.

It could also be that the approach urged here will close the gap between the fields of classification and information retrieval. Scholars of information retrieval increasingly disdain the "bag of words" assumption driving many search techniques: that the concepts being searched for occur independently. They appreciate that users search for combinations of concepts (e.g. Mengle and Goharian 2010, Khoo and Na 2006). Though search engines generally ignore existing classification systems, they might find a classification which stresses such combinations useful. Hjørland (2012), in speculating on the role of classification "after Google," says much that is consonant with the approach recommended here: that information scientists should work on an overall structure that somehow connects domain analyses, that the key is the semantic relations between concepts (though he at times stresses hierarchy), and that documents should be classified not in terms of simple aboutness but rather what a reader would find useful/novel in them.

Last but not least, Börner (2006) suggests that in the near future scholars might just add 'nuggets' or 'nodes' to the web of knowledge. That is, the present practice of writing stand-alone papers will be replaced by a practice of adding insights to a pre-existing structure. She reviews various efforts over the last century to develop links between related bits of information (such as citation indices). New technology creates an opportunity to finally achieve this goal. But search engines are like inserting a needle in a haystack, and usually do not place search results in context: they "fail to equip scholars with a birds-eye view of the global structure and dynamics of schol-

168

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

arly knowledge and expertise" (186). The sort of classification here can be used both to classify works and ideas (a desiderata noted by Gnoli 2008). It would thus be congenial to the sort of shift foreseen by Borner, such that any author's ideas can readily be related to the ideas of other authors. But the structure's fluidity would mean that classification itself does not privilege certain ideas over others.

4.0 Process

The system as outlined above allows the free combination of concepts across any hierarchies (of both things and relationships). How, though, are these hierarchies developed? Though we have reduced the need for hierarchy, it is nevertheless important to ensure that these hierarchies themselves still reflect our goals with respect to social diversity.

Olson has also often used a "slicing pizza" analogy. We can subdivide classes into subclasses in multiple ways. Different groups may wish to slice their pizza in different ways. This conundrum seems insoluble. Any use of hierarchy in classification must of necessity privilege one way of slicing the pizza. And Hjørland (2012) notes that there are no good guidelines on how to pursue an interpretivist approach to classification, once one denies there is one best way to subdivide.

But it is in fact quite straightforward to address this problem within the sort of classification urged here: most if not all possible approaches to slicing any pizza can in practice be addressed simultaneously within a web-of-relations approach. As noted above, the web approach significantly lessens the need for hierarchy. In particular, it eliminates the oft-noted practice (e.g. by Mazzocchi et al 2007), common in all major classifications, of abusing hierarchy such that causal arguments (or other sorts of relationship between phenomena) are treated as if they were a proper subset of some phenomenon. Recycling need not be treated as a subclass of garbage because there is no other place to put it if a synthetic approach to combining things and relationships is taken.

More centrally, when hierarchy is employed in the type of classification recommended here, it is usually subdivision in terms of "type of" (but occasionally "parts of") that is called for. And "type of" is best defined functionally for social phenomena (so that institutions, for example, are classified in terms of their official purpose) and in terms of their essence for natural objects (so that species are organized in terms primarily of genetic inheritance, and chemical compounds in terms of constituent chemicals).

Yet surely this privileges this one slicing strategy? While there are other reasons for pursuing this particular slicing strategy, the one to stress here is that most/all

other ways of slicing the pizza can be easily captured through combinations. It has often been noted, for example, that pharmacologists might want to classify drugs in terms of physiological effect, while chemists will want to classify them by chemical composition. The former can easily be rendered as, say (chemicals)(reduce)(blood pressure). The latter can only be captured through a "type of" approach. In other words, the classification wished for by pharmacologists is a classification of relationships, whereas the classification sought by chemists is a classification of subsidiary types of real things.

We can then proceed to classify chemicals the way chemists would. A pharmacologist seeking to reduce blood pressure can nevertheless readily find all works on (chemicals)(reduce)(blood pressure). Such works will likely report on successes achieved with some chemicals and perhaps failures experienced with others. The pharmacologist may wish to explore chemicals that are similar in some way to those that have proven successful in the past. In doing so they can be guided both by the chemist's understanding of chemical composition and the pharmacologist's understanding of relationships.

The claim here is strong. We have collectively imagined or at least exaggerated the challenge of slicing because we have abused hierarchy in order to capture relationships. Once we handle relationships as relationships, the slicing conundrum is alleviated and may even disappear. We must always be careful of reaching empirical conclusions on the basis of theoretical arguments alone. The arguments made here must be tested in practice. My own efforts to develop the Basic Concepts Classification suggest that we only rarely confront choices about how to slice (see Szostak 2013b). But this conclusion needs to be verified by others who may bring different slicing preferences to the task. What is clear on theoretical grounds is that we can substantially reduce our slicing choices by treating relationships as relationships.

Our success here reflects the simple fact that most scholarly works—and likely most general works as well—express some sort of causal relationship: (chemicals)(reduce)(blood pressure) or (dogs)(biting)(mail carrier). The best way to capture the subject of such works is to synthetically link things and relationships. The minority of works that examine the properties of a particular thing (or perhaps relationship) are also best dealt with synthetically: (steel)(is)(strong).

Mai (2010) argues that information science has long but mistakenly assumed that we were searching for the one best classification, and that general rules and commonalities existed that needed to be identified. As noted above, he argues that bias is inevitable. But it is useful to explore here the precise arguments he makes with respect to what we have termed "slicing the pizza." First, he

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

169

notes that likeness is not a quality of things but a relationship between them; we can find some similarity between any two things (e.g. plum and lawnmower). But what sorts of similarities exist between a plum and a lawnmower? Perhaps color, perhaps uses to which they can be put, perhaps places they are stored. All of these can be captured through relationships. The only singular class of which they are "types of" is "things." It is thus not true that there exists some logical classification of things in which plums and lawnmowers would appear within the same subclass. Likewise, Mai notes that a cow can be treated as a type of food, an animal, and so on. But again we can handle these diverse classifications through relationships. We need again to be careful of leaping to an empirical conclusion, but it must seem that many/most/all types of likeness can be handled through a web-of-relationships approach. And indeed Mai's own words—that likeness is a quality of relationships—suggest that this is so.

Mai then argues, following Hjørland, that a stone in a field has information of different types for different users and thus we cannot hope to classify all of these; no one mapping is the true mapping. But we can clearly use relationships to capture (at least) many of these for they reflect different uses to which the stone can be put: mining, building, skipping, and so on. It is noteworthy that the word "mapping" is used here in the sense of one-to-one mapping when the solution is a map that shows all relationships and thus allows one concept to be mapped to many.

Finally, what about an area where there is intense scholarly controversy, such as in defining types of mental illness? Psychologists disagree about how this is best done (see Cooper 2011). Even here a web-of-relations approach has much merit. Some psychologists would classify in terms of physiological symptoms and others psychological symptoms. Some would look for common causes, others for common effects. Rather than choosing one way of classifying mental illness, it would be better to employ relationships to capture all.

5.0 Concluding remarks

Though the three issues of purpose, structure, and process were addressed separately, the analyses are complementary: there is one approach to classification that addresses all three. It is thus possible to develop a classification that respects and supports social diversity. Such a classification must be universal, must be grounded in the basic concepts that we perceive, must allow the free combination of all things and relationships, and should seek to classify works with respect to authorial perspective. Such a classification serves the goals outlined in the second section: it supports

communication both within and across groups, does not privilege any group, and blends domain analysis with a universal approach. It instantiates the web-of-relations approach advocated in the third section; indeed only a universal approach that involves the free combination of basic concepts can fully instantiate a web-of-relations approach. It thus alleviates many of the problems associated with hierarchy by reducing the need for hierarchy. In particular the synthetic approach addresses the important concern, addressed in the fourth section, that different groups and individuals might prefer different approaches to breaking a class into subclasses.

Social diversity is one of many goals to be pursued in classification. Happily, the approach to classification recommended in this paper also has many other positive attributes. These were briefly reviewed in the third section. There is thus no conflict between supporting social diversity and pursuing various other goals of knowledge organization.

References

Börner, Katy. 2006. Semantic association networks: using semantic web technology to improve scholarly knowledge and expertise management. In Geroimenko, Vladimir and Chen, Chaomei, eds., *Visualizing the semantic web 2nd ed.* Berlin: Springer, pp. 183-98.

Clavier, Viviane and Paganelli, Céline. 2012. Including authorial stance in the indexing of scientific documents. *Knowledge organization* 39: 292-9.

Cooper, Rachel. 2011. Some classifications will be natural. *Knowledge organization* 38: 398-404.

Fann, K. T., 1970. *Peirce's theory of abduction*. The Hague: Martinus Nijhoff.

Feinberg, Melanie. 2011. How information systems communicate as documents: the concept of authorial voice. *Journal of documentation* 67: 1015-37.

Friedman, Alon and Smiraglia, Richard P. 2013. Nodes and arcs: concept map, semiotics, and knowledge organization. *Journal of documentation* 69: 27-48.

Gnoli, Claudio. 2008. Ten long-term research questions in knowledge organization. *Knowledge organization* 35: 137-49.

Gnoli, Claudio. 2012. Metadata about what? Distinguishing between ontic, epistemic, and documental dimensions in knowledge organization. *Knowledge organization* 39: 268-75.

Hjørland, Birger. 2012. Is classification necessary after Google? *Journal of documentation* 68: 299-317.

Julien, Charles-Antoine, Tirilly, Pierre, Dinneen, Jesse and Guastavino, Catherine. 2013. Reducing subject tree browsing complexity. *Journal of the American Society for Information Science & Technology* 64: 2201–23.

170

Knowl. Org. 41(2014)No.2
R. Szostak. Classifying for Social Diversity

Khoo, Christopher S. G. and Na, Jin-Cheon. 2006. Semantic relations in Information Science. *Annual Review of Information Science and Technology* 40: 157-228.

Kleineberg, Michael. 2013. The blind men and the elephant: towards an organization of epistemic contexts. *Knowledge organization* 40: 340-62.

Lambe, Patrick. 2007. *Organising knowledge: taxonomies, knowledge and organisational effectiveness*. Oxford: Chandos.

Langridge, Derek W. 1989. *Subject analysis: principles and procedures*. London: Bowker-Saur.

León Manifesto. 2007. Available www.iskoi.org/ilc/leon.htm

Mai, Jens-Erik. 2010. Classification in a social world: Bias and trust. *Journal of documentation* 66: 627-42

Mai, Jens-Erik. 2011. The modernity of classification. *Journal of documentation* 67: 710-30.

Mazzocchi, Fulvio, Tiberi, Melissa, De Santis, Barbara and Plini, Paolo. 2007. Relational semantics in thesauri: some remarks at theoretical and practical levels. *Knowledge organization* 34: 197-214.

Mengle, Saket S.R., and Goharian, Nazli. 2010. Detecting relationships among categories using text classification. *Journal of the American Society for Information Science & Technology* 61:1046–61.

Olson, Hope. 2007. How we construct subjects: a feminist analysis. *Library trends* 56: 509-41.

Soergel, Dagobert. 1985. *Organizing information: principles of data base and retrieval systems*. San Diego: Academic Press.

Szostak, Rick. 2002. Intuition and interdisciplinarity: a reply to Mackey. *Issues in integrative studies* 20: 131-7.

Szostak, Rick. 2004. *Classifying science: phenomena, data, theory, method, practice*. Dordrecht: Springer.

Szostak, Rick. 2010. Universal and domain-specific classifications from an interdisciplinary perspective. In Gnoli, Claudio and Mazzocchi, Fulvio, eds., *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the 2010 conference of the International Society for Knowledge Organization, Rome, February 2010*. Advances in knowledge organization 12. Würzburg: Erlon Verlag.

Szostak, Rick. 2011. Complex concepts into basic concepts. *Journal of the American Society for Information Science & Technology* 62: 2247-65.

Szostak, Rick. 2012. Classifying relationships. *Knowledge organization* 39: 165-78.

Szostak, Rick. 2013a. Speaking truth to power in classification. *Knowledge organization* 40: 76-7.

Szostak, Rick. 2013b. *Basic concepts classification*. http://www.economics.ualberta.ca/en/FacultyandStaff/~/media/economics/FacultyAndStaff/Szostak/Szostak-Basic-Concept-Classification2.pdf

Szostak, Rick. 2013c. Communicating complex concepts. In O'Rourke, Michael, Crowley, Stephen, Eigenbrode, Sanford D. and Wulfhorst, J. D., eds., *Enhancing communication and collaboration in interdisciplinary research*. Thousand Oaks: Sage.

Szostak, Rick. 2014. Classification, ontology, and the semantic web. Paper presented at the ASIST SIG/CR, Montreal, November, 2013, Advances in classification research, Available https://journals.lib.washington.edu/index.php/acro

Szostak, Rick. 2015. A pluralistic approach to the philosophy of classification. Prepared for a special issue of *Library trends*.

Yi, Kwan and Chan, Lois Mai. 2010. Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. *Journal of the American Society for Information Science & Technology* 61: 677–87.