LEARNING IN SETS

While the network has proved to be a useful structural principle and, in a minimal technical sense, underlies every social form enabled by the Internet and other networked technologies, and groups are well founded in practice and literature, they are not always the most useful way to look at the social structures that emerge in cyberspace. Sometimes we do not know people in any meaningful way, so “network” is too strong a word for our engagement, and sometimes we are not members of shared groups, yet people can make a big difference to our learning. In this chapter we will describe how the set, a simple aggregate of people and the artifacts they produce, can provide meaningful learning opportunities and how it differs from group and net social forms. Unlike previous chapters on nets and groups, there is not a copious bounty of literature to call upon that discusses sets because few, if any, researchers have explored their use in a learning context.

This is uncharted ground, and much of what we write here will be relatively new in academic literature, though the set has not gone unnoticed by the blogging community and popular press, nor by millions of users of social media. Eldon (2011), for example, observes that set-oriented social interest sites such as Twitter, Tumblr, and Pinterest have experienced massive growth. These are still often inaccurately referred to as “interest-based social networks” (Jamison, 2012): just as early network-oriented applications were called as “groupware,” there is a tendency to see systems through the lenses of what we are familiar with, and we are currently familiar with social networks. Though under-researched as a social form, especially for learning, sets are important. It is not accidental that relational database technologies used to store and retrieve information about people and things in the world are based on sets, because categories matter, both to people
and machines. To a significant extent, the ways that we categorize the world shape our experience of it, and represent what we know of it (Hofstadter, 2001; Lakoff, 1987; Wittgenstein, 1965). We do not just know. We know things, which fit into categories, and this is important. As Lakoff (1987) puts it, “Without the ability to categorize, we could not function at all, either in the physical world or in our social and intellectual lives” (p. 6). Categories according to Lakoff, were classically seen as things ‘in the world’ that we could simply identify through their common properties. Wittgenstein (1965) both problematized the issue and slightly sidestepped it by suggesting that, for at least some of the categories that we use, this is simply not true. Instead there are family resemblances in which things we identify through a single category may share some but never all of the same properties, and the boundaries that we put around particular categories are not fixed but socially constructed. More recently, thinkers such as Lakoff and Hofstadter (2001) have shown the deep psychological, social and linguistic complexity of the ways that we categorize things, showing how metaphorical meanings are not just a feature of language but fundamental to understanding, without which we would be unable to build cognitive models of the world around us. Categories allow us to symbolically represent collections of things in ways that are meaningful to our social, intellectual, and practical needs, while letting us extend our understanding across fuzzy boundaries, making connections and drawing analogies from which we construct our knowledge. In many ways, knowing the right names of things is a crucial step towards understanding them. This has an important pragmatic consequence in the context of the current enquiry: when seeking to learn, especially in academic disciplines, we typically begin by thinking of topics, areas, or categories into which our new knowledge can be classed and named.

DEFINING THE SET

Sets as a social form are made up of people with shared attributes. There are indefinitely many attributes that may be shared by individuals, which may be specific or relate to coming within a range of values: location, height, IQ, choice of automobile, and so on. Most of these attributes will be of little value to a learner, but some might. In learning, particularly useful set attributes might include a shared interest in a topic, a shared location, a qualification in a particular subject area, or a shared outlook.

In order to be useful, it should be possible to identify a set and to interact or share with people in it. In this sense, there is a minimal requirement for a
mechanism for sharing and communicating with others in a set. Like groups and nets, sets rely on a substratum in which they are situated and observable.

**Sets are About Topics and Themes**

The notion of the set bridges both people and things. For instance, one may find resources that are part of the set of writings about networked learning, as well as people with an interest in networked learning. The social form of the set simply refers to any collection of people, and in a learning setting this is often related to artifacts that they produce or seek. In typical cases, what causes us to identify the set is the topic, artifact, place, or site around which they aggregate.

A concrete example of this is a page on Wikipedia. While groups and networks can and do develop around Wikipedia pages, the central thing that draws people to both edit and read a Wikipedia article is an interest in the topic it addresses. Beyond that, there need be no social engagement, no direct communication, no exchange of information, not even a shared purpose. The boundaries of this particular social set are the page, and beyond that boundary is everything else. While networks and groups may develop in support of topics and pages, and various inducements are provided by the site to reveal one’s identity, such as greater editing rights, the ability to move articles or participate in elections, they are not a necessary feature of engagement with others on Wikipedia.

People may simply be identified by IP address which can be entirely anonymous (for instance, if an edit is made in an Internet café). This is not an uncommon occurrence. In one survey reported on Wikipedia itself of editors who made 500 or more edits (placing them among the most prolific), 5 out of 67 editors were identified only by IP address, not name (en.wikipedia.org/wiki/User:Statistics#Case_1:_Anon_Surprise.21). In 2005, Voss found that, across different language sites, anonymous edits accounted for between 10% (Italy) and 44% (Japan) of all edits made. It is notable, however, that it is increasingly difficult to find such statistics in recent research papers. The strong academic focus on networking in most research publications on social software means that anonymous edits are often deliberately excluded from results of studies (e.g. Nemoto, Gloor, & Laubacher, 2011; Wöhner, Köhler, & Peters, 2011) which tells us more about the biases of researchers than the nature of Wikipedia.

Similarly, when we create hashtags for public posts in Twitter, they are a signal that defines a set for anyone with an interest in the topic defined by that hashtag. When we search for such a hashtag, we rarely have any particular interest in or knowledge of the people that created it: they are just a set of people who have
tweeted about that topic. Of course, Twitter supports a profound net form as well with the mechanism of following but, through the hashtag, it is equally powerful as a means to support sets.

**Individual Identities are Seldom Important**

Identities of people that are revealed in sets may be hidden, anonymous or, even where they are revealed, of relatively little consequence to others in the set. Those who engage in sets are typically more concerned with the subject than the identities of the people that constitute them. One of the characteristics that tends to be indicative of a set mode of interaction in cyberspace is that names of participants, if available at all, are often abbreviated to usernames, without the associated translations into real names or profiles found in networked and group modes of engagement. This does not mean that everyone in a set is unknown: sets overlap with networks and groups. We may participate in a set with people we recognize and people we do not know, and we may come to know people by their consistent pseudonyms. However, most of the time, the identity of the individual, even when known, is not the most important factor when engaging on a set-oriented social system.

Sets are seldom bound by temporal constraints, nor do they demand the use of particular tools or technologies, though both can be important in certain contexts: without the means to discover things, it would be hard to put them into sets. In the broadest sense, sets are found within networks and groups. Indeed, groups and nets can always be viewed as sets, and subsets of sets. A group is a set of people who are members of the group, and a network is a set of people who are in some way connected with one another through direct or indirect links. Similarly, one may find sets of groups, sets of nets and, of course, sets of sets.

**Sets are not Technologies**

At their simplest, sets are simply assemblages of people with shared attributes. They have borders that are defined by the categories that make them, but while the process of categorization might be considered vaguely technological, this stretches the definition of “technology” further than we would like. There are therefore no innate technologies that are required to engage as a set. Having said that, there are many ways that technologies can play a role in establishing, forming, and facilitating a set, beyond simply providing a real or virtual place where people with shared attributes may congregate. Tools like search engines, tagging systems, databases, and classification tools sometimes play a key role in making set modes of
engagement possible in the first place. Such tools often take the place or augment the capacity of a human to organize and classify people and things.

Why Distinguish Sets from Nets?
The reason for distinguishing the set is twofold. In the first place, the ways in which we interact are different when the attribute(s) forming the set matters, rather than the people with whom we engage or the mission of the group. In the second place, the operations we can perform on sets are quite different from those that we perform on networks and groups, a factor of great significance when we come to talk of collectives. Many collectives are the result of set-based aggregations and transformations.

THE BENEFITS OF ANONYMITY

In some cases, the lack of an easy way to identify an individual who is learning in a set may be beneficial, especially when dealing with sensitive topics that require him or her to reveal things that may be uncomfortable or embarrassing. This may be due to the nature of the topic under discussion. For example, many medical sites, counselling sites, and sites relating to socially difficult things that people do not always want to reveal to their networks or groups take on the set social form. This is even true where the site appears to use the same tools and processes as a network or group site, simply because of the extensive use of more anonymous identities. In other cases, the value of anonymity in the set lies in selective disclosure. Self-determination theory suggests that there are three pre-conditions for intrinsic motivation in a learning task: feeling in control, feeling competent, and feeling relatedness with others (Deci & Ryan, 1985). If people are concerned about their level of competence, then fear of negative reactions from peers and teachers may reduce their inclination to share, leading to a vicious circle of doubt that undermines confidence, contribution, and motivation. In a group setting, one of the roles of a teacher is to reduce that sense of doubt, to offer encouragement and positive reinforcement to build confidence.

In a network, that safety net is often lost, because things released into the network may be seen beyond their original context. The products of learning are usually safe to reveal, but the process may be less so. Where anonymity is allowed, fear of disclosure will be lower. However, this is a double-edged sword, and there is a fine balance between the gains and losses that will vary according to context. Anonymity also reduces the significance of social capital (Nemoto et al., 2011) and
the benefits of knowing one’s peers, as well as feeling pride in a job well done that is recognized by a peer group, thereby reducing motivation on the axis of relatedness. If contributions are truly anonymous (as they are in, say, an anonymous Wikipedia page edit), rather than simply anonymized (as they are when pseudonyms are used on a question-and-answer site) then there are no opportunities to gain social capital by merging set-based interactions into net-based interactions.

IDENTITY AND THE SET: TRIBAL UNDERPINNINGS

While in many cases, membership in a set may have no significant impact on an individual and there are many ways to be a member of a set without even being aware of it, there are also many forms of set membership that are central to a person’s identity. Race, gender, nationality, (dis)ability, sports team supported, fashion preference, profession, religion, and so on are crucial to a person’s sense of being in the world and, much like a group (and unlike in a net), those who self-identify with a set may identify people outside it as “other.” On some occasions, such identity is of little or no consequence: for instance, we may feel a distant kind of camaraderie that makes us wave or honk our horns when we see someone else driving the same kind of car or riding the same kind of bicycle. On other occasions, identification with a set means much more. The starting point for understanding this lies, obtusely, in the realm of groups and group dynamics.

E.O. Wilson (2012) suggested that group evolution has played a large role in our development as a species, and thus we depend on identifying with sets of others, or tribes we belong to. For Wilson, the dual driving forces that form us—individual survival and things done for the good of the group—determine our ethics and social being. The sociality of our species places emphasis on survival as a characteristic of the tribe, band, or larger group rather than the individual. In modern societies, this evolved aspect of our being has become more complex because we do not see ourselves as part of a single set but, typically, of many. Cross-cutting cleavages, diverse sets that intersect across many axes (S. E. Page, 2011), mean that we may feel a sense of identity with more than one set of people—a football team, a nation, a set of people with particular abilities or disabilities, and so on. A heavy metal fan who sees another person wearing a t-shirt advertising their favourite band may treat them as a member of the same “tribe,” making assumptions about other shared attributes that relate to lifestyle, preferences, and behaviours, though those people may also be supporters of hockey teams, believers in a particular religion, or other sets that are also meaningful to their identity.
and create feelings of allegiance. Likewise, those wearing religious symbols such as crosses, turbans, veils, or beads may signify not just membership in a set of religious iconography–wearers but also a complete ethical, social, aesthetic, cultural, ontological, and epistemological outlook, as well as being parts of other sets. Of course, religious tribes do not simply relate to identity but often drift into group-modes of social organization, with hierarchies, prescribed behaviours, and rules of membership—the borders are blurred and variable.

Tribes are equally prominent in academia (Becher & Trowler, 2001): people are self-categorized by and identify with others sharing subject areas, uses of methodologies, schools of thought, interests in particular topics, past membership of institutions, classes of qualification, and many more attributes. Some sets are viewed by their members as mutually exclusive despite cross-cutting cleavages such as shared membership of institutions or professional bodies. What to an outsider may seem like remarkably similar things can be the cause of tribal divisions, to the extent that different languages evolve around them. For example, those who make use of activity theory are typically looking at the same things and using the same words with very similar purposes to those who employ actor network theory, yet seldom do the two tribes meet, and if they do, there are many ways they misunderstand one another. The mutually exclusive sets we belong to, though intersecting with other sets that cross those borders can lead to conflict, creative or otherwise. If they were completely isolated from one another then it would be of no consequence, but the cross-cutting cleavages bring them into juxtaposition.

Tribal sets, which involve many different attributes and a sense of membership, are potentially powerful social forms for organizing, motivating, and coordinating activities of members. Membership in a tribe can help create social confidence: knowing that others in a set share common beliefs or attributes can help to reduce the fear of the unknown that may beset those engaging with an unknown community. Conversely, they carry many associated risks when compared to sets that relate to a single attribute. That strong sense of identification can lead to heightened emotions when those who disagree are involved, especially thanks to the naturally anonymous or impersonal modes of engagement that tend to be found in sets. For instance, challenges to religious or political beliefs, criticisms of bands, sports teams or even tastes for certain cellphones can lead to harmful and bitter flame wars. This is one occasion where a transition from set to more interactive nets or group modes is not always helpful.
The social form of the set resembles that of the net in many ways, but without the social constraints where actions of others can strongly affect learning. Sets offer the greatest freedom of choice of any of the forms (Figure 6.1), though it is important to note that this does not necessarily equate to greater control, because too many choices without guidance or the means to make critical decisions is not control at all (Garrison & Baynton, 1987).

**Figure 6.1** Notional cooperative freedoms in sets.

**Place**  
Like all cyberspace learning, there are usually few limits on where a set-based learner can learn. However, there may be some constraints that depend on the attribute chosen to form the set itself, notably where geographical proximity is a significant factor.

**Content**  
There are few limits on content in a set, and most revolve around content: people who are interested in x, people who know about y, people in a place. However, a major issue affecting all set-based learning is that it is not always easy to find
the appropriate classification scheme to define the set in the first place. There are an indefinite number of ways to categorize anything, and it is an active, learned, social behaviour to do so (Lakoff, 1987; S. E. Page, 2008). The learner is often faced with a “chicken or egg” problem of not knowing which classifications relate to what he or she needs to learn, because he or she does not know what classifications are applied within a given domain.

Pace
There are virtually no constraints over pace in set-based learning, save for those that are intrinsic to the nature of a particular set. For example, those with an interest in sunsets may have limited opportunities or interest in gathering at other times of the day, and the set of those who attend a particular event will not exist long before or after the event.

Method
There are virtually no constraints over choice of method in set-based learning. However, it is very much up to the learner to choose the learning methods that are appropriate, and without much control over delegation, the difficulties for learners lie in finding appropriate methods.

Relationship
Sets are typically highly diffuse and impersonal, even though there is total freedom to choose with whom one may interact. Sets are often conduits into the more personal and social forms of engagement of nets and groups, however.

Technology
The main technology constraints in set-based learning are those of compatibility: the set exists in a particular technological environment or a constrained range of environments. There are therefore constraints imposed by the chosen instantiation of any given set: in order for the set to form, it needs a technological substrate to take hold in, and unlike a network, there cannot be alternative channels to a set that are not provided by its aggregator. That said, it is possible for individuals to amalgamate sets from multiple sources, in effect creating a set of sets or, maybe more accurately, a network of sets. Some kinds of set instantiation demand certain types of technology in order for the set to be visible: those that can aggregate, such as tagging or folksonomy systems, or those that can be aware of location, for example. In certain cases, technologies may determine or at least
partially determine the set: owners of iPhones, for example, or those who use a particular app.

**Medium**

Medium form is irrelevant to set-based learning, unless the medium itself defines it, as in a set of writings, videos, songs, or media-constrained attributes such as colour or loudness. A set can, in principle, consist of any number of different media with shared attributes such as subject or theme.

**Time**

Because sets are about attributes of people and things, there are few if any time constraints affecting engagement in a set.

**Delegation**

While it may be possible to find people who are interested in something—a piece of software, a place, an idea, and so on—it is not always easy to sort out the valuable from the peripheral, misleading, or useless. Without even the social capital available in networks to guide people in a set, all content and dialogue is potentially suspect, and lacking other mechanisms, either net- or collective-based, there is no one to whom control can reliably be delegated. Sets provide a lot of choices, but the information required to exercise those choices may be limited.

**Disclosure**

The relative anonymity of sets means that people making use of them are able to retain some measure of anonymity and, on the whole, can be extremely selective about what they disclose and to whom. Having said that, sets only have value insofar as people do disclose knowledge and information, so while personal disclosure is highly controllable, it is necessary for people to reveal information in order for them to function at all.

**Transactional Distance and Control in Sets**

In a set, everyone is equally distant from everyone else in terms of communication, unless it is formed around a teaching presence: for instance, a Khan Academy tutorial creates a very high transactional distance between the tutorial creator and the learner who is using it, though this can be reduced if the creator of the tutorial engages in activities designed to feign a type of interaction by, for example,
asking questions of oneself as if they originated from a live student or engaging in asynchronous discussions around the video tutorial. In such a case, that particular interaction drifts firmly into the networked social form with known individuals, albeit held together by weak and transitory ties in dialogue with one another. Within the set itself—that is, the people who are the discussants in the tutorial—transactional control, in the sense of the learner’s ability to choose what to do next, is absolute: a set is defined by intentional engagement around a topic. While there may be some dependencies on whether or not a reaction is forthcoming when a problem or concern is posted to a set, sets are decided upon and identified by the learner, who is free to seek people with shared interests. There is neither the overt or implicit coercion of the group, nor the social coercion of the network.

Dialogue is, in most senses, freely possible and strongly encouraged, and therefore the *communication* aspect of transactional distance between learners in the set is very low, though it can vary considerably in intensity and volume and, like in the net, become a distributed aggregate value. For example, an online forum or bulletin board makes the process of exchanging messages very straightforward and largely unconstrained. However, the *psychological* gulf between one learner and another is typically very high, because those in the set may neither know nor care much about one another. While caring can be an important attribute in both group and net social forms, in a set the person as a distinct human individual seldom matters at a personal level. If they are visible at all, people often become ciphers, anonymous or near-anonymous agents with which to interact. Most importantly, the great number of choices available to set users does not always equate to control. Whether sufficient help is given with making choices or not depends on the nature of the others in the set, the topic, the degree of familiarity that the learner has with it, and many other factors. Transactional *control* may therefore not be as great as the number of choices suggest. Transactional distance in the set is a complex phenomenon that, as in the net, is difficult to pin down.

**LEARNING IN SETS**

**Sets and Focused Problem-Solving**

Sets are most useful to learners who are fairly sure of what they wish to know or at least the broad area of interest. Much set-based learning occurs “just in time,” concerned with finding out something of value to the learner now, rather than a continuing path. For instance, we may visit Wikipedia, a Q&A site, or Twitter in
order to discover an answer from the set of people who have posted on this topic to a question or perhaps establish a starting point for further investigation.

**Sets and Focused Discovery**

Another common use of sets is to maintain knowledge and currency in a topic or area of interest. For instance, we may subscribe to a feed on a site such as Reddit or Slashdot in order to get a sense of the buzz around a certain topic. The majority of people who use such sites are not actively engaged with the network, but visit or subscribe to them because of an interest in the areas that they discuss. Because such sites are socially enabled, we may contribute ideas, pose problems, seek clarification, and use the other contributors to construct our knowledge, thus helping us to become experts within a subject area, not just to find answers to particular questions or suit specific needs.

**Sets and Serendipitous Discovery**

Beyond that, just as we find overlapping networks, we also find overlapping sets. It is a rare set-based interaction that keeps within the precise limits of the topic of interest, because people have many and diverse interests, often revealed through exposure to cross-cutting cleavages. Thus, as we find with networks, sets sometimes provide opportunities for serendipitous discovery beyond the immediate area of interest. This is frequently enhanced through the use of collectives, especially by recommender systems that suggest other articles, posts, or discussions that may be of interest.

Another way that sets can aid serendipitous discovery is when we spot trends or patterns in behaviour. For example, if one were sitting indoors and noticed that everyone outside was using an umbrella, he or she can learn from the set that it is raining. Similar things happen online: an aggregated RSS feed, for instance, might contain multiple versions of a trending story, which might therefore pique one’s interest. We may discover in a set-based conversation subtleties and areas of interest in a subject we were not formerly aware of. There are subtle blurs here, however, between sets, nets, and collectives. Such trends may be spread through social networks as memes, or be generated automatically by aggregators that combine set behaviours and that, consequently, drive the trend.

**Sets and Multiple Perspectives**

The vastness of cyberspace means it is rare to find only one site or page connected to a particular set. Topics are typically represented in different ways in various
places and often present multiple perspectives, points of view, and ontologies, going far beyond the diversity found in nets (where we might see bias due to affiliation and similarity with others to whom we are connected). This has value in many ways. Every learner is different from every other, with different prior knowledge and experience and different preferences for learning, so the presence of multiple perspectives makes it more likely that one or more will fit with cognitive needs.

Perhaps more significantly, multiple perspectives require learners to make judgments, choose between alternative views, or reconcile them. This active process of sense-making is one of the cornerstones of connectivist approaches to learning: differences are embraced and nurtured because the result is a richer connection and more deeply embedded learning. Differences require us to establish our own points of view, and to better know why we hold them. Multiple perspectives also broaden our outlook, enabling us to see connections that a single point of view, such as one we gain from an intentional teacher, may obscure. For example, to one individual, the set of things connected with e-learning may be limited to what can be found on the World Wide Web, whereas to another it covers any computer-enabled learning activity, while for yet another it refers to pedagogies of cyberspace. By combining these perspectives, a learner may find a valuable intersection or broaden his or her outlook and discover other related issues and areas of interest. The flip side of this benefit is that, much as in a net, it is up to the learner to make sense of conflicting views that he or she discovers. This can be a powerful and creative learning opportunity or, if the area is new or complex, may increase confusion and reduce motivation.

Sets can Support Formal Learning
Sets are of value as part of an individual’s self-paced learning journey, even in a formal setting. For example, at Athabasca University, undergraduate students start work on courses at any time and follow their own schedules within a six-month contract period. They seldom know other students in their course, and though the course itself is highly structured and led by tutors and teachers, the social form for student–student interaction is far more akin to a set than that of a group. There are few social interactions, no process–driven group engagement, few social norms, and few (if any) rules of engagement with other course members. They are not a cohort. They are just a collection of people bound together by the attribute of working over the same period on the same course. While students are not directly working with others or at the same time as others, they often benefit from the presence
of others either directly (through contributions to question and answer sites), or through artifacts that others have shared. Course discussion forums provide both a repository of prior questions and answers, and a place to pose and answer such questions, though in our experience we find that set-based learners rarely engage in extended discussions. We should observe that, though very close to sets, these are tribal groups: there are still norms, expectations, and regulations as well as membership exclusions that make them set-like groups rather than pure sets.

**Breadth Versus Depth**

Broad sets are useful when learning is exploratory and the questions themselves may be unknown. A set of students in an Athabasca University course or a subscriber to an RSS feed from a popular gadget review site will be open to a broad number of ideas and content that fall within a range determined by the shared attribute. At the other end of the spectrum, a person in search of an answer to a single question may turn to a social set-oriented site such as Wikipedia for answers that rely on the set’s specificity, or a site that is so broad there is likely to be someone who knows the answer to any question. For a specific problem, the perfect set would be the global set of everyone. However, it is important that the two sets—people with specific problems and people willing and able to give specific answers—intersect, and that they can find each other. Where a site or service is specific and narrow, this is achieved by being in the same virtual location. For a more general purpose site, it is common for experts to classify themselves into sets, and/or for the site itself to be divided by classifications, often hierarchically organized or with a folksonomic, tag-based approach for identifying subsets. Once again, search engines play an important role in filtering out specific subsets of interest.

**Categories of Things**

Sets are defined by shared characteristics. They are communities of homophily. Sometimes they are intentional, and sometimes they are latent in what is shared. For example, as I look out of my window now, I see a set of people who are currently sharing the same general space as me. Most are pedestrians walking by with whom I do not and will never share a connection beyond, at this moment, being in a shared space. However, if some event occurred (perhaps a whale poking its head out of the water) then that attribute of shared space may become significant because it would enable learning to occur. We would probably talk about what we
were seeing and, in the process, learn. Someone might identify the whale, some-
one else might mention previous sightings, and another might say how unusual it
is to see one in these waters. Others, seeing the set of people gathering and sharing
the attribute of staring at the whale, might come and join us, perhaps contributing
to the shared learning moment. For a transient few minutes, we would become
a learning community, ad hoc and fleeting. When the whale leaves, the signifi-
cance of the space recedes. Some may perhaps make connections and become
networked as a result, but as a collection of people learning together, our shared
context would no longer matter. In rare cases, the set may even coalesce into a
group that continues to gather at other times and locations as whale watchers.
Similar processes happen all the time across cyberspace.

We search for answers and solutions based on their attributes such as subject,
keywords, and tags, or explore topics in Wikipedia, brushing against those with
shared interests, knowledge, and learning, and then moving on. Indeed, a set-based
way of learning has been the norm since the invention of writing. As soon as the
volume of available material became impossible for one human to track, we relied
on classification systems to discover books, papers, and reports, and latterly other
forms of media. Writers, especially of non-fiction, have a set of attributes in mind
when writing books: subject, expected level of ability, background, language and
so on define the sets for whom something is written. The same is true for all
media used for learning.

CATEGORIES AND TAXONOMIES

Categories are ways of putting things into sets and are one of our primary means
of sense-making. To a large extent, how we think is determined by how we cate-
gorize the world (Lakoff, 1987). Our categories evolve as we learn. Expertise can
be seen as an increased ability to both ignore attributes that are insignificant and
to subdivide things that, to non-experts, appear to occupy the same categories (S.
E. Page, 2008). Some of the work of a teacher is involved with helping learners
to identify and focus on categories that are significant in a subject or skill being
taught, to see both big patterns and small distinctions. Traditionally, categoriza-
tions of learning content tended to be performed by trained or otherwise know-
ledgeable individuals who would classify books, papers, journals, and media for
easy discovery and organization. The builders of taxonomies created ordered sets
of things, sorting them into easily identified clusters and groupings.
For the most part, taxonomies have a tendency to be hierarchical. It is no accident that ontologies used in the Semantic Web, though capable of taking any network form, are typically hierarchical in nature as they refer to sets, subsets, and further subsets of objects that are relatively easy for both humans and computers to navigate and understand. However, the world is not always so easily categorized. Many sets intersect, and connections are often more in a network structure than a hierarchy. For this reason, faceted approaches to classification, browsing, and navigation have gained much ground in recent years. Faceted classification allows objects, people, or data to be classified in any number of “facets” from which different combinations of set attributes can be selected for various classification purposes. Ranganathan’s facets (2006) have found particular favour in the library community, offering a structured schema that takes full advantage of the intersection of multiple sets to find things we seek. Although it can cause difficulties when allocating objects in a physically ordered space such as library shelves, a faceted classification scheme lends itself well to computer-based organization. Perhaps more significantly from a learner perspective, facets provide ways of seeing the same things differently. By breaking out of a networked or hierarchical model of thinking, facets encourage a set-based view of the world where multiple orientations can be explored. If experts define such facets, then they offer a means of seeing the world from the perspective of different experts. However, when defined by a diverse crowd, facets may actually offer greater value.

S. E. Page (2008) argues, using fundamental logic and empirical data, that a random set of people will frequently provide better problem-solving in aggregate than a set of experts because of the greater diversity of perspectives, heuristics, interpretations, and predictive models they share. For Page, interpretations equate loosely to categorizations—they are ways of dividing up the world by lumping things together. Combined with predictive models, they provide a means of describing the world and, more significantly, taking effective actions. On the social web, interpretations are reified in the form of tags, metadata supplied by creators and users of content that help others to interpret and discover sets. In combination, the aggregate of such tagging is known as a folksonomy (Vander Wal, 2007).

FOLKSONOMIES

The growth of social media has concurrently seen the growth of a bottom-up method of faceted classification in the form of social tagging, whereby any resource (bookmarks, photos, videos, blogs, and so on) is tagged by one or more
individuals. A machine to enable discovery of similarly tagged resources that others can find aggregates their classifications. These folksonomies define sets of things with shared attributes most commonly known as tags, and they can be used to guide a learning journey. Because of the diversity of interpretations of the world that such tags represent, they are a powerful way for learners to identify and explore both the vocabulary associated with a given subject area and the different ways that the area is conceptualized. Anticipating our discussion of the power of the collective in the next chapter, when combined in a weighted list such as a tag cloud where tags that are more frequently used are shown with greater weight through visual cues such as size, font, or colour, they can indicate not just the range of interpretations of the world that the crowd uses but also the relative importance of such interpretations in aggregate. Kevin Kelly has identified tags and the hyperlink as the two most important inventions of the last 50 years (2007, p. 75).

There are many set-oriented uses of tags in which learners help others to learn. Twitter hashtags help us to find discussions, snippets of knowledge, and hyperlinks to further resources from which we may learn. Flickr Commons (http://flickr.com/commons/) is an exercise in mass tagging, involving tens of thousands of people categorizing public domain photos for the benefit of themselves and others, allowing users to easily find relevant photos in huge collections. The cataloguing and discovery of images is a wickedly complex problem, because even the simplest of holiday snaps can be categorized in an indefinite number of ways (Enser, 2008). The social tagging in Flickr Commons is a great example of how a large, anonymous set of people can create value for others without any kind of social interaction. Some photos in the public domain collection have been tagged thousands of times, with tags identifying people, places, objects, themes, subjects, concepts, colours, and hundreds of other attributes that may be used to split objects into sets. Bookmark sharing sites such as Delicious, Furl, and Diigo are heavily dependent on tags that people provide to categorize websites of interest according to topic.

As well as enabling the set to help its members make sense of the world interpreted by others, the act of tagging itself is a metacognitive tool that encourages the tagger to think about the things that matter to him or her, helping the process of sense-making, embedding reflection in the process of creation, and thus enhancing learning (Argyris & Schön, 1974). This process may be aided by systems that suggest additional tags, previously applied by others, similar to tags first chosen, which helps to decrease a potential multiplicity of synonyms from becoming tags,
but also limits variability with both positive and negative results. We will return to other downsides of tagging later in this chapter.

**Tools for Sets**

There are many tools available that offer and enhance set-like modes of learning. Typically, most set-oriented applications are not exclusively dedicated to the set, also providing tools to branch into networks and, in some cases, groups. We describe a few of the main examples of the genre below in order to provide a sense of the range of tools and systems that can be used in set-oriented learning.

**Listservs, Usenet News, Open Forums, and Mailing Lists**

For decades before the invention of the World Wide Web, people engaged in posting on bulletin boards, anonymous FTP servers, newsgroups, and other topic-oriented services with great enthusiasm. Though many of these developed into rich networked and group communities, with emergent or imposed hierarchies and complex economies driven by social capital, several others celebrated open engagement around subjects and themes without significant social ties. Such services are still very common today in the form of social interest sites—Pinterest, Wikia, and learn.ist being prime examples—sites dedicated to different kinds of software and hardware, and many more.

**Socially-augmented Publications**

It is rare to find any form of publication in the wild that does not allow some level of anonymous user interaction—newspapers, magazines, public blogs, and the like, all offer engagement at a public level, frequently anonymous or where the identity of the person making comments is irrelevant, concealed, or ambiguous. There is a fine dividing line between the anonymous set orientation of these and the networked mode of engagement, and many combine the two. Sometimes, networks are explicit in trackbacks, where one blog comment leads to a different blog site, or through engagement in a conversation by known individuals. Much of the time, the comments are from people that no one else in the dialogue knows, nor wishes to know.

**Tags, Categories, and Tag Clouds**

Folksonomic classification, where bottom-up processes are used to tag content, are archetypally set-oriented. When using tags to find content, our concern is
not with the individuals who create them but with the topics that they refer to. Hashtags in Twitter, tags in Delicious, Flickr, and many other systems provide a set-oriented way of cooperative resource discovery. Sometimes, sites will use a combination of top-down categories and bottom-up folksonomies. For instance, Slashdot, Reddit, Digg, and StackOverflow provide ranges of common topic areas around which posts occur.

Search Terms
When we enter a search term into a search engine, we are typically seeking a set of things that share the attributes of the keywords or phrases we enter. What we get back, if all has gone well, is a list of items where others have used those terms. Thus, the search engine mediates between creator and seeker, enabling a simple form of one-to-one dialogue between them. However, the intentions of the creator may be very far removed from the intentions of the seeker, even when he or she is skilled in the art of searching. Unfortunately, as we have already observed, expertise is in part a result of being able to use categories effectively and a learner will be unlikely to know which terms are most appropriate to his or her needs in a novel field of interest. The sets returned, in such cases, may be highly tangential and confusing. For example, if a learner enters a search for “evolution” with the intention of learning more about the theory, then the list of results are likely to include many ideologically driven creationist sites (often deliberately manipulated through search-engine optimization to appear on the list), sites using the word in the pre-Darwinian sense (like the evolution of a design or concept), a film by Charlie Kaufman, a number of beauty products, and plenty more results of little value. Like the tag, the search term is highly susceptible to various forms of ambiguity. Unlike most tagging systems, search terms may be refined. A search for “Darwin’s theory of evolution” will result in a more focused set of results, but again, the anonymity of the set will mean that the learner is in conversation with not only evolutionary theorists and historians but also creationists. Bearing in mind that our hypothetical learner knows little or nothing about evolution, this places him or her in great danger. Without a theoretical framework to understand the manifold weaknesses and failings of the creationist point of view, he or she may learn inaccurate ideas that will make understanding the correct theory more difficult. Complexity theorists might view the potential range of useful and less useful results as a rugged landscape: there are many possible solutions or “peaks” that may be fit for the purpose, but climbing one (even a low one), will make it significantly harder to move from there to a higher, more useful peak (Kauffman, 1995).
While most search engines follow the logic of the set in an abstract sense, many make use of the set of people more explicitly in algorithms that mine similarities between searchers. Some, such as Google’s use of PageRank, also use networks to help provide relevant results. We shall return to this powerful use of the set in our chapter on collectives.

**Social Interest Sites and Content Curation**

Sites such as Pinterest, Learnin.st, Wikia, Scoop.it, etc., allow people to share collections of related content—in brief, sets. Curated content can be created by individuals, groups, and networks as well as sets of people, and can be directly authored and/or collected from elsewhere, but however it is created, it provides a set of resources that are clustered around a topic of interest. Many more general social sites provide tools for the aggregation of content around a topic or theme: YouTube Channels and Facebook Pages, for example, provide thematically organized content where the set is at least as important as the network or group that is associated with it. Though the genre has been common throughout the history of the social net, going back to (at least) Usenet News and bulletin boards, in recent years there has been a significant growth in social curation sites, not to mention sustained growth in older social bookmarking sites like Delicious, Diigo, and Furl, sharing options for personal curation tools like Evernote or Pocket (formerly ReadItLater), and ways of using more general-purpose tools like Facebook Pages or Google Sites to assemble and share information on a topic. Curated sites or areas of sites are concerned with niches—areas of interest that are often very narrow—for instance, food (e.g., Foodspotting.com) or fitness (e.g., Fitocracy.com). While most niche sites can be used by groups and often involve nets, publicly available niche sites based around topics are deeply set-based in nature.

The vast majority of niche sites make extensive use of folksonomies for organization, often combined with a more top-down and hierarchical categorization system. From a learning perspective, curated sites combine many of the advantages of a traditional, teacher-created content-based behaviourist-cognitivist learning resource with the added value of sets, and optionally, nets and groups. Social curation sites, as the name implies, embed the ability to tag, rate, discuss, and comment. Not only that, most curated content can be re-curated, mashed up, and aggregated, extending the value by recontextualizing it for different communities and needs. Thus, different kinds of conversation can develop around the same content, new connections can be made between different topic areas, and the value of diverse perspectives and interpretations can be heavily exploited.
Shared Media

Many rich media sites share tutorials and exemplars, some user-generated, some more top-down but with associated discussion or comment options. YouTube, TeacherTube, The Khan Academy, Flickr, Instructables, and many other sites offer rich learning content around which set-oriented discussions and learning can evolve. Media act as anchors for learning a particular topic. Wikis are flagship set-based tools. Wikipedia, Mediawiki Commons, Wiki Educator, and a host of other reference and sharing sites are based around categorized content. While many wikis do support sets and networks, the primary engagement in a wiki is nearly always focused around content rather than social interaction.

Arguably the poster child for set-based learning, Wikipedia is without a doubt the most consulted encyclopedia ever written, and one of the top two tools for learning on the Internet today, the other being Google Search. If ever anyone expresses doubt that online learning has a future, we have only to ask him or her to what they turn to first when seeking to learn something new. In many cases, the answer is “Wikipedia” or “Google Search.” Wikipedia organization is complex and highly social, yet it has few identifiable groups and very little in the way of networks. The vast majority of interaction is indirect, mediated through edits to pages by a largely anonymous or unknown crowd; most editing or visiting a page because they are interested in the topic it describes. In other words, they are part of a set with the shared attribute of interest in a topic.

With a similarly vast number of users, YouTube is another set-based system that is extremely popular for a wide range of uses, many educational in nature. Social networking in YouTube is not its main feature, and much of the interaction that occurs is centred on specific videos or clusters of videos (collections) rather than people known to one another. While the number of educational videos on YouTube greatly outnumbers those found on any other site, including Facebook, other similar sites like TeacherTube and SchoolTube provide services that are focused specifically on education. The benefit of such sites is their greater focus on formal learning, making it easier for learners to identify reliable and useful resources without the distractions of Lolcats and music videos. They are niche sites that contain further sub-niches or subsets categorized in ways designed to link learners with content and consequent interaction. Thus, the choice of the site itself acts as a means of classifying and organizing learning resources along set lines.
Locative Systems
Places are attributes shared by people who are in the same location. A wide range of social applications have been designed to take advantage of geographical co-location, from restaurant finders (e.g., Yell, Around-me, Google Latitude), to game playing as a means of discovering one’s locale (Geotagging, FourSquare) to cooperative shopping and dining (Groupon). Many mobile apps make use of location information to both discover and post information relating to the locale: FourSquare, Google Latitude, Geotagging, and many more tools allow persistent interactions to occur around a place. Locations thus become augmented by the activities of people who inhabit them, with the location serving as the defining attribute of the set of people who visit geographical spaces.

Augmented Reality
2D bar codes such as Semacode, QR codes, and similar technologies enable physical objects to be tagged. These bar codes are used for advertising, allowing people to snap photos of codes using cellphones or similar devices and receive either small snippets of information, or more commonly, hyperlinks to websites providing further information. While these have some potentially valuable educational applications, they are not usually socially enabled. However, a particularly promising approach to learning as a set in a location is to provide virtual information via cellphone, tablet, or more sophisticated devices such as Google Glass, and to allow people to leave virtual cairns or tags that others may discover in the space if equipped with a suitable device.

Crowdsourcing
A particularly powerful use of sets in learning is found in question-and-answer sites and other approaches to crowdsourcing work, problem-solving, and creative construction. From simple Q&A sites such as Quora to more complex brokerages for skills and services, the crowdsourced solution to learning problems is popular and thriving. Again, many of these sites shift between network and set modes, sometimes intentionally, sometimes seamlessly. For example, Amazon’s Mechanical Turk or Innocentive both provide a mediating role between those with problems and those able to provide solutions, typically using set-based characteristics to match the two, and facilitate the exchange of money between the parties. Other systems, such as Yahoo Answers and Quora, are less obviously incentive-driven: while social capital often plays a role, in which case interactions drift toward network-based models, many people contribute answers because
they can. Altruism is a deep-seated human characteristic that has evolved in our species: one need look no further than the fact that people frequently risk their own lives to save those of strangers to see this fundamental urge in action (E. O. Wilson, 2012).

One of the most obvious ways to exploit the wisdom of crowds is to ask a question. Assuming the question is meaningful and has a correct answer, there is likely to be someone somewhere in cyberspace who knows it. Two giants of networking have tackled this opportunity in quite different ways.

Yahoo Answers is one of the older user-generated answer sites. Modelled after the wildly successful Korean site Naver Knowledge iN (www.naver.com), Yahoo Answers allows users to post and answer questions with no fees or concrete rewards. Questions and their responses are categorized and lightly filtered to remove obnoxious or nonsensical material. Users provide answers, and the questioner decides or allows the crowd to select the best one. Obviously, the site provides some value to users who can search or browse the archives for answers to relevant questions. Like all social sites, Answers gains value in proportion to the number of users. To support and encourage participation, Yahoo offers “points” for contribution. Five months after its launch in December 2005, Yahoo Answers was publishing nearly a half million questions per month, which generated nearly 4 million answers, an average of 8.25 answers per question (Gyongyi, Pedersen, Koutrika, & Garcia-Molina, 2008).

As in many publicly available sites, Yahoo Answers contains a great deal of “noise,” or questions and responses that can charitably be classified as silly or inane. Interestingly, many of the questions seem to be posted to stimulate discussion as much as to obtain a definitive answer. A question posed by the user Gothic Girl illustrates both noise and a discussion stimulator: “What is your favorite food?? (it can be candy too, i say that’s food)” received 41 answers! Alternatively, a question by Katie R. in the Math section, “If I calculate the variance of a collection of data to be .235214, does this tell me that there is large variance (that the data is spread out) or that there is relatively little variance?” received a comprehensive answer with examples from a top contributor whose profile explains “by education and profession, I am a statistician.”

Rival answer sites such as Answerbag.com and Quora, a more network-oriented Q&A site, are developing rules and practices that attempt to better organize questions and answers and support the development of communities among their members. For example, they allow members to develop searchable profiles and engage in discussion via comments to either questions or comments. Google
took a more traditional approach for Google Answers, a more commercially oriented service, allowing users to post bounties between $2 and $200 for solutions. Rafaeli, Raban, and Ravid (2007) analyzed all questions and answers submitted between 2002 and 2004, and found that over half of the 78,000 questions asked were successfully answered with an average payout of $20.10. After four years of operation, Google discontinued accepting questions and answers, and described the project as an interesting experiment. Its failure in the face of Yahoo’s continuing success has raised an interesting debate in the blogosphere. It seems that many want to ask questions, a few want to answer, but few want to pay and even fewer want to handle the logistics of accounting, curtailing spam, and all the other issues that challenge Web ventures. This also speaks to the dangers of extrinsic motivation reducing the motivation to answer (Kohn, 1999). It is a very notable feature of most surviving Q&A sites that the rewards are intrinsic, and often provided for completely altruistic reasons, with no hope of even social capital being accrued.

In recent years, StackOverload sites have become extremely popular because they offer not only set-based interaction but also a collective-based method of identifying useful answers, organized by those perceived as being the most accurate or beneficial.

The use of answer sites creates an additional option for teachers and learners that provides a more current social resource than more traditional web or print sources. This query of the crowd is however less definitive and reliable than more traditional reference resources including those such as Wikipedia, which garner much more critical and comprehensive review by peers for accuracy, connectiveness, relevance, and authority. Some learners use answer services merely as a means to lighten their workload, and as a consequence, likely diminish their learning by posting homework questions in search of “easy answers.” Not surprisingly, this abuse of the crowd has given rise to the dyoh (Do Your Own Homework) movement.

Nonetheless, question and answer sites may prove useful for topical questions where discussion of especially socially constructed issues among answerers may be a forum to generate knowledge not available in more traditional resources. A review of the popular sites also reveals examples of explicit content that would be offensive and inappropriate for many learners.

TeachthePeople.com is another startup site that provides “experts” with server space to which they can upload teaching and learning materials in many formats, into “learning communities.” The site shares ad revenues with “teachers” that are dependent upon the number of learners who access the site.
Crowdfunding

Increasingly, learners are funding their learning with the aid of the crowd. Crowdfunding sites for students such as Upstart (www.upstart.com) or Scolaris (www.scolaris.ca) match sets of people interested in funding learners with donors. While many still rely on group forms for this role (governments, families, companies, and so on), the set has proven to be surprisingly effective for connecting those in need with those who wish to give. Because such applications tend to be one-off requests, networks have little or nothing to add, save in helping to verify identity and, occasionally, allowing prospective funders to find out more about students seeking funds.

Risks of Set-based Learning

Reliability

The relative anonymity of sets makes it significantly harder to gain a strong sense of the reliability of content produced by the crowd than it does in groups and networks. The Internet is notoriously filled with distortions, lies, and falsehoods of many kinds, but even when data is accurate and meaningful, it does not mean that it will be of great value to a particular learner at a particular point in his or her learning trajectory. The problem is made worse by the fact that, sometimes, people deliberately mislead or distort the truth.

In the absence of cues such as the presence of advertising, an excess of exclamation marks, or a lack of references, there are three distinct ways that reliability of knowledge gained through sets can be ascertained inherent in the social form. The first is correlation: if more than one similar answer to a problem can be found in a set, then it increases the probability that the answer is reliable. The nature of sets, however, makes this a risky approach, because people in sets influence one another and it is very common for falsehoods to be propagated through and across them, each wrong solution reinforcing those that come before. The second is disagreement: where multiple perspectives and solutions are presented, this typically leads to argument, and by analyzing the strengths and weaknesses of the arguments, the learner can come to a more informed opinion about the correct solution. Disagreement is usually a good thing for learners in sets, because it encourages reflection on the issues and concepts involved, enabling learners to form a more cohesive view of a topic. Third, beyond the inherent capabilities of the social form, other social forms can play an important role in establishing veracity: we may, for instance, trust opinions voiced in our networks, turn to a group for discussion, or
as we shall see, make use of the collective to establish reputation or reliability of information provided in the set.

Anonymity

On the whole, the relative anonymity of the set has notable benefits to the learner. There can be greater openness and keenness to participate, especially when topics involve sensitive personal disclosure. Where the crowd is contributing to, editing, and evolving a resource started by others (e.g., a Wikipedia article) the anonymity makes it far easier to make edits because editors are unlikely to feel as beholden to earlier authors as they would in a group or network. When using wikis in a group, we have found that the strong ties, roles, social capital, and the politeness that this leads to can significantly deter members from editing what others have laboured to produce. This may be a particularly strong tendency in the authors’ two native countries, Canada and the UK, both known for cultures of politeness, but it seems likely that the more learners know one another, the less inclined they will be to modify one another’s work in the peculiarly mediated world of the wiki, at least without extensive use of associated discussion pages or other dialogue options. However, the flip side of relative anonymity is that it makes it more likely for people to be treated impersonally, as ciphers, with feelings that can be ignored or, as we see in the case of Internet trolls, manipulated for fun. From the early days of Usenet News and bulletin boards, we have seen large anonymous communities brought down by flame wars and trolling.

Another drawback of anonymity is that the motivation to participate is significantly lower than in groups or networks. If individuals are not recognized and identifiable, there is sometimes less social capital to be gained, and there is no sense of being beholden to other individuals, either because they are known directly to us or because of the written or unwritten rules of a group. Size can play an important role in overcoming this limitation. Where many people are engaged, such as might be found on a large social site like Twitter or Wikipedia, there are more likely to be others willing to share and participate at any given time. The Long Tail (C. Anderson, 2004) means that someone, somewhere, is likely to share the same concerns, no matter how minor the interest.

The Trouble with Tags

Tags are a useful way to harness the collective wisdom of the crowd, and we will return to more advanced ways that they can be used in the next chapter.
on collectives. However, folksonomies suffer from a range of related issues and concerns.

**Context and Ambiguity**

Especially when learning, the meaning of tags may be closely connected with the context of use. The same word in a different context can mean something different, even though the dictionary definition is the same. For example, if an expert tags something as “simple,” it means something quite different than if the same term was used by a beginner. Equally, “black” might designate a colour, a race, or a kind of humour, among many other things. “#YEG” is a hashtag commonly used by residents in Edmonton to refer in Twitter posts to the city, yet it also is the designation for the Edmonton International Airport. The word “chemistry” used about an image might refer to the subject of chemistry, or equally to the bond between two lovers in a different context. In some cases, the same word may have multiple distinct meanings in a dictionary. Context is also important when dealing with lexical and syntactic ambiguities where longer descriptions are applied. For example, “Outside of a dog, a book is a man’s best friend; inside, it’s too hard to read” (attributed to Groucho Marx (van Gelderen, 2010, p. 42)) or “they passed the port at midnight.”

Bruza and Song (2000) describe a diverse set of categories that might become tags: S-about (subjective-about, broadly scalar qualities), O-about (objective-about, broad binary classifications), and R-about (contextualized to a group of users). R-about is particularly interesting, as it suggests that different communities may use the same terms differently. This is confirmed by Michlmayr, Graf, Siberski, and Nejdl (2005), who looked at the properties of tags describing bookmarked sites on the Web obtained from Delicious. They postulated that those who bookmarked similar sites and described them with similar tags would share other tags, interests, and perhaps, already belong to, or be interested in developing, existing networks or groups. They found, however, that users who tagged similar sites did not have large intersections of other resources that they tagged. An average of 84% of sites bookmarked by users who share a common site were not bookmarked by other users sharing a common bookmark. Furthermore, they found surprisingly little correlation between folksonomic tags and those developed as a component of the more formal tagging systems developed by the Open Directory Project (www.dmoz.org). This suggests that folksonomic classification may serve personal and perhaps group needs, but beyond showing popularity and tag cloud images, the
extent to which inferences can be drawn based on folksonomic tags or the taggers is limited without further examination of context.

**Homonymy**

Sometimes, especially in English, the same word means more than one thing. These are subcategorized as homographs, heteronyms, and homophones. Homographs are spelled the same but with different meanings: for instance, bat (an animal) and bat (a stick for hitting balls). When the pronunciation is different, they are usually referred to as heteronyms: for instance, “bow” (a ribbon tied in your hair) and “bow” (to lower your head). Equally, homonyms may be homophones (sounding the same but spelled differently), for instance “through” and “threw.”

**Synonymy**

Even where terms are distinct, more than one term may be used to tag the same thing. Some are obvious: for instance, “people,” “persons,” and “person” refer to very similar resources. Stemming dictionaries and tools like WordNet can deal effectively with such simple cases. In other cases, the words have quite distinct and precise meanings that are not synonymous, but will typically be used to describe the same object: for example, e-learning, online learning, and networked learning, at least for some, refer to the same set of objects. This can be a particular problem when using metonyms—for instance, “Hollywood” to refer to the US film industry and the place where it is most concentrated—where the term is not only a synonym but also is ambiguous.

**Binary versus Scalar Tags**

Nearly all tag-based systems treat tags as simple binary classifications which, in some instances, are what is needed. However, many tags are fuzzy and constitute fuzzy sets (Kosko, 1994): something may be fun or less fun, red or more red, cute or less cute (Dron, 2008). Golder and Huberman (2006) list seven distinct varieties of tag: identifying what (or who) a resource refers to, identifying what it is, identifying who owns it, refining categories, identifying qualities or characteristics, self-reference, and task organizing. Very few systems, notably those created by author Dron, make use of fuzzy tags that allow degrees of membership in a set (Dron, 2008; Dron, Mitchell, Boyne, & Siviter, 2000). We hope to see more such systems appearing in future, but they are beset by the inevitable complications of entering and using fuzzy tags. Binary tags take little effort to create, and are typically a comma-separated list of words. Fuzzy tags require not only the tag
but also its perceived value to be entered, and raise further issues as to how they are presented and aggregated—for instance, should the values be simply averaged, or should there be some form of weighting based on number of uses too? Such problems also beset simple rating systems on, things like review sites, and the solutions are similarly imperfect: showing numbers of ratings separately, for example.

Lack of Correlation
These and other related concerns matter considerably when learning in sets, because a learner may find it harder than an expert to distinguish context and ambiguity, not be aware of relevant synonyms, or fail to observe closely related but distinct homonyms. While it can be argued that the process of discovering such uncertainties is an effective way to become adept in a given subject area, this may equally reduce motivation and increase the time needed to learn something new.

SETS IN THE ONLINE CLASSROOM
Within a formal, group-based educational setting where cohorts of students work in lock-step with one another on shared activities, set-based tools and communities can provide great augmentative value.

While traditionalists throw up their hands in horror at the problems that emerge from students using Wikipedia in traditional courses, citing concerns about reliability, superficiality, and plagiarism, the online encyclopedia has a place in almost any learning transaction. It is a wonderful way to enter into a topic, providing not only a fairly reliable overview (especially in academic topics) but also links, references, and further reading that can greatly assist the exploration of a subject area.

Moreover, many teachers have reported success in encouraging students to make active contributions to the site: they create pages, correct errors, and engage in the often rich discussions that emerge around a particular page. However, volunteer Wikipedia experts have also complained about the mess of forked (or unrelated) articles, and poorly written or incomplete edits that some students have left. In true wiki spirit, there is an editable page on Wikipedia (en.wikipedia.org/wiki/Wikipedia:Assignments_for_student_editors) discussing how to make the most effective use of a Wikipedia article as a writing assignment for students.

Similarly, tutorials available through sites such as the Khan Academy, eHow, WikiHow, HowStuffWorks, provide not only useful supplements to classroom learning but also a chance to engage with others, to see how they conceptualize
and mis-conceptualize subjects and topics, and gain a sense of their own knowledge in relation to others. Within a formal setting, the widespread availability of varying quality resources that can take the place of some of the traditional roles of a teacher makes it possible to “flip” the classroom (Strayer, 2007), a term that describes what many teachers have always done: leave content for self-guided homework and concentrate on richer learning activities in the classroom. Content discovery and activities that in more traditional settings form the material of the learning process, whether online or not, can be offloaded to the set, allowing the teacher to concentrate on social knowledge construction processes that are more appropriate to a grouped mode of learning.

### Teaching Set Use

We have noted that one of the major problems with set modes of interaction, as well as one of the greatest opportunities, is anonymity. This means that it is vital for users of sets to develop well-honed skills in identifying quality, relevance, and reliability of both people and resources. Teachers in conventional courses can play an important role here, modelling good practice, providing feedback, recommending strategies, and offering opportunities for safe practice.

Self-referentially, the set itself can provide resources and clues about the reliability of information found within it, particularly if it incorporates collective tools that emphasize reputation, provide ratings, or show other visualizations that give hints about the value of a contribution or individual. Even where that is not the case, it is often possible to follow conversations and identify which participants hold the upper hand in controversies or disagreements.

One important role for the teacher wishing to make use of sets is to define or identify relevant vocabularies and narrow down the attributes by which sets are classified. This may simply be a question of sharing vocabularies, identifying relevant search terms, and providing exercises that use the appropriate wording. However, the diversity of views and vocabularies that may be discovered also open up many opportunities to explore the ontological assumptions of a subject area, and much can be gained from comparing and contrasting different ways of seeing the world as a result.

The choice of appropriate sets is an important one, and relates to the purpose and context of the learner. A diverse crowd may be useful in solving some problems and less effective in others. Generally, when learning, a set of experts is better than a random set, or one made up of beginners, or things they come up with will be entirely random. But too narrow a focus may mean they will not meet
Learning in Sets

the needs of the learner. Sometimes, proximal development is an issue. A set of subject experts is probably not useful to help learn the basics of a subject because the vocabulary and assumed knowledge of the set may not just render the subject incomprehensible but actually demotivate the learner. For beginners, it is better to find a set of expert teachers, explainers, demonstrators, and co-learners, each of whom has a certain amount of knowledge. The set will represent a range of perspectives and views of the subject, which together will offer diverse opportunities to connect existing knowledge to new discoveries.

Designing and Selecting Set-oriented Applications

There are two main issues that a set-oriented system needs to deal with: publication (or sharing), and discovery (or finding). On the one hand, there needs to be sufficient data organized effectively so that sets can be discovered and formed in the first place. On the other, it should be possible to use tools to find, organize, and make use of them.

Unless a networked application or site is highly focused on a finely differentiated subset, it is almost a defining characteristic for a set-oriented application to have the means of classifying content. The most popular approaches to this are to offer top-down categories or topics, bottom-up tags, or both; some go further in providing RDF-based ontologies or faceted classification schema. Search tools are also vital, in some cases circumventing the need for explicit categorization, though use of metatags, keywords in titles, and other cues still play a strong role in helping the search system to find what you are looking for. A richer search system is often valuable: at its most extreme, this might take the form of a visual query tool that generates SQL or similar commands to extract data from a relational database.

Curation tools are of particular value in set-oriented applications. Users should be provided with the means to collect and assemble content, and to create it. This may be as simple as a wiki—the popular Wikia site, for example, which is making great efforts to be a social networking site and build group-like communities, is a predominantly set-oriented application almost entirely wiki-based. It allows people to create tagged wikis and provide anonymous edits, much like Wikipedia. Other tools, such as learni.st and Pinterest, provide tools for aggregation that allow people to assemble content around particular topics, with a focus on presentation and classification. RSS feeds and other push technologies that provide channels, such as listservs or mobile apps making use of social site APIs, can be very valuable in certain kinds of set-oriented, curated content application, allowing a learner to identify a particular set or subset, which can feed him or her
with a stream of information. This is especially relevant to broad sets that provide rich content around a subject area. Such aggregation may be less important on question and answer sites or similarly narrow-focus social systems, where engagement is unlikely to persist beyond dialogue relating to the presenting problem.

Curation tools gain value if they are able to use common standards such as HTTP and RSS to retrieve content and metadata. Where access to otherwise restricted content is needed, such as from a closed network system, it is also valuable to provide the means to access them through their APIs. For our own Elgg-based site, Athabasca Landing, we created tools to use and provide authenticated RSS feeds, tools for importing feeds into different site media (such as wikis, blogs, and shared bookmarks), and tools to embed Google Gadgets.

Beyond the set, site analytics that monitor usage and hits on various pages or artifacts can also be useful in providing feedback, indices of value, and even fodder for advertising services to a set curator.

Relational databases are ideally suited to set modes of interaction because of their formal basis in set theory. However, looser kinds of database management systems may have greater value for some kinds of set data, especially where either very high performance trumps the need for accurate classification, or classifications are fuzzy, unspecified, or shifting.

Like all other social applications, communication and sharing tools are a prerequisite in set-based systems, with a greater emphasis on sharing than that found in network or group social systems. Because of the sporadic and bursty nature of set interactions, tools to notify people via other systems such as email or SMS are useful.

Verifiable identification of an individual in a set-oriented application is seldom as important as it is in networked and group applications, though profiles that reveal interests, skills, and purposes are very helpful in filtering for useful topics of interest. That said, one of the biggest difficulties when dealing with sets is determining the, accuracy, truthfulness, and trustworthiness of others in the set, so it is helpful to provide a means for allowing people to reveal some kind of persistent identity, even if it is pseudonymous and shifts between one set and another.

Another range of potentially valuable tools for set-oriented applications are those that provide controllable filtering. Given that there may be diverse viewpoints, and that some content may be boring or disagreeable to some members of the set, it is important to allow features such as the blocking of individuals, filtering based on keywords, and tools that enable learners to focus on specific things—again, curation tools are useful, as are personal “dashboards” that enable...
a learner to assemble collections of content and dialogue. It should be noted that filtering is a potentially double-edged sword. Though well-suited to anonymous engagement in a set, in network or group applications it can impose implicit censorship on members and thus play a powerful role in shaping the community and reinforcing its values, creating an echo chamber or filter bubble (Pariser, 2011) that may have harmful and unforeseen effects. Because sets, by definition, do not involve any distinct community, filter bubbles are less problematic, assuming that other sets addressing similar concerns are available for those that find their interests or beliefs are excluded.

Associated with the relative anonymity of their members and perhaps more than in any other social form, sets are frequently intertwined with collectives. It is rare to find a set-oriented application without at least some collective features and/or a large amount of editorial control. Rather than dwell on this in detail here, we will return to it in the next chapter.

CONCLUSION

Sets are a ubiquitous social form we all engage in both on and off the Internet. The characteristic forms of social engagement that emerge in sets in a learning context typically have to do with cooperation rather than collaboration. Set-based learning is about sharing ideas, resources, tools, media, and knowledge, and engaging with others on an ad hoc, transient basis. On many occasions, others will make use of what we have shared without our knowledge or consent: the value of the set therefore grows over time. Once persistent dialogues start to occur, set-based systems blur into net-based systems: one of the most notable uses of sets is as a means for forming networks and, occasionally, groups.

Arguably the greatest value from sets comes when they are the social form behind collectives, and the most effective sets make extensive use of collectives by creating structure and dynamic processes to drive them and capitalize on their features. We turn to collectives in the next chapter.