Opinion

# Categorization in the wild

Robert J. Glushko[1], Paul P. Maglio[2], Teenie Matlock[3] and Lawrence W. Barsalou[4]

[1] School of Information, University of California, Berkeley, CA, 94720, USA
[2] IBM Almaden Research Center, 650 Harry Road, San Jose, CA, 95120-6099, USA
[3] Cognitive Science Program, University of California, Merced, CA, 95344, USA
[4] Department of Psychology, Emory University, Atlanta, GA, 30322, USA

**In studying categorization, cognitive science has focused primarily on cultural categorization, ignoring individual and institutional categorization. Because recent technological developments have made individual and institutional classification systems much more available and powerful, our understanding of the cognitive and social mechanisms that produce these systems is increasingly important. Furthermore, key aspects of categorization that have received little previous attention emerge from considering diverse types of categorization together, such as the social factors that create stability in classification systems, and the interoperability that shared conceptual systems establish between agents. Finally, the profound impact of recent technological developments on classification systems indicates that basic categorization mechanisms are highly adaptive, producing new classification systems as the situations in which they operate change.**

Categorization research focuses on the acquisition and use of categories shared by a culture and associated with language – what we will call 'cultural categorization'. Cultural categories exist for objects, events, settings, mental states, properties, relations and other components of experience (e.g. birds, weddings, parks, serenity, blue and above). Typically, these categories are acquired through normal exposure to caregivers and culture with little explicit instruction. Large literatures across the cognitive sciences – in psychology, neuroscience, linguistics, anthropology, philosophy and artificial intelligence – document major progress in understanding cultural categorization, along with controlled laboratory paradigms that use artificial categories to isolate, model and implement categorization mechanisms [1–6].

While focusing on cultural categorization, cognitive science largely has ignored two other forms of categorization 'in the wild', what we will call 'individual categorization' and 'institutional categorization'. Although individual and institutional categorization have existed for millennia, modern technological developments have made them significantly more available and powerful. As a result, individual and institutional categorization are becoming increasingly important in human activity and deserve scientific study alongside cultural categorization. Studying all three forms together is likely to produce greater understanding of categorization mechanisms.

*Corresponding author:* Barsalou, L.W. (barsalou@emory.edu).

## Individual categorization

Individual categorization occurs when someone creates an idiosyncratic classification system primarily for his or her own use, for example, when creating categories to organize locations where food can be gathered, objects in a garage, CDs in a music collection, websites in the favorites list of a browser, etc. Often one creates an individual classification system with little input from others and doesn't share it.

Prior to modern Web technology, individuals relied on memory and writing to develop individual classification systems. Modern technology, however, makes developing these systems much easier. The Del.icio.us Web service, for example, allows users to store and organize Uniform Resource Locators (URLs) for Web resources and to label these URLs however they please (e.g. labeling URLs for books to read with 'read'). These labels are called 'tags', and URLs tagged in the same way form an individual category. As users assign whatever tags they like to URLs, they create an individual classification system that reflects their interests. Because tags are stored automatically, recording individual categories is much easier than using memory and writing. Because tags can be used to retrieve associated categories, retrieval is easy and powerful (e.g. retrieving URLs tagged with 'read'). As a result of these technological innovations, individual classification systems are becoming increasingly pervasive and important in everyday life. Box 1 describes another popular tagging application, Flickr, and illustrates the social dimension that some individual classification systems take.

A rapidly growing research literature primarily addresses statistical properties and behavioral demographics of tagging practice. As the number of tagged objects increases, the overall set of tags used also increases; however, some tags are used far more than others, and the number of tags applied to a given entity tends to be small [7]. Users differ in their proclivity to tag, tags differ in their use over time, and tagged objects differ in how they are tagged [8]. Users vary in their reasons for tagging [9]. Tagging for others produces different tagging than tagging for oneself, with factual tags used for others and subjective tags used for oneself [10]. The overlap in tags for the same object by different uses is extremely low [11]. Why tags co-occur when tagging a given object is not yet understood [12]. Much remains to be learned about tagging, and what we conclude is likely to change rapidly given that the practice is an evolving system and far from stable.

1364-6613/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2008.01.007 Available online 17 March 2008

**Box 1. Flickr: creating individual and social classification systems for photos**

Flickr is a website that allows users to upload photos from a computer, camera or mobile phone. Individual photos can be tagged manually, thereby categorizing them implicitly and making them accessible later by using the tags as retrieval cues. Users can generate their own tags or they can draw on tags of others. Photos can be linked directly into a GPS-based map so that photos of specific locations can be stored and retrieved by GPS coordinates. Clearly, Flickr offers considerable advances in creating individual classification systems for photos compared with physical photo albums.

Flickr is more than an individual classification system that organizes and tag photos. Many people use Flickr for social purposes. Users can join public groups to share photo interests, or they can create private groups such as family distribution lists for family-related photos. Groups often develop explicit tagging schemes that reflect their shared experience (e.g. 'stanfordalum'), similar to how speakers of pidgins and creoles develop hybrid phrases in natural language. Over time, users gain expertise in group tagging practices. As collections of tagged photos evolve for a group, the ability to retrieve relevant subsets becomes increasingly powerful.

Flickr illustrates how individual classification systems can evolve beyond a single individual to a group. Even when an individual classification system becomes shared, it is not shared widely like cultural classification systems are. Nevertheless, the social use of Flickr illustrates that classification systems often don't fall neatly into one type of classification system. Instead, a classification system is often a hybrid to some extent, primarily belonging to one type of system but partially exhibiting properties of other types.

## Institutional categorization

Institutions engineer classification systems explicitly to serve institutional goals, typically requiring considerable time and resources to develop, maintain and apply. Again, modern Web technology greatly enhances institutional classification systems, although groups probably have developed them for millennia. Business, industry, law and science couldn't function without these systems.

An institutional classification system increases interoperability within a group by establishing a common set of categories that allow different agents to share information effectively in pursuing goals. By creating and enforcing shared categories, an institution streamlines interactions and transactions such that consistency, fairness and higher yields can result. Because payoffs are high, expending the resources to create and maintain an institutional classification system is justified.

Two types of institutional classification systems – institutional taxonomies and institutional semantics – are common. Examples of institutional taxonomies include the Dewey decimal system for classifying books and the United Nations Standard Products and Services Code for classifying products and services. The International Organization for Standardization provides even more general classification systems, ranging from agriculture to health care to mathematics. Professions and institutions typically develop taxonomies for their domains of expertise, such as the Diagnostic and Statistical Manual of Mental Disorders (the DSM-IV) in clinical psychology [13]. Figure 1 illustrates fragments of institutional taxonomies.

An institutional taxonomy increases the likelihood that an institution's agents will classify relevant entities the same way, such as when different libraries place books in the same categories and different doctors assign patients to the same diagnostic and insurance categories. Such standardization reduces transaction costs, enhances networking and achieves many other useful outcomes. In other areas, such as science, institutional taxonomies function as engines of progress. In chemistry, for example, the periodic table has driven the search for discovering new elements. In genetics, mapping the human genome has driven not only the search for specific genes but also their role in producing phenotypic outcomes.

The other common form of an institutional classification system – institutional semantics – defines common abstractions that underlie transactions [14]. In commerce, for example, shared abstractions are required that allow buyers, sellers, banks and shippers to coordinate their activities in bringing a purchased product to a buyer. The data that a buyer provides in the fields of a Web form must have semantic equivalents in the seller's databases for customers and products, which in turn must have semantic equivalents in the applications that banks use for financial transactions and that shippers use for deliveries. To create interoperability between all parties, a common set of abstractions about all relevant aspects of transactions must be developed explicitly. Furthermore, this common set must handle diverse instances of transactions, along with the vagaries of inconsistency that occur. Once these abstractions are in place, they create an interface between parties that achieves interoperability; namely, all parties can align various aspects of the transaction. Figure 2 illustrates fragments of institutional semantics.

Institutional taxonomies and semantics are often developed by groups of individuals. Much is known about this process within institutional contexts [15] (Oasis – Technical Committee Process, [http://www.oasis-open.org/committees/process.php]), although little is known in basic science about the cognitive processes that allow groups to develop, maintain and apply institutional classification systems. Given how central these systems are to human activity, it is essential to understand the cognition mechanisms that produce them. What mechanisms allow individuals to identify abstractions that enable interoperability in an institutional domain?

Increasingly, computer scientists develop tools that automatically construct institutional classification systems, bypassing the need for humans to construct them. In some cases, rule-based systems are handcrafted to map semantic equivalents across different Web-based applications into each other (e.g. descriptions of real estate on the websites of different real estate companies). Additionally, however, programs learn these rules from mining the Web [16]. By attempting to map the attributes of two Web objects into each other, such systems develop mapping rules and models that can determine whether two Web objects are of the same type [17]. These programs use domain constraints [18], construct complex mappings [19], use probabilistic representations [20] and identify supporting evidence for mappings [21]. These programs also use machine learning to map Web ontologies (taxonomies) into each other [22].

**(a)**

532 Fluid mechanics Liquid mechanics
533 Gas mechanics
534 Sound and related vibrations
535 Light and paraphoticphenomena
536 Heat
537 Electricity and electronics
538 Magnetism
539 Modernphysics
540 Chemistry and allied sciences
541 Physical and theoretical chemistry
542 Techniques,equipment,materials
543 Analytical chemistry
544 Qualitative analysis
545 Quantitative analysis
546 Inorganic chemistry
547 Organic chemistry
548 Crystallography
549 Mineralogy

**(b)**

10101901 Butterflies
10101902 Beetles
10101903 Bees
10101904 Silkworms
10102000 Wild animals
10102001 Elephants
10102002 Live foxes
10110000 Domestic pet products
10111300 Domestic pet treatments and accessories and equipment
10111301 Pet toys
10111302 Pet grooming products
10111303 Pet litter or equipment for pet waste management
10111304 Pet food bowls or equipment
10111305 Medicated pet treatments
10111306 Domestic pet training kits
10111307 Pet blankets

**(c)**

11.040.10 Anaesthetic, respiratory and reanimation equipment
*Including medical gas installations*
11.040.20 Transfusion,infusion and injection equipment
*Including blood packs*
*Syringes, needles and catheters,see11.040.25*
11.040.25 Syringes,needles and catheters
11.040.30 Surgical instruments and materials
*Including surgical dressings,sutures,etc.*
11.040.40 Implants for surgery,prosthetics and orthotics
*Including pacemakers*
*Ophthalmic implants,see11.040.70*
11.040.50 Radiographic equipment
*Including radiographic diagnostic and therapy equipment*
*Dental, medical and industrial radiographic films, see37.040.25*

TRENDS in Cognitive Sciences

**Figure 1**. Fragments of institutional taxonomies. **(a)** Partial listings for 'Natural Sciences and Mathematics' from the Dewey decimal Classification system. **(b)** Partial listings for 'Live Plant and Animal Material and Accessories and Supplies' from the United Nations Standard Products and Services Code. **(c)** Partial listings for 'Medical Equipment' from the International Organization for Standardization.

Interestingly, individual classification plays an increasingly important role in the development of institutional classification systems. Although computer science tools provide considerable power in developing institutional classification systems, tagging within these systems by actual users provides important sources of information that automation can't provide. Box 2 describes several examples.

## Integrating cultural, individual and institutional categorization

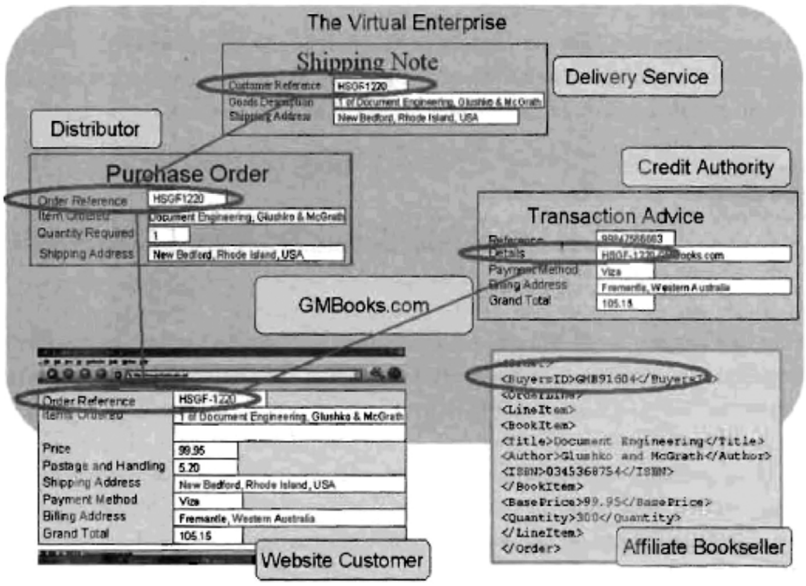
As described at the outset, large literatures across the cognitive sciences address the acquisition and use of cultural classification systems. Extensive research has made much progress understanding the mechanisms that underlie the acquisition and use of cultural categories. By contrast, relatively little is known about individual and institutional categorization in these research communities. How might these communities explain them?

### *Cultural categorization*

One possibility is that cultural categorization is the core form of categorization in humans from which individual and institutional categorization develop optionally. At early ages, children acquire cultural categories universally, effortlessly, with little instruction, simultaneously with language. Furthermore, evolutionary adaptations anticipate the acquisition of cultural categories [23]. Feature areas in the brain anticipate the features of important categories for objects, settings, events and mental states. Similarly, conjunctive biases in association areas anticipate likely correlations among features for important evolutionary categories [24]. Clearly, category learning is an epigenetic process, given that categories vary significantly across cultures [5]. Nevertheless, neural architecture anticipates important types of universal categories, thereby making it easier for all individuals to acquire them [25].

Furthermore, a variety of social and cultural mechanisms ensures that children acquire cultural categories. Social and linguistic interactions between children and caregivers play powerful roles in transmitting these categories to children [26]. The fact that words for cultural categories exist in language further ensures their acquisition. These and other sociocultural mechanisms ensure that tens of thousands of cultural concepts are transmitted from one generation to another. Sociocultural mechanisms also stabilize concepts across

(a)




(b)

```
</xs:complexType>
<xs:complexType name="PartyNameType">
   <xs:sequence>
      <xs:element name="Name" type="xs:string" minOccurs="0"/>
   </xs:sequence>
</xs:complexType>
<xs:complexType name="AddressType">
   <xs:sequence>
      <xs:element name="Room" type="xs:string"/>
      <xs:element name="BuildingNumber" type="xs:string"/>
      <xs:element name="StreetName" type="xs:string"/>
      <xs:element name="CityName" type="xs:string"/>
      <xs:element name="PostalZone" type="xs:string"/>
      <xs:element name="CountrySubentity" type="xs:string"/>
      <xs:element name="Country" type="xs:string"/>
   </xs:sequence>
</xs:complexType>
<xs:complexType name="OrderLineType">
   <xs:sequence>
      <xs:element name="LineItem" type="LineItemType"/>
   </xs:sequence>
</xs:complexType>
<xs:complexType name="LineItemType">
   <xs:sequence>
      <xs:element name="BookItem" type="BookItemType"/>
      <xs:element name="BasePrice" type="xs:decimal"/>
```

TRENDS in Cognitive Sciences

**Figure 2.** Fragments of institutional semantics from Glushko and McGrath [14]. **(a)** Fields across forms for the customer, vendor, distributor, deliverer and creditor that must be interoperable for a transaction to succeed. **(b)** Examples of abstractions (e.g. 'Name') in a transaction model that establish interoperability between analogous fields for the customer, vendor, distributor, deliverer and creditor.

individuals and introduce slow conceptual change as culture, technology and institutions evolve.

### *Individual categorization*

An individual classification system is not a core system for two reasons: not all individuals acquire them, and individuals vary widely in the systems they develop. When individuals do create these systems, they do so to support idiosyncratic goals [27–29]. Clearly, an individual classification system draws heavily on a cultural classification system, beginning with a subset of existing cultural categories. Nevertheless, variants of these categories are tailored to the specific instances categorized. For example, when developing a tagging system for pictures, existing cultural categories for objects and locations are used to categorize pictures. In the process, however, these categories acquire senses tailored to the tagging context, such as when combining cultural categories for the words 'Stanford' and 'alumnae' to form the tag 'stanfordalum'. Even when individual classification systems become

**Box 2. Individual tagging in institutional classification systems**

Some institutions use individual tagging to develop their institutional classification systems in ways not possible with standard computer science tools. A large technology corporation, for example, uses individual tagging to keep its institutional directory accurate [30]. Although employees are supposed to keep their profiles current, often they don't. To remedy this problem an individual tagging program, Fringe, allows one employee to update another employee's profile. Essentially, Fringe capitalizes on the fact that a few highly active taggers typically contribute the bulk of tags to a tagging system. By providing a tagging mechanism that supports these users' interests and skills, the corporation keeps its institutional directory more current than it would be otherwise.

Analogously, DogEar allows a large technology corporation to create a rich database of public bookmarks [7]. As employees store bookmarks, they tag them freely, using whatever categories they like. Further associated with each employee's bookmarks and tags is a profile of the employee, describing his or her position, skill set, etc. Thus, if an employee has an interest in a particular topic, say Java programming, 'Java' can be entered into DogEar to find other employees who have used 'Java' as a tag. In this manner, employees can find other employees who share their expertise or who can provide expertise they don't have. By integrating bookmarks and tags across employees a powerful institutional system of URLs develops.

In a particularly intriguing approach to mixing individual and institutional categorization, interactive online games produce effective tags for web-based resources [31]. Because these games are personally and socially rewarding, they induce Web users to provide extensive amounts of tagging data that solve large-scale institutional problems. If these games were made available on major internet sites, for example, they could provide effective tags for all images on the Web in a few weeks [32]. Similarly, these tagging games can localize focal objects in web images [33], provide text descriptions of images [34] and collect common sense knowledge useful for computer science applications [35].

These examples illustrate the considerable potential of integrating individual and institutional classifications. By capitalizing on the strengths of individual classification, institutional systems can solve computational problems that are impossible to address with standard computer science tools.

shared with other people, they remain relatively idiosyncratic and far from universal, thereby never achieving the core status of a cultural system (e.g. a shared picture-tagging system doesn't become shared by an entire culture).

The rapid explosion of tagging systems on the Web reflects the availability of powerful technology that makes it easier to develop, maintain and use individual systems. This illustrates a fundamentally important principle of human categorization mechanisms: as the context changes in which human categorization mechanisms operate, they produce new types of classification systems. When new technological tools become available, categorization mechanisms adapt quickly and new classification systems result. Rather than categorization being a fixed process, it evolves dynamically as situational constraints change.

### *Institutional categorization*

An institutional classification system is not a core system, again because not all individuals acquire it and because individuals vary widely in the systems they acquire. When groups of individuals do create a system, they do so to support group goals, such as increased quality, precision, production, marketing, distribution, etc. Clearly, an institutional classification system draws heavily on the background cultural system, beginning with a subset of existing cultural categories. Again, however, variants develop that reflect institutional constraints. Interestingly, institutional categories often feed back into cultural systems, as when the scientific concept of 'mammal' changed the categorization of whales as fish. Indeed, some institutional systems become so central to a culture that they eventually become part of the cultural classification system. Examples include the periodic table, basic number systems and basic scientific theories. Over time, institutional classification systems that become central to a culture's conceptual framework enter its shared classification system.

Developing an institutional system typically requires considerable effort. Engineering a system typically requires relatively high levels of cognitive ability and training, and typically occurs within hierarchical, authoritative, and authenticating social structures, as explicit social forces constrain its construction, maintenance, transmission and application. As a result, institutional categories often achieve greater precision, relative to cultural and individual categories.

The rapid explosion of Web technology is having transformational effects on institutional classification systems. Much more powerful institutional taxonomies and semantics result from increasingly powerful abilities to specify and share these systems. Again, the context in which basic categorization mechanisms operate affects the classification systems produced.

### *Dimensions of variability across classification systems*

Comparing cultural, individual and institutional classification systems side by side suggests dimensions on which they differ, presented in Box 3. Notably, when cultural categorization is considered in isolation, these dimensions are not particularly salient but do become salient when all three types are considered together.

Across dimensions, social factors appear to be one central difference among the three categorization types. In cultural categorization, strong social forces create stability in classification systems across individuals implicitly. Parent–child interactions, peer interactions, the media, and cultural events all contribute to this stability. Conversely, explicit social forces shape institutional classification systems, originating in institutional structure, policy and practice. Finally, social factors are relatively absent in individual categorization. Even though these individual systems can become shared, the role and impact of social factors appears considerably weaker than in cultural and institutional systems. Most importantly, viewing these three types of categorization together highlights social factors, suggesting that research should direct more attention to the social factors that create, transmit and maintain classification systems.

Interoperability also appears central to differences among categorization types. In institutions, establishing interoperability is an explicit goal when constructing classification systems. In cultures, shared categories enable different minds that reflect different experience

**Box 3. Dimensions of variation across classification systems**

**Explicitness: the amount of awareness required to acquire and use a classification system.** Cultural classification systems appear largely implicit given that children probably have relatively little awareness about their acquisition. Conversely, individual, and institutional classification systems appear largely explicit, given that individuals construct them intentionally.

**Effort: the amount of effort required to create and use a classification system.** Whereas the amount of effort is relatively low for cultural systems, it is relatively high for institutional systems and lies in between for individual systems.

**Precision: the degree of precision in specifying and applying category criteria.** Whereas exemplars and prototypes often specify cultural categories statistically, rules often specify institutional categories precisely. Individual categories take either form.

**Goals: whether a classification system serves individual or group goals.** Individual classification systems primarily serve an individual's goals, although they also can serve the goals of friends and family. Cultural systems serve an individual's goals by categorizing experience in the world effectively such that individual goals can be achieved but serve the culture's goals by creating interoperability during communication and interaction. Institutional systems primarily serve the goals for information organization and interoperability within an institution, or among a group of insitutions with common interests.

**Interoperability: the type of conceptual interface that a classification system provides.** A cultural system provides an interface between individuals in a culture. An institutional system provides an interface between institutional agents. An individual system provides an interface within an individual across events.

**Reuse: how much a classification system is used once created.** Reuse appears highest in cultural classification systems (given that the same core system evolves across generations) but appears lowest in individual systems (given that they expire with individuals). Institutional systems fall in between, depending on how long they persist within institutions and how long institutions themselves persist.

**Change: how quickly a classification system changes.** Whereas individual systems are likely to change fastest, cultural systems are likely to change slowest, with institutional systems falling in between.

**Feedback: the significance of the consequences for making a mistake.** Mistakes during cultural categorization produce negative consequences that range from violated expectations to social disapproval to death. Mistakes during institutional categorization produce lowered quality, precision, production, marketing, distribution, etc. Mistakes during individual categorization appear least significant, such as failing to find something in a personal collection.

to communicate effectively. Finally, individual categories create interoperability over time within an individual, making it possible to coordinate experience across different events.

## Implications for cognitive science research

Our treatment of individual and institutional categorization, along with their relation to cultural categorization, barely scratches the surface. Clearly much research is necessary to develop satisfactory accounts. In this brief opinion article, our goal has been to bring individual and institutional categorization to the attention of researchers who study cultural categorization. We hope that our speculation piques their interest and outlines useful directions for productive investigation.

Several issues and goals strike us as worth pursuing in the near future. First, to what extent do the mechanisms that underlie cultural categorization produce individual and institutional categorization? What additional mechanisms are important? Second, how do categorization mechanisms produce new types of classification systems as the context in which they operate changes? How does evolving human technology produce new classification systems and categorization behaviors? Third, although relatively pure forms of cultural, individual and institutional classification systems appear possible, many hybrids exist, such as individual tagging systems that become social and institutional classification systems that become cultural. What does the dynamical production of hybrids tell us about human categorization? Finally, what will examining cultural, individual and institutional categorization together teach us that wouldn't be possible from studying cultural categorization alone? We suspect that viewing categorization broadly will yield important insights not possible otherwise. In particular, we believe that understanding the roles of social factors and interoperability has the potential to produce fundamental new insights into categorization.

## Acknowledgements

This article summarizes a symposium presented by the four authors at the August 2007 meeting of the Cognitive Science Society in Nashville, TN, USA. We are grateful to Robert Goldstone and two anonymous reviewers for extremely helpful comments on this article.

## References

1 Smith, E.E. and Medin, D.L. (1981) *Categories and Concepts,* Harvard University Press
2 Murphy, G.L. (2002) *The Big Book of Concepts,* MIT Press
3 Prinz, J. (2002) *Furnishing the Mind: Concepts and Their Perceptual Basis,* MIT Press
4 Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45
5 Atran, S. and Medin, D.L. *The Native Mind and the Cultural Construction of Nature*, MIT Press (in press)
6 Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigation,* University of Chicago Press
7 Millen, D.R. *et al.* (2006) Dogear: social bookmarking in the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Grinter, R. *et al.*, eds), pp. 111–120, Association for Computing Machinery
8 Golder, S. and Huberman, B.A. (2006) Usage patterns of collaborative tagging systems. *J. Inform. Sci.* 32, 198–208
9 Ames, M. and Naaman, M. (2007) Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Rosson, M.B. and Gilmore, D., eds), pp. 971–980, Association for Computing Machinery
10 Sen, S. *et al.* (2006) Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Hinds, P. and Martin, D., eds), pp. 181–190, Association for Computing Machinery
11 Marlow, C. *et al.* (2006) HT06, tagging paper, taxonomy, Flickr, academic article, ToRead. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (Will, U.K. *et al.*, eds), pp. 31–40, Association for Computing Machinery
12 Kipp, M.E.I. and Campbell, D. (2006) Patterns and inconsistencies in collaborative tagging practices: an examination of tagging practices. In *Proceedings of the Annual General Meeting of the American Society for Information Science and Technology*, Austin, TX, USA (http://eprints.rclis.org/archive/00008315/)
13 American Psychiatric Association (1994) *The Diagnostic and Statistical Manual of Mental Disorders,* (4th edn), American Psychiatric Publishing, Inc
14 Glushko, R.J. and McGrath, T. (2005) *Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services,* MIT Press

15 Greenstein, S. and Stango, V., eds (2007) *Standards and Public Policy*, Cambridge University Press
16 Doan, A. and Halevy, A. (2005) Semantic integration research in the database community: a brief survey. *AI Mag.* 26, 83–94
17 Tejada, S. *et al.* (2001) Learning object identification rules for information integration. *Inf. Syst.* 26, 607–633
18 Doan, A. *et al.* (2001) Reconciling schemas of disparate data sources: a machine learning approach. In *Proceedings of the ACM SIGMOD Conference on Management of Data* (Mehrotra, S. and Sellis, T., eds), pp. 509–520, Association for Computing Machinery
19 Dhamankar, R. *et al.* (2004) iMAP: discovering complex matches between database schemas. In *Proceedings of the ACM SIGMOD Conference* (Valduriez, P. *et al.*, eds), pp. 383–394
20 Mimno, D. *et al*. (2007) Probabilistic representations for integrating unreliable data sources. In *Information Integration on the Web (IIWeb), Papers from the 2007 AAAI Workshop* (Ullas, N. and Nie, Z., eds), pp. 74–79, AAAI Press
21 Michalowski, M. *et al.* (2005) Automatically utilizing secondary sources to align information across sources. *AI Mag.* 26, 33–43
22 Doan, A. *et al.* (2003) Learning to match ontologies on the semantic web. *Intl. J. Very Large Data Bases* 12, 303–319
23 Caramazza, A. and Shelton, J.R. (1998) Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *J. Neurosci Cogn.* 10, 1–34
24 Simmons, W.K. and Barsalou, L.W. (2003) The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20, 451–486
25 Malt, B.C. (1995) Category coherence in cross-cultural perspective. *Cognit. Psychol.* 29, 85–148
26 Gelman, S.A. *et al.* (1998) Beyond labeling: the role of maternal input in the acquisition of richly structured categories. *Monogr. Soc. Res. Child Dev.* 63, 1–148
27 Barsalou, L.W. (1983) Ad hoc categories. *Mem. Cognit.* 11, 211–227
28 Barsalou, L.W. (1991) Deriving categories to achieve goals. In *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 27) (Bower, G.H., ed.), In pp. 1–64, Academic Press
29 Barsalou, L.W. (2003) Situated simulation in the human conceptual system. *Lang. Cogn. Process.* 18, 513–562
30 Farrell, S. *et al.* (2007) Socially augmenting employee profiles with people-tagging. In *Proceedings of the 20th Annual ACM Symposium on User interface Software and Technology* (Chen, C. *et al.*, eds), pp. 91–100, Association for Computing Machinery
31 von Ahn, L. (2006) Games with a purpose. *IEEE Computer Magazine* 39, 96–98
32 von Ahn, L. and Dabbish, L. (2004) Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)* (Dykstra-Erickson, E. and Tscheligi, M., eds), pp. 319–326, Association for Computing Machinery
33 von Ahn, L. *et al.* (2006) Peekaboom: A game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)* (Grinter, R. *et al.*, eds), pp. 55–64, Association for Computing Machinery
34 von Ahn, L. *et al.* (2006) Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI Notes 2006)* (Olson, G. and Jeffries, R., eds), pp. 79–82, Association for Computing Machinery
35 von Ahn, L. *et al.* (2006) Verbosity: a game for collecting common-sense knowledge. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI Notes 2006)* (Olson, G. and Jeffries, R., eds), pp. 75–78, Association for Computing Machinery