The Semantic Web and Human Inference: A Lesson from Cognitive Science

Takashi Yamauchi

Mail Stop 4235, Department of Psychology,
Texas A&M University College Station, TX 77843 USA
tya@psyc.tamu.edu

Abstract. For the development of Semantic Web technology, researchers and developers in the Semantic Web community need to focus on the areas in which human reasoning is particularly difficult. Two studies in this paper demonstrate that people are predisposed to use class-inclusion labels for inductive judgments. This tendency appears to stem from a general characteristic of human reasoning – using heuristics to solve problems. The inference engines and interface designs that incorporate human reasoning need to integrate this general characteristic underlying human induction.

1 Introduction

In order for the Semantic Web to remain technologically viable, it must either drastically expand human abilities or create totally new experiences in our lives. Many inventions have achieved this: the automobile expanded people's mobility; TV elevated the scope of our experience; E-mail facilitated the ability to communicate. Unless Semantic Web can change human experience and ability in a fundamental manner, its technological innovation will be short-lived.

How can the Semantic Web change our lives? One goal of the Semantic Web is to aide complex inferential tasks such as medical diagnosis, business decisions, or investment predictions by making the vast amount of data now available on the World Wide Web useful for the machine [1]. However, to develop successful inference agents, semantic web researchers need to know exactly what is needed to supplant human inferential behavior – how people make inferential reasoning, when their reasoning becomes irrational and fallible, and how a computer program can help fulfill our shortcomings.

The question therefore boils down to the following: are the existing approaches, such as RDF, RDFS, OWL, and other inferential agents, sound enough in principle to support human reasoning abilities? If not, what is needed for the evolution of the Semantic Web? Study in human cognitive psychology can make an important contribution in this regard.

In this paper, I will describe two cognitive studies conducted in my laboratory and illustrate how people (college students) solve simple inferential questions using ontological "tags." In so doing, I aim to provide insight into the development of

inferential agents in the Semantic Web environment from the perspective of human cognitive psychology.

In a nutshell, the present studies show two things: one, that people display a significant tendency to use class-inclusion labels to make inferential judgments, and two, that they separate category labels from property labels and make "rule-like" reasoning. This disproportional reliance on categorical class labels appears to stem from two sources: (a) human reasoning is intertwined with language, thus linguistic categories, such as count nouns and adjectives, influence the way people make inductive reasoning; (b) because of processing constraints, people often apply simple heuristics, such as using class-inclusion labels, to solve inferential problems [2]. The inference engines and interface designs that incorporate human reasoning need to take advantage of these fundamental characteristics of human reasoning in order to be successful. Specifically, it may be useful to separate ontological status of class-information (applicable for count nouns) and property information (applicable for adjectives). Furthermore, applications of Semantic Web technology should focus on the areas in which human reasoning is particularly vulnerable.

2   Study 1

Consider a simple reasoning experiment, in which subjects are shown schematic pictures of cartoon insects side by side and are asked to predict a hidden value of an attribute of a test insect on the basis of a sample insect (Fig. 1). One selection is consistent with the attribute value shown in the sample insect (e.g., long horns in Fig. 1a). The other selection is inconsistent with the feature shown in the sample insect (e.g., short horns in Fig. 1a).
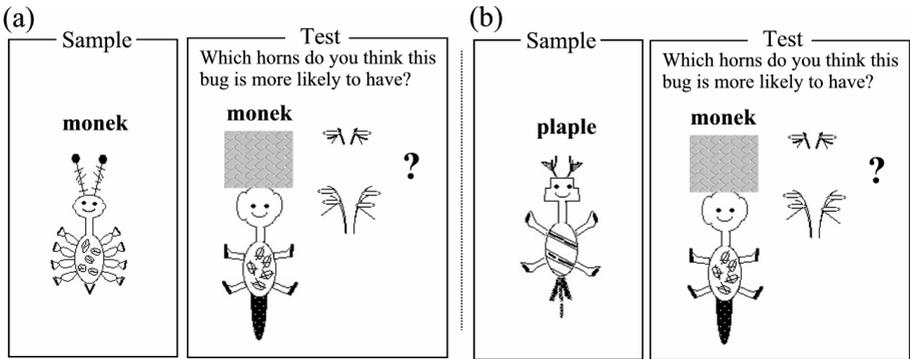

Fig. 1. Two examples of a stimulus frame in the inference task in Studies 1 and 2. In one group of trials, both sample and test stimuli had the same labels ("monek" and "monek"). In the other group of trials, sample and test stimuli had different labels ("plaple" and "monek"). Subjects were asked to predict the hidden value of body parts in test stimuli on the basis of sample stimuli shown right next to the test stimuli. In this manner, this experiment assessed the extent to which the attribute value of one stimulus (sample stimulus) is projected to other stimuli.

Above each insect, an arbitrary tag ("monek" or "plaple") is placed. The main question is how the probability of selecting the consistent attribute (selecting the long horns in Fig. 1a) would change as the ontological status of the arbitrary tags is modified. As subjects rely on the labels to make judgments, the probability of making a consistent choice would go up or down sharply depending on whether sample and test stimuli have the same tags (Fig. 1a; "monek" vs. "monek") or different tags (Fig. 1b; "plaple" vs. "monek"). For example, many subjects would select long horns in Fig. 1a (consistent choice), but few subjects would choose short horns in Fig. 1b. The question is how people's inferential behavior would change when these arbitrary tags carry different ontological information, class inclusion information or property possession information.

In four independent experiments, a group of undergraduate students received the same stimuli and answered the same inferential questions (60 trials in total). However, the instructions they received were modified slightly so that these arbitrary labels ("monek" and "plale") represented different ontological information. In a class-tag condition, the instructions characterized the two arbitrary labels ('monek" and "plaple") as representing two *types* that these insects belong to. In a property-tag condition, the instructions characterized the same arbitrary labels as representing shapes of wings hidden underneath the insect's body. In a pictorial class-tag condition and a pictorial property-tag condition, the labels were removed and replaced with pictorial symbols (Fig. 2).
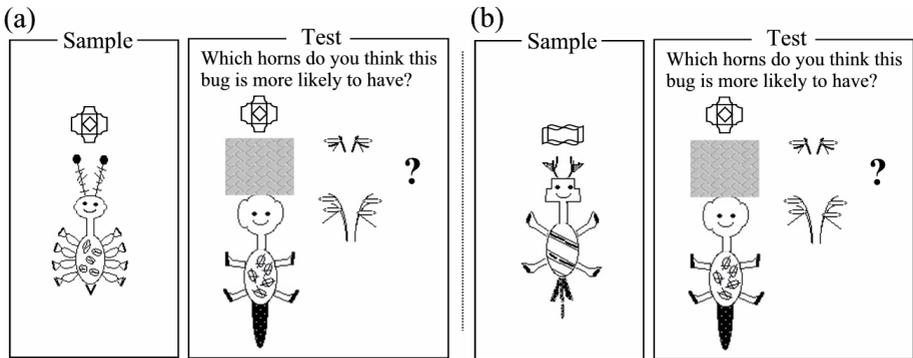

Fig. 2. Two examples of the stimulus frames used in the pictorial class-tag and pictorial property-tag conditions in Study 1. Note that the verbal labels ("monek" and "plaple") in Fig. 1 are replaced with pictorial signs in these conditions.

In this manner, the class-tag condition characterized these labels with class inclusion information, and the property-tag condition characterized these labels with "property" information. In both pictorial class-tag and pictorial property-tag conditions, the two verbal labels ("monek" and "plaple") were replaced with pictorial signs. These pictorial signs were described as representing two "types" of insects (pictorial class-tag condition) or different shapes of insects (pictorial property-tag condition). Except for these points, all subjects received the same stimuli and answered the same inferential questions.

The probability of selecting attributes consistent with sample stimuli would go up or down drastically depending on whether sample and test stimuli have the same or different labels (Fig. 1 and Fig. 3). However, this shift would depend on the ontological information that these tags represent. For example, when the tags represent names of the "types" that these insects belong to, the probability of selecting attribute values consistent with the sample stimuli (i.e., "consistent choice") would go up. On the other hand, when the same tags represent a property that these insects have, the probability of consistent choice would go down. In this manner, this study helped identify the extent to which different ontological tags (class information vs. property information; verbal labels vs. pictorial signs) would influence the inference of unknown attributes [3].
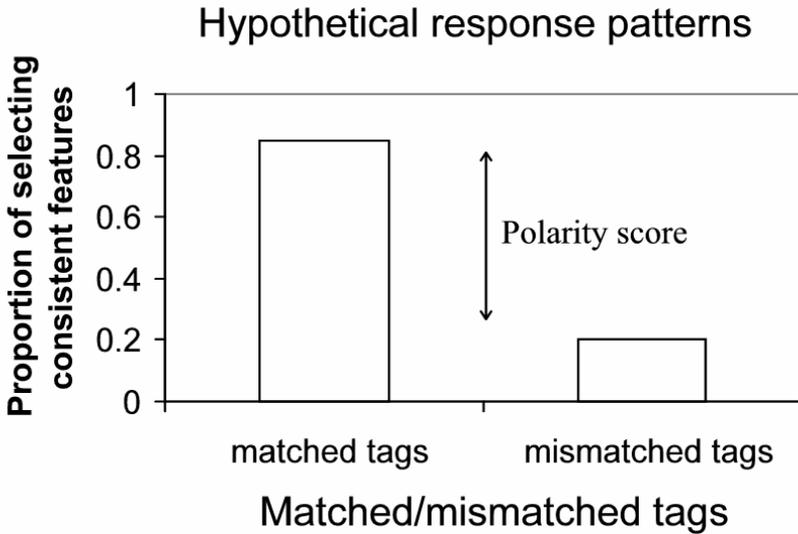

Fig. 3. A hypothetical response pattern. This response pattern shows a heavy reliance on tags. As the sample and test stimuli have the same tags (see Fig. 1a), the proportion of making consistent choices is high. However, when the sample and test stimuli have different labels, the proportion of making consistent choices goes down sharply. The difference between the two proportions is defined as a "polarity score."

In Study 1, I present the results from four experiments, and show a clear indication that people tend to rely on verbal labels representing class-inclusion information to make inferential projections. Study 2 further shows that this bias is present even when "class" division does not have proper meaning, suggesting that using class-inclusion labels is likely to be a default strategy for human reasoning.

2.1 Method

2.1.1 Participants
A total of 175 undergraduate students participated in this study (the class-tag condition, $n=49$; the pictorial class-tag condition, $n=51$; the property-tag condition,
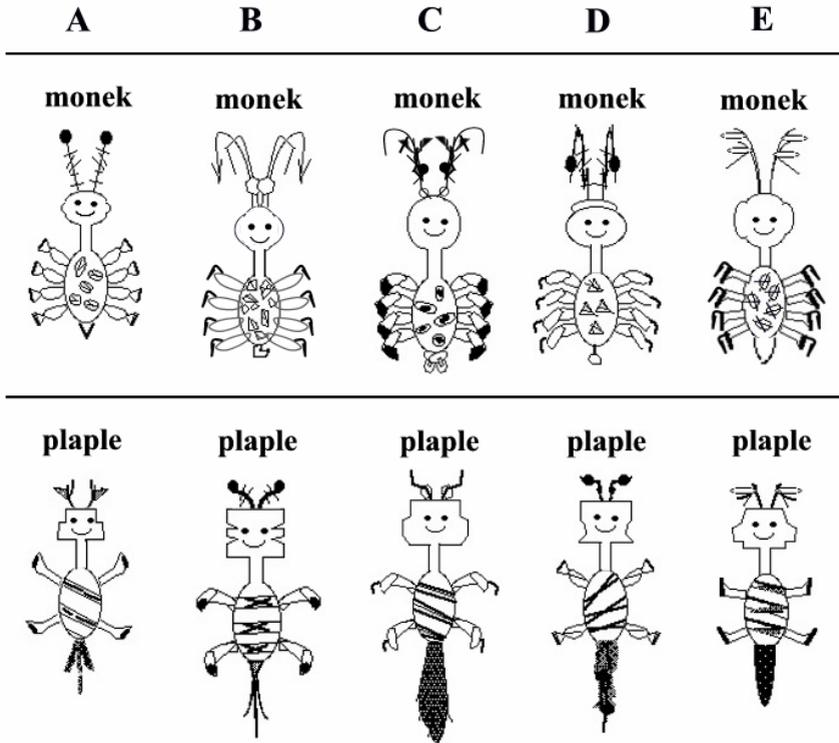
Fig. 4. Five sets of prototypes used to produce test stimuli. One hundred test stimuli were produced by swapping two feature dimensions of two corresponding categories, "monek" and "plaple." See also Table 1.

n=41; the pictorial property-tag condition, n=34). The participants were recruited from the Psychology Subject Pool administered at Texas A&M University.

2.1.2 Materials

The stimulus materials were schematic illustrations of cartoon insects produced from five sets (A, B, C, D, and E) of prototypes (Fig. 4). Each stimulus was composed of a combination of five feature dimensions with binary values (horns=long/short, head=round/angular, torso=dotted/striped, legs=8 legs/4 legs, tail=short/long), along with category labels ('monek'/'plaple') (Table 1). The depiction of these features varied across the five stimulus sets, while they maintained abstract commonalties. For example, the monek prototypes in the five sets all had long horns, round heads, dotted torsos, 8 legs and short tails, while the exact appearance of these components was different across the sets.

Individual trials consisted of pairs of a sample stimulus and a test stimulus (Fig. 1). We created 100 test stimuli from the five sets, A, B, C, D, and E (20 stimuli from

each set – Fig. 4). These test stimuli were produced systematically by exchanging in-
dividual components of the two prototypes. All test stimuli had 2 features consistent
with the prototype of one category and 2 features consistent with the prototype of the
other category, and 1 feature was masked for an inference question (Table 1). One
version of the stimulus materials was produced from prototype sets A, B, and C. The
other version was produced from prototype sets A, D, and E (Fig. 4). The sample
stimuli, which were shown right next to the test stimuli, were the two prototype stim-
uli of set A.

Table 1. The structure of the test stimuli used in Studies 1 and 2. (1, 0)=Horns(long horns,
short horns), Head(round, angular), Body(dotted, striped), Legs(8 legs, 4 legs), Tail(short,
long), Labels(monek, plaple). *?/1* refers to the dimension queried in each trial, and the "consis-
tent response" for that question. For example, given test stimulus M1, the horns dimension was
queried, and the selection with 1 (long horns) is defined as a "consistent response."

	Horns	Head	Body	Legs	Tail	Labels
M1	*?/1*	1	1	0	0	1
M2	1	1	0	0	*?/1*	1
M3	1	0	0	*?/1*	1	1
M4	0	0	*?/1*	1	1	1
M5	0	*?/1*	1	1	0	1
M0 (sample stimulus)	1	1	1	1	1	1
P1	*?/0*	0	0	1	1	0
P2	0	0	1	1	*?/0*	0
P3	0	1	1	*?/0*	0	0
P4	1	1	*?/0*	0	0	0
P5	1	*?/0*	0	0	1	0
P0 (sample stimulus)	0	0	0	0	0	0

2.1.3 Procedure

For each trial, participants were shown a pair of sample and test stimuli on a com-
puter screen, and were instructed to select one of two feature values for the body
part in question. Every participant received 30 test stimuli twice (60 trials in total).
In one case, a test stimulus was paired with the prototype of the corresponding

category (i.e., *match* condition) (Fig. 1a). In the other case, the same test stimulus was paired with the prototype of the other category (i.e., *mismatch* condition) (Fig. 1b). For example, stimulus M1 in Table 1 was shown twice, once with the sample stimulus M0 (this is called a *matched* trial because M1 and M0 had the same label 'monek') and once with the sample stimulus P0 (this is called a *mismatched* trial because M1 and P0 had different labels, 'monek' and 'plaple'). Each participant received a total of 60 trials (20 test stimuli each from sets A, B, and C or sets A, D, and E – Fig. 4).

Participants indicated their responses by clicking one of the two designated buttons. The order of presenting stimuli was determined randomly for each participant.

2.2 Results and Discussion

Each condition was treated as an independent study (no participants participated in these experiments more than once). Because these experiments were given in separate semesters, the four experiments were compared by a meta-analytic procedure [4]. Specifically, we compared the effect size *r* of the polarity score (Fig. 3) obtained in each experiment with the following equations:

$$\text{Effect size } r = \sqrt{\frac{t^2}{t^2 + df}} \tag{1}$$

$$\text{Fisher } Z_r = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) \tag{2}$$

$$\frac{Z_{r1} - Z_{r2}}{\sqrt{\dfrac{1}{N_1 - 3} + \dfrac{1}{N_2 - 3}}} \tag{3}$$

where N_1 and N_2 represent the number of subjects in each of two experiments, and (3) is distributed as Z.

Fig. 5 and Table 2 summarize the major results from Study 1. As Fig. 5 shows, the impact of matched/mismatched labels was considerably larger when the labels were characterized with class inclusion information (class-tag condition). When sample and test stimuli had the same tags, the proportion of making consistent responses was high. When sample and test stimuli had different tags, the proportion of making consistent responses declined sharply. This tendency decreased substantially when the same labels were characterized with property information. It is also evident that verbal labels, but not pictorial signs, were more important for inferential projections. When the verbal tags were replaced with pictorial signs, the polarity score declined substantially.
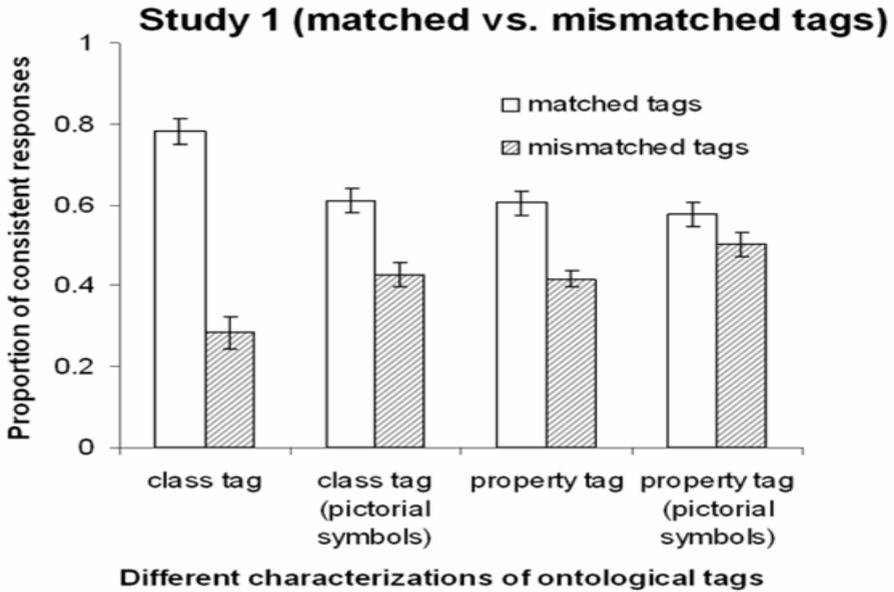
Fig. 5. A summary of Study 1. The error bars represent two standard error units obtained in each condition.

Table 2. Main results from Study 1

	class tag	property tag	class tag (pictorial symbols)	property tag (pictorial symbols)
effect size (r)	0.79	0.50	0.52	0.36
Fisher Z	1.08	0.55	0.58	0.38

Comparing Fisher Z's in all pairs (p-value)

	class tag	property tag	class tag (pictorial symbols)	property tag (pictorial symbols)
class tag	-			
property tag	0.01	-		
class tag (pictorial symbols)	0.01	0.14	-	
property tag (pictorial symbols)	0.001	0.71	0.89	-

Overall, the effect size comparing the impact of matched and mismatched tags was particularly large in the class-tag conditions as compared to the other conditions; Z's>2.37, p's<0.01 (Table 2), suggesting a clear inclination to use category labels for the prediction of body parts.

3 Study 2

Study 1 indicates a strong preference for using class-inclusion tags for inferential predictions. It is likely that using these categorical labels is a default strategy for inferential reasoning, and Study 2 tested this idea. In Study 2, the instructions in one condition (a random-class tag condition) explicitly stated that the two "classes" ("monek" and "plaple") of these cartoon insects were determined randomly by a coin toss, and there was no particular meaning attached to these types. This condition was compared to other property-tag conditions, in which the same arbitrary tags were characterized as representing names of diseases that these insects carry, or names of the islands on which these insects live. If using class-inclusion tags is a default reasoning strategy, the polarity score obtained in the random-class tag condition should be still larger than those obtained in the other property-tag conditions.

3.1 Method

3.1.1 Participants
A total of 164 undergraduate students participated in the experiment for course credit (random-class tag, n=57, property-tag (disease), n=49; property-tag (island), n=58).

3.1.2 Materials and Procedure
One version of the stimulus materials was produced from prototype sets A and B. The other version was produced from prototype sets A and C (Fig. 4). In total, each participant received 40 test trials. In the random-class tag condition, the instructions specified the arbitrary tags representing two types of the insects that were determined randomly by a coin toss. In the other property-tag conditions (disease and island), the instructions stated that these tags represent names of diseases or islands that these imaginary insects carry or live on. Except for this single point, the three conditions were identical in their procedures, materials, and designs. Individual participants participated in one of the 3 experiments, and no participants participated in the experiment more than once.

3.2 Results and Discussion

As in Study 1, each experiment was treated as an independent study (no participants participated in these experiments more than once), and these studies were examined with a meta-analytic procedure by comparing the effect sizes of the polarity scores obtained in each experiment.

Fig. 6 and Table 3 summarize the major results from Study 2. The most important result in Study 2 is that the polarity score obtained in the random-class tag condition is still substantially higher than those obtained in the other property-tag conditions, suggesting that using class-inclusion labels is somewhat automatic for many participants.
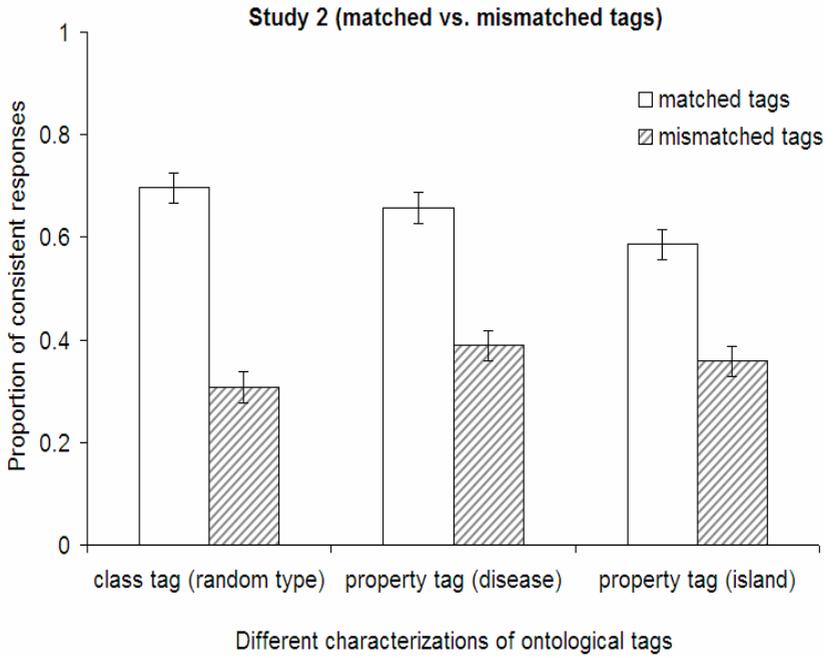
Fig. 6. A summary of Study 2. The error bars represent standard errors obtained in each condition.

Table 3. Main results from Study 2

	class tag (random type)	property tag (disease)	property tag (island)
effect size (r)	0.74	0.55	0.54
Fisher Z	0.95	0.62	0.61

Comparing Fisher Z in every pair (p-value)

	class tag (random type)	property tag (disease)	property tag (island)
class tag (random type)	-		
property tag (disease)	0.05	-	
property tag (island)	0.04	0.48	

As Fig. 6 shows, the impact of matched/mismatched labels was considerably larger when the labels were characterized with class inclusion information. When sample and test stimuli had the same tags, the proportion of making consistent responses was high. When sample and test stimuli had different tags, the proportion of making consistent responses declined sharply. This tendency decreased substantially when the same labels were characterized with property information.

Overall, the effect size obtained in the random-class tag condition was significantly larger than that in the property tag (island) condition; $Z=1.70$, $p<0.05$. The effect size difference between the random-class tag condition and the property-tag (disease) condition was also substantial; $Z=1.61$, $p=0.053$ (Table 3), suggesting a sizable advantage for a random-class tag over the two property tags. Taken together, these results confirm that there is a strong tendency to use class-labels to make predictions.

4   General Discussion

We manipulated the meanings associated with arbitrary labels attached to imaginary insects, and examined how the manipulation would influence the way the subjects predict the attributes of these insects. In both Study 1 and Study 2, the college students displayed a strong tendency to use class inclusion labels to make their inferential judgments. This tendency was reduced significantly when the verbal tags were replaced with pictorial signs, or when the tags represented various kinds of property information, such as the disease, location, or the shape of wings that these imaginary insects carry, live, or possess.

Why does human reasoning heavily rely on verbal labels representing class inclusion information? One source of this bias seems to come from the need to maintain *cognitive economy* [5]. One of the most compelling findings in the cognitive psychology of inductive inference is that people make a predictive inference on the basis of the categorical information that is immediately recognizable [6]-[10]. By arranging concepts categorically, we treat individual objects as a group and deal with the characteristics of the group as a whole, rather than individual objects separately. "Grouped" representation can expedite many cognitive tasks, and help overcome the processing limitations of the human brain [11]. As a result, we often gloss over individual differences and draw erroneous conclusions about individuals (e.g., stereotyping).

Another important source of the over-reliance on class inclusion labels is likely to stem from the influence of language. Category labels generally correspond to count nouns (e.g., animals, dogs, cats, or apples). They are subject to linguistic constraints to a larger degree than adjectives, because count nouns vastly outnumber adjectives in linguistic communications. This may be an important reason why noun labels are fundamentally different from attribute labels and affect our inductive inference in a significant manner [12]. For example, noun labels can be used as a metaphor ("my job is a *jail*"), or to help activate a particular aspect of a concept ("the *pen* is mightier than the *sword*"). When noun labels are used metaphorically ("my job is a *jail*"), the label *jail* does not represent a collection of individual instances of "jail." Rather, the label accentuates one aspect of the concept (e.g., being captive). In this manner, the meaning associated with noun labels is determined in the context of communication, and categorical noun labels can flexibly influence our inferential behavior.

Implications for Semantic Web research. What does the current finding tell us in terms of the development of the Semantic Web? There are several important implications. First, the developers and researchers in the Semantic Web community may be better off by focusing on the areas in which human reasoning is particularly vulnerable. Our strong inclination to use class inclusion labels comes from the need to process information quickly in light of our processing constraints. This bias occurs particularly when processing information is mentally cumbersome. Unlike computers, the human brain is far limited in its processing capacity. When reasoning involves many options and conditions to weigh, we rely on simple heuristics, such as the recognizablility and representativeness of a selection.

The application of the Semantic Web can be especially important in fields such as product selection, medical diagnosis, and human resources management. For example, in selecting a home insurance policy, we need to process a large amount of regulations, conditions, criteria and so on. Insurance companies often deliberately fill their policies with many unrecognizable options, making it difficult for lay people to grasp necessary information. Selecting appropriate drugs also require extensive processing of the information related to their ingredients, possible side effects, and effectiveness. The Semantic Web can help people make rational decisions by helping reduce processing overload. A similar application can be developed in many areas of product selection, such as the selection of infant foods, automobiles, schools, and houses. Clearly, machines appear to do much better than humans in these areas.

Another potentially promising area of Semantic Web application can be medical diagnosis. According to recent statistics, 1 in every 47 diabetic patients (5.7 million US citizens) were misdiagnosed.[1] Misdiagnosis occurs primarily due to the failure to take a variety of patient information into account, such as his/her family history, lifestyle, age and racial marker. With the advent of genome research, individuals' genomic information will also become available to clinicians in the future, and doctors will be increasingly fraught with individuated information. By developing inference agents tailored for specific patents, medical diagnosis can be facilitated significantly by using the Semantic Web technology [13] [14].

Human resources management can be another promising area. Selecting appropriate employees, and providing sound services to individual employees (e.g., health insurance and retirement benefits, and employee training) requires a vast amount of individuated information. This process can be facilitated by utilizing semantic web agents that are appropriately geared to analyze the information about individual employees.

Finally, the present study indicates that OWL researchers and developers may need to consider vocabularies that are aligned with linguistic classes, such as count nouns, adjectives and verbs. Likewise, algorithms for ontology matching may also need to integrate these linguistic classes. Many agents in ontology matching measure the semantic similarity between ontologies by a weighted sum of individuated "similarity factors" [15] [16] [17]. These linguistic classes can be a dominant factor in determining the degree of matching between ontologies. Because human knowledge

---

[1] This statistics is taken from The National Women's Health Information Center, U. S. Department of Health and Human Services and reported at
http://www.wrongdiagnosis.com/intro/notdiagcommon.htm.

is primarily expressed by language, our conceptual activities (e.g., inferential reasoning) are also subject to the structure of language [18] [19]. Researchers in the Semantic Web community may need to pay attention to this relationship.

References

1. Walton, C.D.: Agency and the Semantic Web. Oxford University Press, Oxford (2007)
2. Murphy, G.L.: The Big Book of Concepts. MIT Press, Cambridge (2002)
3. Yamauchi, T., Kohn, N., Yu, N.: Tracking mouse movement in feature inference: Category labels are different from feature labels. Mem Cognition 35, 852–863 (2007)
4. Rosenthal, R.: Meta-analytic Procedures for Social Research. Sage Publications, London (1984)
5. Rosch, E.: Principles of Categorization. In: Rosch, E., Lloyed, B.B. (eds.) Cognition and Categorization, pp. 27–48. Lawrence Erlbaum, Hillsdale, NJ (1978)
6. Lagnado, D.A., Shanks, D.R.: The Influence of Hierarchy on Probability Judgment. Cognition 89, 157–178 (2003)
7. Murphy, G.L., Ross, B.H.: Predictions from Uncertain Categorizations. Cognitive Pscychol. 27, 148–193 (1994)
8. Ross, B.H., Murphy, G.L.: Category-based Predictions: Influence of Uncertainty and Feature Associations. J. Exp. Psychol. Learn 22, 753–763 (1996)
9. Yamauchi, T., Markman, A.B.: Inference Using Categories. J. Exp. Psychol. Learn 26, 776–795 (2000)
10. Yamauchi, T.: Labeling Bias and Categorical Induction: Generative Aspects of Category Information. J. Exp. Psychol. Learn 31, 553–568 (2005)
11. Gigerenzer, G., Todd, P.M.: Simple Heuristics That Make Us Smart. Oxford University Press, New York (1999)
12. Langacker, R.W.: Conceptualization, Symbolization, and Grammar. In: Tomasello, M. (ed.) The New Psychology of Language: Cognitive and Functional Approach to Language Structure, pp. 1–39. Lawrence Erlbaum, Mahwah (1998)
13. Shankar, R.D., Martin, S.B., O'Connor, M.J., Parish, D.B., Das, A.K.: Towards Semantic Interoperability in a Clinical Trials Management System. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 901–912. Springer, Heidelberg (2006)
14. Sheth, A., Agrawal, S., Lathem, J., Oldham, N., Wingate, H., Yadav, P., Gallagher, K.: Active Semantic Electronic Medical Record. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 913–926. Springer, Heidelberg (2006)
15. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. In: Spaccapietra, S. (ed.) Journal on Data Semantics IV. LNCS, vol. 3730, pp. 146–171. Springer, Heidelberg (2005)
16. Euzenat, J., Loup, D., Touzani, M., Valtchev, P.: Ontology alignment with OLA. In: Proceedings of the 3rd International Workshop on the Evaluation of Ontology-based Tools 2004 (2004)

17. Doan, A., Madhavan, J., Domingos, P., Halevy, A.: Ontology matching: A machine learning approach. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 385–403. Springer, Heidelberg (2004)
18. Genter, D., Goldin-Meadow, S.: Whiter Whorf. In: Genter, D., Goldin-Meadow, S. (eds.) Language in Mind: Advances in the Study of Language and Thought, pp. 3–14. MIT Press, Cambridge (2003)
19. Yamauchi, T.: Linking Syntax and Inductive Reasoning: Categorical Labeling and Generic Noun Phrases. Psychologia (in press)