

War* vs 励志† in Forrest Gump: Cultural Effects in Tagging Communities

Zhenhua Dong¹, Chuan Shi², Shilad Sen³, Loren Terveen², John Riedl²

¹Dept. of Information Technical Science
Nankai University Tianjin
Tianjin, China
dongzh@mail.nankai.edu.cn

²GroupLens Research
Department of Computer Science
University of Minnesota
Minneapolis, Minnesota
{chshi,terveen,riedl}@cs.umn.edu

³MSCS Department
Macalester College
St. Paul, Minnesota
ssen@macalester.edu

Abstract

People from different cultures vary in cognition, emotion, and behavior. We explore cultural differences in a tagging system. We developed a model of cultural differences and performed a controlled empirical study with American and Chinese subjects to investigate questions that arise from the model. American and Chinese subjects differed in many ways: the number and types of tags they applied; the extent to which they applied suggested tags or entered new tags of their own; and how often they applied tags that originated from a different culture. Our results are consistent with theories of cultural differences between Asian and Western cultures. Our findings suggest new opportunities and mechanisms for shaping user behavior to produce useful tag repositories.

1 Introduction

Tagging is a ubiquitous feature in social web sites. Users can associate short text labels of their choice with items such as movies, electronics, photos, web pages, and blog posts. Features like tag clouds help people understand and explore the space of items. Users' individual tendencies drive their tagging choices: one user might tag the movie "The King's Speech" as a "period drama" while another tags it "abdication." However, many sites also use auto-complete tag suggestions to encourage coherent tag repositories. For example, a user who begins typing "period drama" may instead choose "period piece" if the system tells them other taggers prefer it (both tags have been applied to "The King's Speech" on Amazon.com).

Culture is a set of attitudes, beliefs, values, and practices shared by a people. We can speak of the culture of a nation, an ethnic group, an organization, etc. A culture's core values shape how its people make decisions, interpret the world, and relate to others (Hofstede, Hofstede, and Minkov 2010). Not surprisingly, work in CSCW has found that culture plays a significant role in how people use social systems. Such studies are crucial, because they help in properly interpreting and generalizing findings and techniques.

* 战争, in Chinese

† *inspirational*, in English

Our research studies the role of culture in tagging. We specifically conducted an online survey in which subjects were given the opportunity to tag 20 movies. We administered the survey to English-speaking American and Mandarin-speaking Chinese participants (each in their own native language), and we selected suggested tags for each movie from either an American or Chinese source. This study design lets us pose three research questions:

1. Do users from these two cultures have distinct personal tagging tendencies? Left to their own devices, do they tag differently?
2. Do users from these two cultures respond differently to tag suggestions?
3. Do users from these two cultures respond differently when the source of a suggested tag is from their own culture or the other culture?

To preview our results, the answer to all these questions is "yes". These results are important for both researchers and designers of tagging systems. They help researchers properly scope and interpret observed patterns of user behavior, for example, in identifying factors that cause the behavior. They offer guidance to designers who want to meet user needs and shape user behavior, for example, by suggesting new techniques for tag suggestion algorithms.

The remainder of the paper is organized as follows. We first survey related work, detailing how we draw on and extend it. We then present our research framework, including a model of how culture relates to the tagging process. We next present our research results and then discuss their implications. Finally, we close with a brief summary.

2 Related Work

Since we study tagging behavior of Chinese and Americans, we first survey relevant work on differences between Western and Asian culture. We then review studies of culture in online communities and collaborative interaction, then narrow our focus to studies of cultural influences on tagging.

Cultural Differences A large body of social science research has identified systematic differences between cultures. We focus on research comparing Asian and Western cultures. **Individualism vs. Collectivism.** Asian cultures emphasize the fundamental relatedness of individuals to each other, with a focus on living harmoniously with

others (Hofstede, Hofstede, and Minkov 2010; Markus and Kitayama 1991; Nisbett et al. 2001). On the other hand, American culture emphasizes connectedness among individuals less. Instead, individuals seek to maintain their independence by seeking out and expressing their unique inner attributes (Lee, Joshi, and McIvor 2007). **Analytic vs. Holistic.** (Nisbett et al. 2001) finds that Westerners tend to favor context-independent, analytic thought, focusing on salient objects independent of context. Asians, tend to favor context-dependent, holistic thought, attending to the relationship between the object and its context. These studies inspired our analyses and guided the interpretation of our results. Specifically, we applied Nisbett et al.'s holistic vs. analytic distinction to create a novel tag categorization scheme.

Cultural differences in online communities As computing and communication technologies have spread around the world, researchers have studied cultural differences in technology use. (Chau et al. 2002) found that consumers from different ethnic origins use the Web for different purposes and formed different impressions of the same Web sites. (Lee, Joshi, and McIvor 2007) studied how culture influences online consumer satisfaction. (Kayam, Fussell, and Setlock 2006) found significant differences in the use of instant messaging between Asians and North Americans: Asians used multi-party chat, emoticons and audio-video chatting more than Americans. These results were consistent with general distinctions between Western (individualistic, low-context) culture and Eastern (collectivistic, high-context) culture. (Pfeil, Zaphiris, and Ang 2006) found differences in edits to Wikipedia among people from different cultures. (Wang, Fussell, and Setlock 2009) found that working in a mixed-culture group led Chinese but not American participants to adapt their communication style, and (Wang, Fussell, and Cosley 2011) found that cultural diversity can be used to enhance task outcomes through the use of appropriate technological mediation. Our work provides specific new insights about cultural differences in tagging systems, a ubiquitous feature of social online communities.

Culture and tagging (Sen et al. 2006) investigated factors that influence how users choose to apply tags, including their personal preferences and tags applied by other members of the community that they observe. As detailed below, we build on this work to account for cultural differences.

A few researchers have studied tagging in cross-cultural settings. (Dong and Fu 2010) found order differences in how Chinese and Americans tagged images. Americans tended to tag the main objects in an image first, and background objects and overall properties later, while Chinese tended to apply tags about the overall properties of the image at first, and the objects in the image later. They attribute these differences to the analytic/holistic distinction. (Peesapati, Wang, and Cosley 2010) studied photo tagging, presenting subjects with photos from their own culture and from other two other cultures. They found that subjects “felt closer” to photos from their own culture and that there were cultural differences in how subjects tagged. Compared to these studies, we conduct a more extensive user study and reveal more specific cultural differences in tagging behavior.

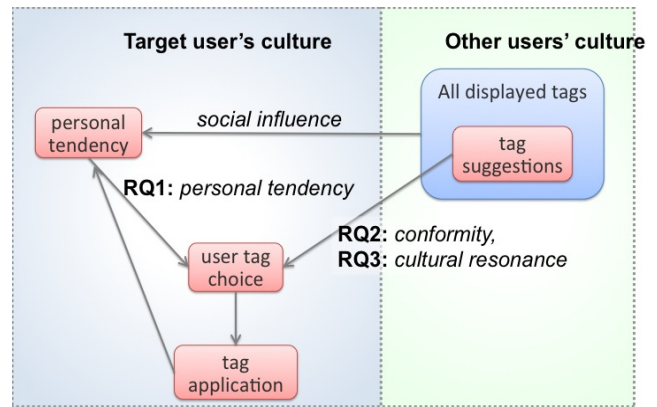


Figure 1: A model of cultural differences in tagging behavior. Users are grounded in a culture’s values, which lead to initial personal tendencies in tagging behavior. Personal tendencies evolve via two forces: social influence through exposure to other users’ tags, and through an investment in one’s own tags. Users may choose tags based on their personal tendencies, or they may choose context-specific tag suggestions offered by the system. Users’ adoption of tag suggestions is mediated by their desire to conform and the consistency of suggested tags with their own beliefs and attitudes (cultural resonance).

(Chen and Tsoi 2011) studied tagging on two different music sites, one popular with Chinese users (SongTaste) and another popular with European and American users (Last.fm). The sets of tags on the two sites exhibited several distinct patterns, including differing in the types of tags applied and the usefulness of the tags. Our work differed significantly in the methods used. We did a controlled study: we knew the cultural and linguistic background of our subjects, all subjects used the same interface, and we carefully controlled contextual information available to them. Chen and Tsoi compared differences between the global sets of tags on two different web sites, with completely different user interfaces and unknown user demographics.

3 Research Framework

To organize our research, we modified the model of (Sen et al. 2006) to account for cultural differences in the tagging process, specifically we study how a person’s culture relates to their *personal tendency*, *conformity*, and *cultural resonance*. Figure 1 shows the model. The model focuses on cultural elements of tagging and does not attempt to include all social and technical aspects of tagging systems. For example, the tags a user chooses are obviously based heavily on the characteristics of a particular item (e.g. a movie may be a *drama*). However, such aspects less related to culture are not included. We next detail the model’s components and motivate the research questions we ask about it.

Users’ culture implicitly guides how they interpret items and apply tags. Clearly, it is simplistic to assume that a person has a single culture: what about someone who is born and grows up in India, studies in the United States, then works in Germany? However, as Hofstede et al. note, “...it is

immensely easier to obtain data for nations than for organic homogeneous societies (Hofstede, Hofstede, and Minkov 2010).” Therefore, to recruit clear experimental groups, we selected members of two distinct cultures and language groups: native speakers of Mandarin Chinese living in China and native speakers of English living in the United States.

Personal tendency is a tagger’s set of preferences and beliefs about tags and their use (Sen et al. 2006). Users bring a pre-existing personal tendency, which affects the number and type of tags they choose to apply, and which is partially determined by their cultural background (Chau et al. 2002). Over time, their personal tendency evolves as they interact with the items and tags in a system. Our first research question studies how culture affects this process:

RQ1: Personal-Tendency: Do users from different cultures exhibit different tagging tendencies?

Hofstede et al. differentiates between culture’s affect on a person’s *values*, the underlying beliefs and attitudes acquired early in life, and *practices*, which can be externally observed (Hofstede, Hofstede, and Minkov 2010). Since values are difficult or impossible to measure accurately, our study focused on differences in tagging practices.

Tag Suggestions Prior work showed users create tags similar to those they have viewed (Sen et al. 2006; Suchanek, Vojnovic, and Gunawardena 2008). We study cultural differences in two mechanisms affecting a user’s adoption of suggested tags: conformity and cultural resonance.

Conformity One reason users apply suggested tags is the psychological tendency towards conformity (Allen and Porter 1983). Research shows that people from different cultures have different degrees of conformity (Kim and Markus 1999; Markus and Kitayama 1991; Nisbett et al. 2001). We thus pose a second research question studying the relationship between conformity and culture:

RQ2: Conformity: Do users from different cultures vary in their degree of conformity to suggested tags?

Cultural Resonance. The tags suggested by a system are applied by people from a specific culture and may therefore reflect the culture of the original tagger. A user may be more likely to select tags that reflect their own cultural beliefs and attitudes, even if they are not aware the tag came from their own culture. We call this effect *cultural resonance*, and our third research question studies it:

RQ3: Resonance: Are people from one culture more likely to adopt tags from their own or another culture?

4 Experimental design

To study our research questions, we designed a survey in which subjects were shown a sequence of movies and asked to apply tags to each of the movies. In some conditions, subjects were shown a set of suggested tags for each movie (the common practice in online tagging communities), while in a control condition, no suggested tags were presented.



Figure 2: Tagging Interface. Users may apply tags by typing them or (in appropriate conditions) clicking a suggested tag.

4.1 Experimental Conditions

We refined our basic survey idea into a 2 x 4 between-subjects design.

The first factor, *subject culture*, had two values: American and Chinese. As mentioned earlier, we define cultural groups based on a subject’s nationality. We selected these nationalities for several reasons. First, they are historically quite distinct. Second, the dominant languages in each culture, English and Mandarin Chinese, are very different. Third, our research team included native speakers of English and Chinese (the Chinese speakers also speak English), which greatly facilitated our research.

The second factor, *tag suggestion type*, had four values:

- **None.** No suggested tags were presented to subjects.
- **Own.** When subjects were presented a movie, they were shown the 8 most popular tags for that movie from a movie web site from their own culture and language. We used MovieLens.org as the American source and Douban.com as the Chinese source. Subjects were not told the source of suggested tags.
- **Other.** For each movie, subjects were shown the 8 most popular tags from the web site of the other culture: Chinese subjects saw tags from MovieLens, and American subjects saw tags from Douban. Tags were translated into the native language of the subjects (see below for details).
- **Combined.** Subjects were shown 4 popular tags from MovieLens and 4 popular tags from Douban. We detail below how we select tags for this condition.

This gives 8 experimental conditions, which we refer to by the subject culture and tag suggestion type, i.e. Chinese-None, American-None, Chinese-Own, American-Own, etc.

4.2 Survey Design

We administered the survey via the Web. Subjects were presented with an initial screen describing the experiment. They then were shown a sequence of 20 screens, one for each of 20 movies. Subjects were asked whether they had seen each movie. They were free to apply as many tags as they wanted to each movie they had seen. After viewing

all the movies, on the final screen they were asked for three optional items, their country, native language, and gender.

4.3 Screen Design / Tagging Interface Design

Since we wanted to study how suggested tags influenced the tags subjects chose to apply, we wanted to minimize other textual content that might influence their choice. Thus, we did not display the type of metadata commonly available for movies, such as the director, actors, plot summary, or reviews. On the other hand, we wanted to present information that would “jog the memory” of our subjects, so they could recall whether they had seen a movie and be able to select appropriate tags.

We thus decided to present the title and five photos for each movie. The five photos showed two posters for the movie, two images associated with important plot elements of the movie, and one image showing the main actor(s). We located photos using IMDB and Google Image Search. Figure 2 shows the photos for the movie “Forrest Gump”.

When subjects indicated they had seen a movie, they were shown a screen where they could apply tags (Figure 2). Users in the None condition were only able to type in tags. Users in all other conditions could either type in a tag or click on a suggested tag to apply it. All entered tags were displayed at the bottom of the screen.

4.4 Selecting movies

We selected the 20 movies to include in the survey based on four principles. The first three principles ensured that users would be familiar with many movies. The fourth ensured that we balanced foreign and U.S. movies.

Popularity. The movie should be popular in both China and the United States. We measured popularity by the number of rating in MovieLens and Douban.

Genre. We selected movies across a range of genres.

Year. We selected movies ranging from 1972 to 2010.

Language/Culture. American movies produced in English are popular in many parts of the world, including China. On the other hand, relatively Americans see relatively few movies produced outside the United States in other languages. We wanted to balance American/English and non-American/non-English movies as much as possible. Therefore, while respecting the Popularity principle, we chose 10 American/English movies and 10 non-American/non-English movies (including two Chinese).

4.5 Selecting and translating suggested tags

As explained above, subjects in the tag suggestion conditions were shown 8 suggested tags for each movie, either from their own culture, the other culture, or combined cultures. This is a primary source of cultural influence on subjects’ tagging choices. We also realized that even if a tag is used in both our source sites, Douban and MovieLens, it might be much more popular in one than the other; this too, is a cultural difference. Therefore, we also represented the popularity of each tag as a number in parentheses; for example, Figure 2 shows that “Vietnam” has a popularity of 10 for the movie “Forrest Gump”.

Condition	Subjects	Distinct tags	Tag apps
American-None	26	615	1108
Chinese-None	54	692	1110
American-Own	23	310	1164
Chinese-Own	29	217	1011
American-Other	15	286	755
Chinese-Other	26	140	929
American-Combined	14	198	500
Chinese-Combined	23	159	880

Table 1: Subjects and basic descriptive statistics.

This method raised three issues: (a) how to compute popularity consistently across two different sites, (b) how to build a list of tags from both MovieLens and Douban for the combined condition, and (c) how to translate tags.

Computing popularity The obvious starting point for the popularity of a tag is the number of times it was applied. However, Douban has many more users and therefore many more tag applications than MovieLens. Therefore, we normalized tag popularity as follows. For each chosen movie, we summed the number of applications of the 8 most popular tags. Then for each of the 8 tags, we calculated the percent of the sum the tag accounted for and used that number. The average normalized popularity across each tag position was nearly identical for both cultures (adjusted- $R^2 = 0.99$).

Combining tags from MovieLens and Douban In the Combined conditions, we presented users with 4 popular tags from the MovieLens and 4 popular tags from Douban. We began by selecting the 4 most popular tags from each source; if there were no common tags in the two sets (after translation), we had our 8 tags. If there were any common tags, we proceeded as follows. If there was a single common tag, this left us with only 7 unique popular tags. We therefore selected the 5th had been selected, we continued down the list of popular tags from the site until we found one that had not been selected. If there were two common tags, we did this procedure for each of the sites. And if there were more than two common tags, we continued this procedure in the obvious way. Finally, we computed a tag’s combined normalized popularity score as the average of the normalized popularity scores of the tag on the two sites. About 18% of the suggested tags in the combined group were common tags chosen in this manner.

Translating tags The Other and Combined conditions required us to show tags produced in Chinese to American subjects and tags produced in English to Chinese subjects. The bilingual members of our research team did the translation in collaboration with a bilingual professor of Chinese Literature at our university and a bilingual graduate student who studies Chinese film.

4.6 Subjects and recruitment

We created two versions of our survey, one in English and one in Chinese. Subjects in each language version were randomly assigned to one of the tag suggestion type conditions. We solicited participants through posts to relevant email lists and social network sites such as Twitter, Facebook, Douban,

and Renren. The survey was live from March 8 through April 18 of 2011.

499 people filled out at least part of the survey. However, to test our conditions clearly, we took a conservative approach and analyzed data only for subjects who (a) made it to the end of the survey, and (b) took the Chinese survey and indicated their country was China and native language was Chinese, or took the English survey and indicated their country was the United States and native language was English. This resulted in a total of 78 American subjects and 132 Chinese subjects. Both Chinese and American subjects added tags to an average of about 10 movies ($\mu=9.8$, $\mu=10.6$ respectively). Subjects also tagged similar numbers of American movies ($\mu=6.3$, $\mu=6.7$) and foreign movies ($\mu=3.4$, $\mu=4.9$). Table 1 shows the number of subjects in each condition and basic descriptive statistics.

A limitation of our survey is the possible selection-bias differences between Chinese and American subjects. Characteristics such as age and tagging experience level may differ between subjects in each culture. These differences may influence tagging behavior or overall reaction to online experiments in each subject group. However, the use of similar recruitment channels between cultures should help mitigate these effects.

5 RQ1: Personal-Tendency

Do users from different cultures exhibit different personal tagging tendencies?

We measured two outcomes of taggers' personal tagging tendencies that contribute to the usefulness of a tag repository and that might differ across cultures: the *number* (tag volume) and *type* (tag class) of tags they apply. We computed the tag class metric only for users in the None conditions since the presence of suggested tags in the other conditions would have confounded the results.

5.1 Tag Volume

Prior research found that Western users applied more tags to describe images and music than Chinese users (Dong and Fu 2010; Kayan, Fussell, and Setlock 2006; Peesapati, Wang, and Cosley 2010). We found similar results: in each tag suggestion type condition (None, Own, Other, and Combined), American subjects applied significantly more tags per movie than their Chinese counterparts¹ (Table 2).

5.2 Tag Class

People from different cultures perceive and describe items in different ways (Dong and Fu 2010; Kayan, Fussell, and Setlock 2006). To understand these differences, we classified tags with two classification schemes. The Factual, Subjective, Personal scheme (FSP) (Sen et al. 2006) is an established tag description scheme. We developed the Holistic, Analytic, Genre (HAG) scheme based on social science research describing cultural differences between Eastern and Western thought.

¹We excluded as outliers subjects whose average number of tags/movie was more than 5 times the overall average. Two users (both American-None) were outliers under this criterion.

Tag suggestion type	Avg num tags per user per movie		p-value
	Americans	Chinese	
None	3.27	1.96	< 0.01
Own	4.85	3.19	< 0.01
Other	4.37	3.05	< 0.01
Combined	4.58	3.20	< 0.01

Table 2: Comparing the number of tags Applied by American and Chinese subjects in each tag suggestion type condition.

	Americans	Chinese	p-value
Factual tags	77.78%	51.20%	< 0.0001
Subjective tags	21.47%	47.67%	< 0.0001
Personal tags	0.76%	0.80%	0.95

Table 3: How Chinese and American subjects differed in their use of Factual, Subjective, and Personal tags.

FSP Classification Schema The FSP scheme was defined by (Sen et al. 2006). *Factual* tags identify objective properties of an item, *Subjective* tags express a user's opinions related to items, and *Personal* tags organize a user's own items. We manually classified each distinct tag into one of the three classes. Tags that do not fit the FSP scheme are classified as other (unless mentioned otherwise, we ignore "other" tags in our research).

Four researchers did the classification, which was divided into two stages. In stage one, two native English speakers classified the English tags from Group 1, and two native Chinese speakers classified the Chinese tags from Group 2. Inter-rater agreement metrics in stage one indicated that coders categorized tags reliably: Chinese and American coders agreed on 84% and 90% of their respective tag classifications. Cohen's kappa values also showed reliable coding (Chinese = 0.80, English = 0.63) (Landis and Koch 1977). In stage two, the coders for each language discussed and tried to resolve any disagreements. If the coders could not agree on a tag's classification, it was not included in our analysis. After stage two, English and Chinese coders reached consensus for 97% and 96% of their respective tags.

Table 3 shows the average proportion of tags in each class for the American-None and Chinese-None groups. Chinese subjects preferred subjective tags, and American subjects preferred factual tags. Figure 3 details these differences by showing the proportion of subjective tags for different users in each culture. Most Chinese users apply similar numbers of factual and subjective tags, and the overall distribution looks relatively normal. However, the distribution of American users looks very different ($p < .00001$, χ^2 test). The 26 American users divide into two clear groups: those who use 80% or more factual tags (22 users), and those who use 80% or more subjective tags (4 users). The graph for factual tags shows the same pattern, but reversed.

As far as we know, (Kayan, Fussell, and Setlock 2006) are the only other researchers to use the FSP classification to analyze cultural differences in tagging. They found that a larger proportion of factual tags in a site popular with Chi-

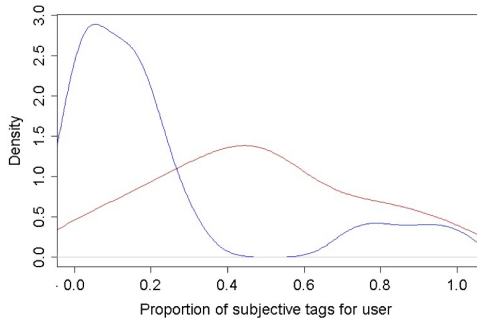


Figure 3: Proportion of subjective tags for different users of Chinese-None (one peak) and American-None (two-peaks).

nese users (SongTaste) and a larger proportion of subjective tags in a site popular with Europeans and Americans (Last.fm). These results appear to be contrary to ours. However, as we described earlier, major differences in study methodologies mean that the results are not directly comparable. For example, user interface differences between SongTaste and Last.fm and unclear user demographics of the two sites may account for the differences.

HAG Classification Scheme We reviewed Nisbett’s work (see above) on holistic vs. analytic thought to create the HAG categorization scheme:

- *Holistic* tags describe a movie as whole, relationships between objects in a movie, or context-dependent aspects of the objects in a movie.
- *Analytic* tags describe specific parts or aspects of a movie.

We first had to define criteria for classifying tags as Holistic or Analytic. One English coder and one Chinese coder worked together to create the criteria. Their work revealed a problem: it was difficult to classify genre tags like *drama*, *horror*, or *comedy* since they have both Holistic and Analytic aspects. Like Holistic tags, genre tags describe an item as a whole; like analytic tags, genre tags describe attributes of an item. Rather than force genre tags into the Holistic or Analytic category, we defined a new category:

- *Genre* tags describe a movie’s genre. We used genres from IMDB (www.imdb.org), which is popular in both China and the U.S.

We classified tags using the HAG schema similarly to the FSP schema. One key difference is that we found it necessary to classify (tag, movie) pairs in the HAG schema, because the same tag may be used in different senses for different movies. For example, the tag “war” was applied to both “Braveheart” and “Forrest Gump” in our survey. We judged the former a Genre tag, because “Braveheart” is a war movie. However, we judged the latter an Analytic tag, because while one portion of “Forrest Gump” relates to the Vietnam War, it is not a war movie.

As with FSP, coders were reliable in stage one codings according to both percentage agreement (Chinese = 89%, English = 84%) and Cohen’s kappa (Landis and Koch 1977) (Chinese = 0.82, English = 0.69). After stage two, English

Holistic	Analytic
Epic (7)	Holocaust (11)
classic (7)	prison (8)
historical (5)	fish (7)
surreal (5)	Pixar (7)
visual-treat (3)	Leonardo DiCaprio (6)
romantic (2)	French (5)
multiple storylines (1)	Australia (3)

Table 4: Examples of holistic and analytic tags.

	Americans	Chinese	p-value
Holistic	27.94%	56.01%	< 0.0001
Analytic	52.94%	35.94%	< 0.0001
Genre	19.02%	8.05%	0.01

Table 5: How Chinese and American subjects differed in their use of Holistic vs. Analytic vs. Genre tags.

and Chinese coders reached consensus on categorizations for 98% and 99% of their respective (tag, movie) pairs.

Table 4 gives a few examples of tags that we classified as Analytic and tags we classified as Holistic. Recall that we actually classify (tag, movie) pairs; however, for simplicity of exposition, Table 4 shows tags that were classified consistently as Holistic or consistently as Analytic.

Table 5 shows the average proportion of tags in each class for the American-None and Chinese-None groups. Chinese subjects applied more Holistic tags, while American users were more likely to use Analytic tags and Genre tags. The proportion of analytic tags for each Chinese and American tagger (Figure 4) resembles the distributions for subjective tags (Figure 3). Most Chinese users apply a balance of Holistic and Analytic tags. The American users split into two groups. The same four users account for the right-hand peaks in Figures 3 and 4. This reflects the overall tendency of holistic tags to be subjective.

The differences in Holistic and Analytic usage are consistent with Nisbett (Nisbett et al. 2001). The higher usage of Genre tags by Americans reflects Nisbett’s finding that Western cultures prefer to use rules and categorization methods to organize objects.

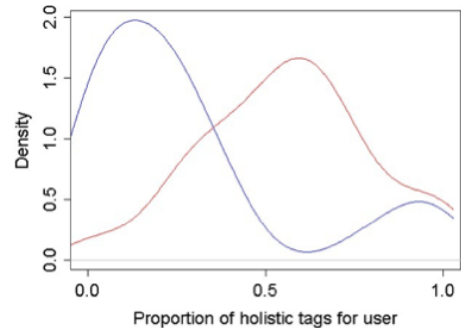


Figure 4: Proportion of Analytic tags for different users of Chinese-None (one peak) and American-None (two-peaks).

Tag suggestion type	Percent conformity	
	Americans	Chinese
Own	81.5%	96.7%
Other	69.2%	91.0%
Combined	83.4%	95.6%

Table 6: Comparing the conformity of American and Chinese subjects in each tag suggestion type condition.

Tag suggestion type	Percent “entirely conformant” users		p-value
	Americans	Chinese	
Own	22%	73%	0.001
Other	0%	55%	0.001
Combined	29%	74%	0.02

Table 7: Percent of subjects who only applied suggested tags in each experimental condition. P-values are calculated using a χ^2 test.

6 RQ2: Conformity

Do users from different cultures vary in their degree of conformity to suggested tags?

Conformity in tagging systems can be measured by the extent to which users apply suggested tags. We measure conformity using *percent-conformity*, defined as follows:

For a user u , consider a movie m that u tagged. Calculate the percent of tags u applied to m that were suggested by the system. Average this value for all the movies tagged by u to get the *percent-conformity* for u . Finally, to compare experimental conditions, average this value for all users in each condition.

We compare the conformity of American and Chinese subjects for the three tag suggestion type conditions: (a) Own: users are suggested tags from their own culture, (b) Other: users are suggested tags from the other culture, and (c) Combined: users are suggested tags from both cultures.

As Table 6 shows, Chinese subjects exhibited stronger conformity than American users: in each of the three tag suggestion type conditions, Chinese subjects applied more suggested tags than Americans.² Some users were “entirely conformant”; they only applied tags suggested by the system. Table 7 shows a dramatic cultural difference in entirely conformant users: while a majority of Chinese users only apply suggested tags, relatively few Americans did so ($p < 0.02$, χ^2 test). These results are consistent with Western and Asian emphasis on individualism and collectivism.

We also wondered whether users from one culture were more likely to choose popular tags. To analyze this, we measured the Pearson correlation between the displayed popularity of a tag and the likelihood of a user to adopt the tag. Although we found stronger correlations with tags from a user’s own culture (own = 0.66, other = 0.42), we did not find significant differences between cultures.

²We do not report significance because the users who only applied suggested tags led to a right-skewed distribution.

Subject Culture	Adoption of:		p-value
	Own tags	Other tags	
American	40.4%	23.7%	< 0.01
Chinese	47.5%	24.4%	< 0.01

Table 8: Cultural Resonance: Subjects who got tag suggestions originating from both cultures adopted more suggested tags from their own culture.

7 RQ3: Resonance

Are people from one culture more likely to adopt tags from their own or another culture?

We investigated this question in two ways: (a) a between-subjects comparison for the Own and Other groups in each culture, and (b) a within-subjects comparison for the Combined group in each culture.

Own vs. Other We again use percent-conformity as our metric. Table 6 shows that for both cultures, subjects who were exposed to tags that originated from their own culture adopted more suggested tags than those who were exposed to tags that originated from the other culture.

Note that 6 also illustrates the greater influence of conformity on Chinese subjects’ tagging choices³: they adopted a higher proportion of suggested tags from an American source than Americans did!

Combined As described above, the Combined conditions use suggested tags from both American and Chinese sources, with all tags translated into subjects’ native language. Therefore, for the combined conditions, we compared the proportion of applied tags that were tag suggestions from subjects’ own culture and the proportion of applied tags that were suggestions from the other culture (we ignore suggestions that were common to both cultures).

Table 7 shows that both American- and Chinese- Combined subjects adopted more suggested tags that originated from their own culture even though they did not know the cultural origin of suggested tags.

8 Discussion

Our results suggest implications for designers of tagging systems and for future research. To situate the implications, we recall a primary goal for a tagging system: to collect a useful repository of tags. To make the goal more focused, we also must specify useful *for what*. As (Sen et al. 2006) found, different classes of tags (in the FSP scheme) are useful for different purposes; for example, factual tags are most useful for finding items, while subjective tags are most helpful for evaluating items. We should also specify useful *for whom*. Our results suggest that users from different cultures may have different preferences regarding tags. Thus, we can restate the goal as: to collect a repository of tags that support the users of a system in performing relevant tasks.

System designers can intervene to achieve this goal at two points: tag application time and task execution time. Systems can shape tag applications by carefully selecting dis-

³And likewise the greater influence of individualism on American subjects’ choices.

played and suggested tags (this work and (Sen et al. 2006)). Likewise, systems may optimize tasks by designing intelligent algorithms and interfaces to be culture-aware.

With this context we can state several design guidelines and several areas for future work. First, designers should consider what tagging vocabulary “deficits” might result from users’ cultural tendencies. For example, a system with mainly Chinese users may produce a low proportion of factual tags, which may reduce the effectiveness of tag searches (Sen et al. 2006). In response, a designer might encourage factual tags by favoring them as suggested tags or by offering “reputation points” for applying them. Second, designers should consider when to suggest tags from the target user’s own culture. For example, a system may increase tag adoption if it suggests Chinese tags to Chinese users. These two guidelines highlight an interesting tension for system designers. The first focuses on improving the overall tagging vocabulary, and the second focuses on matching the natural tendencies of individual users. Supporting users’ natural tendencies may increase their short-term satisfaction, but produce a problematic overall tagging vocabulary.

We propose several areas of future work to establish a firm basis for these guidelines, we. First, while we showed that people from two cultures can apply different tags, we also would like to see whether the tags are useful for different specific tasks. Our intuition is that prior results concerning tag usefulness (Sen et al. 2006) generally will hold for non-Westerners because the results are based on inherent properties of the tasks. For example, factual tags may be useful in finding items because people agree on them and apply them consistently. However, this should be verified. Second, there are no data about the relative usefulness of Holistic, Analytic, Genre tags for the key tagging tasks. This too should be studied for users from various cultures. Third, existing research exploring incentive mechanisms that influence the amount and type of work people do in online communities has focused on American contexts. Since these mechanisms offer potential for eliciting useful tags, their effectiveness for other cultures should be explored.

9 Acknowledgements

We would like to thank Joseph Allen, Jessica Ka Yee Chan, and Brandon Maus for the help in coding tags. We are also grateful for the support of the members of GroupLens Research, particularly Jesse Vig and Daniel Kluver who provided feedback on the study design. This work is funded by the NSF (grants IIS 09-64697, IIS 09-64695 IIS 08-08692) and by the China Scholarship Council.

References

Allen, R., and Porter, L. 1983. *Organizational influence processes*. Scott, Foresman.

Chau, P.; Cole, M.; Massey, A.; Montoya-Weiss, M.; and O’Keefe, R. 2002. Cultural differences in the online behavior of consumers. *Communications of the ACM* 45(10):138–143.

Chen, L., and Tsoi, H. K. 2011. Analysis of user tags in social music sites: Implications for cultural differences. In

Extended Abstracts of ACM Conference on Computer Supported Cooperative Work (CSCW’11). ACM.

Dong, W., and Fu, W. 2010. Cultural difference in image tagging. In *Proceedings of the 28th international conference on Human factors in computing systems*, 981–984. ACM.

Hofstede, G.; Hofstede, G.; and Minkov, M. 2010. *Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival*. McGraw-Hill Professional.

Kayan, S.; Fussell, S.; and Setlock, L. 2006. Cultural differences in the use of instant messaging in asia and north america. In *Proceedings of the 2006 conference on Computer supported cooperative work*, 525–528. ACM.

Kim, H., and Markus, H. 1999. Deviance or uniqueness, harmony or conformity? a cultural analysis. *Journal of Personality and Social Psychology* 77(4):785.

Landis, J., and Koch, G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 159–174.

Lee, K.; Joshi, K.; and McIvor, R. 2007. Understanding multicultural differences in online satisfaction. In *Proceedings of the 2007 ACM SIGMIS CPR conference on Computer personnel research*, 209–212. ACM.

Markus, H., and Kitayama, S. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review* 98(2):224.

Nisbett, R.; Peng, K.; Choi, I.; and Norenzayan, A. 2001. Culture and systems of thought: Holistic versus analytic cognition. *Psychological review* 108(2):291.

Peesapati, S.; Wang, H.; and Cosley, D. 2010. Intercultural human-photo encounters: how cultural similarity affects perceiving and tagging photographs. In *Proceedings of the 3rd international conference on Intercultural collaboration*, 203–206. ACM.

Pfeil, U.; Zaphiris, P.; and Ang, C. 2006. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication* 12(1):88–113.

Sen, S.; Lam, S.; Rashid, A.; Cosley, D.; Frankowski, D.; Osterhouse, J.; Harper, F.; and Riedl, J. 2006. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 conference on Computer supported cooperative work*, 181–190. ACM.

Suchanek, F.; Vojnovic, M.; and Gunawardena, D. 2008. Social tags: meaning and suggestions. In *Proceeding of the 17th ACM conference on Information and knowledge management*, 223–232. ACM.

Wang, H.; Fussell, S.; and Cosley, D. 2011. From diversity to creativity: Stimulating group brainstorming with cultural differences and conversationally-retrieved pictures. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 265–274. ACM.

Wang, H.; Fussell, S.; and Setlock, L. 2009. Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. In *Proceedings of the 27th international conference on Human factors in computing systems*, 669–678. ACM.