Examining the Impact of Collaborative Tagging on Sensemaking in Nutrition Management

Lena Mamykina[1], Andrew D. Miller[2], Yevgeniy Medynskiy[2], Catherine Grevet[2], Patricia R. Davidson[3], Michael A. Terry[4], Elizabeth D. Mynatt[2]

[1] Department of Biomedical Informatics, Columbia University 622 West 168th St. VC5 New York, NY 10032

[2] GVU Center, Georgia Institute of Technology 85 5th Street NW, Atlanta, GA 30332

[3] The Cardiovascular Care Group, Belleville, NJ, Saint Claires Hospital, Dover, NJ

[4] David R. Cheriton School of Computer Science University of Waterloo Waterloo, Canada

ABSTRACT

Collaborative tagging mechanisms are integral to social computing applications in a variety of domains. Their expected benefits include simplified retrieval of digital content, as well as enhanced ability of a community to makes sense of the shared content. We examine the impact of collaborative tagging in context of nutrition management. In a controlled experiment we asked individuals to assess the nutritional value of meals based on photographic images and observed the impact of different types of tags and tagging mechanisms on individuals nutritional sensemaking. The results of the study show that tags enhance individuals' ability to remember the viewed meals. However, we found that some types of tags can be detrimental to sensemaking, rather than supporting it. These findings stress the importance of tagging vocabularies and suggest a need for expert moderation of community sensemaking.

ACM Classification Keywords: H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.
Author Keywords: Collaborative tagging, collective sensemaking, nutrition, wellness
General Terms: Experimentation

INTRODUCTION

In many areas of human enterprise, individuals rely on the collective wisdom of their peers rather than on expert advice alone. In the domain of health and wellness management, social applications such as TuDiabetes [1], Patients Like Me [2], and Eat.ly [3] help thousands of individuals share their experiences and receive advice and guidance from others. Compensating for the limited access to experts, these distributed communities help individuals fill gaps in their knowledge and understanding, and to contribute in kind.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

One mechanism integral to these knowledge-sharing applications is *tagging*. Tagging allows individuals to assign free-form keywords to digitally stored content. For example, in the field of nutrition management, tags can be assigned to pictures of foods as part of a nutrition management program. The process of *collaborative tagging* allows individuals to share their tags with others, thus forming reusable, communal *tagging vocabularies*.

Tagging is generally thought to provide two types of benefits to users. At the most basic level, assigning meaningful keywords to digital data simplifies its retrieval. However, early advocates of collaborative tagging argued that tagging also promotes collaborative sensemaking and consensus building [21]. Sensemaking is a process through which individuals comprehend meaning of events and situations and decide on their course of action. When individuals tag items or events they assign them explicit labels and classify them into categories; both of these are essential steps of making sense of the world [23]. In collaborative tagging, explicit labeling allows individuals to compare their emergent understanding, and reach a consensus, thus enabling collective sensemaking.

In the context of health and wellness communities, these latter benefits would be highly desirable: They could enable development of a richer understanding of individuals' particular health concerns and help them create a shared vocabulary for health issues. Both of these effects could greatly enhance individual health management practices. In nutrition management, tags could highlight nutritional properties of meals, such as "good portion size," "poor proportions," or even "heart attack on a plate." Seeing these tags assigned to pictures of meals could help less experienced members adopt a shared operational vocabulary. Comparing tag assignments across pictures could also help individuals refine their own nutritional judgment, potentially leading to more informed choices in the future.

Despite these expectations, empirical studies of tagging in the context of text recognition and recall in educational settings have not demonstrated the expected benefits [4]. Thus, questions remain as to the potential benefits of

tagging and whether tagging can facilitate comprehension, sensemaking, and consensus building.

In this paper we present results from an investigation of collaborative tagging mechanisms and their impact on nutritional judgment. Our primary research questions include the following:

1. How do individuals use social tagging mechanisms in context of nutrition management? What types of tags do they generate spontaneously?
2. Does assigning tags to meals affect individuals' sensemaking in the context of nutrition management? Are certain types of tags and tagging mechanisms more beneficial than others? Specifically, does tagging improve individuals' ability to remember meals and/or their ability to critically assess their nutritional value?
3. Does a socially (or otherwise) created tagging vocabulary have an impact on individuals' own tagging practices? How sustained is that impact?

To answer these questions, we conducted a controlled experiment. Participants were asked to view pictures of meals displayed on a computer screen and tag these meals using different tagging strategies, then answer a set of questions about the meals. Using a mixed-factors repeated measures design we assessed the general impact of tags, and differentiated between different tagging strategies, and types of tags.

Our experiment suggests that tagging helps individuals remember tagged content. However, we found that the *type* of tags suggested in a tagging vocabulary has a significant impact on individuals' ability to critically assess the nutritional value of these meals. In particular, *descriptive tags* that simply listed meals' ingredients (such as "meat", "potatoes") were actually detrimental to the individuals' ability to assess the meals. In contrast, *assessment tags* that focused participants' attention on nutritional properties of meals (such as "poor proportions" and "lacks vegetables") led to a slight improvement in their assessment abilities. Further, we found that individuals are more likely to spontaneously generate descriptive tags, rather than assessment tags. However, exposure to a tagging vocabulary containing assessment tags encouraged adoption of these more beneficial tagging strategies in individuals' own practice.

The results of this study have direct implications for the design of tagging mechanisms in general and for wellness and nutrition management sites in particular. For example, most of these sites allow individuals to collectively create tagging vocabularies without any supervision or moderation. Our study shows, however, that if these social vocabularies accrue only descriptive tags, they may inhibit sensemaking rather than promote it. To avoid this, such sites might consider interjecting expert-created assessment tags into the social vocabularies.

In the rest of this paper we discuss related work in the area of social tagging, enumerate our specific research questions, then present our experimental study, our findings, and their implications for the HCI community.

BACKGROUND AND RELATED WORK

Tagging and Collaborative Tagging

Collaborative tagging has become a popular mechanism for social web-based services. Examples of early communities that embraced tagging include Flikr, Del.icio.us, CiteULike, among many others. Tagging primarily allows individuals to assign free-form keywords to digital content, such as images or URLs.

One of the main advantages of collaborative tagging systems is to allow their users to converge on a shared meaning of the content that they tag. When multiple people use identical tags for items, aggregation of such tags leads to the emergence of a user-created categorization scheme, known as a "folksonomy"[22].

Tags present an alternative to the traditional method of an expert or group of experts developing a categorization scheme. Past research argues that expert-created taxonomies are not capable of responding to the changing understanding of the world, and result in loss of data when items in-between established categories [21]. In contrast, social tagging is seen as flexible and fluid, continuously responding to the changing understanding of the larger community. In addition, it allows for the emergence of multi-layer taxonomies, from more general categories to very low-level idiosyncratic ones, with no loss of data. However, there is some concern that folksonomies could reinforce misconceptions and stereotypes, and produce herding behavior [20].

Several researchers attempted to map out a theoretical model as well as design space for collaborative tagging systems. Marlow et al [15] present a three-tiered conceptual model, which includes such elements as resources, users, and tags. In the model, tags serve as connections between users and resources. The authors identify a number of design dimensions for social tagging mechanisms, including, for example, whether individuals are provided with tagging suggestions or whether they are asked to tag blindly, without access to tags by other people. The experimental conditions in our study were generally inspired by these variations in tagging mechanisms.

Encoding and Retrieval

While earlier writings on collaborative tagging examined their ability to foster collective sense-making, much of the recent literature focuses on these systems' capacity to simplify access to and retrieval of digital content. For example, Chi and Mytkovic [5] suggest that social tagging is "fundamentally a method for organizing objects for later use. It is a process of *encoding* objects with keywords so as to later *retrieve* those very same documents" (emphasis by authors). Within this perspective, researchers have focused on mathematically describing the dynamics of users'

tagging behaviors, and on the evolution of tags and tagging vocabularies. In this research, the phrase "social bookmarking" often replaces "collaborative tagging," further stressing this work's focus on retrieval.

A common conclusion of this line of research (for example Golder and Huberman [7], and Halpin et al [8]) is the stabilization of tagging systems over time, essentially supporting the notion of consensus building. However, this research primarily focuses on the extrinsic value of the community-developed tagging vocabularies, namely their benefits for retrieval of content. What is not considered is how tags and the act of tagging affect their creators.

Sensemaking

Drawing inspiration from Karl Weick [23], we define sensemaking as a "process of transforming circumstances into a situation that is comprehended explicitly in words and that serves as a springboard to action". Weick includes bracketing (or classification) and labeling (or assigning explicit names) as two essential steps in the sensemaking process.

In nutrition sensemaking, individuals assess properties of a meal in the context of their existing knowledge, their nutritional goals, and their desires of the moment, and explicitly express their comprehension of the situation in words. This comprehension leads to their choice of action. On social networking sites dedicated to nutrition management (e.g., eat.ly), digital images of meals serve as a proxy of real meals. Users tag these meals, thus expressing their comprehension of the meal's value in words. In this scenario tagging becomes a mechanism for bracketing and labeling, and thus for sensemaking.

This view is consistent with other researchers' assessment of the potential of social tagging. For example, Golder and Huberman (2006) [7] assert, "Tagging is fundamentally about sensemaking. Sensemaking is a process in which information is categorized and labeled, and, critically, through which meaning emerges." They discuss the benefits of collaborative tagging in helping communities organize knowledge and coordinate action. Similarly, Pauen et al view tagging as sensemaking aid: "The act of assigning tags to categorize an object is an act of knowledge production as it makes apparent the mental models, or internal representations of knowledge, that one uses to associate with the object" [18].

Several researchers have explored the potential of social tagging to help communities of learners. For example, Yew et al [24] studied students taking an undergraduate Business Information Technology class. As part of class curriculum, the students were required to compose and tag blog posts discussing topics relevant to the class. The authors found a strong correlation between blogging frequency and higher overall grades. However they do not report on the impact of tags.

Oleksik et. al. [17] looked at how tagging supports information and activity management practices. They discuss four main tag usage scenarios: managing short-term/transient activities, revisiting resources and resuming activities, filtering resources and creating meta-organization of resources. Of these uses, the last one in particular could be seen as evidence for individuals use of tags to simplify retrieval, but also as sensemaking tools that helped them see patterns in their digital resources.

Hsieh et al examined the use of tags for synchronous collaboration in a workspace [9]. In their plugin to Lotus Sametime, users could assign tags to tasks. They found that tags could be used for temporal coordination.

Note Taking

While situated in a different context, research examining note-taking in educational settings is potentially relevant to tagging. Research in this space is extensive (see [10] for a review of the work in this area). Despite the different contexts, there are many parallels between the two activities. In both cases, individuals aim to externalize and capture in textual form their emerging understanding of particular concepts. Both are also relatively unstructured and emerge as part of individuals' evolving experience with the content. Finally, both are expected to aid in comprehension and later information retrieval.

Research in note-taking identifies a set of concepts that help structure an understanding of tagging as a cognitive activity. For example, it makes a distinction between note-taking *process* and its *product* (i.e., the actual notes), both of which are seen to have their benefits. In particular, there is evidence that the process of taking notes leads to a better understanding of the material, even if the notes are never reviewed [12]. Reviewing the notes at a later time also aids comprehension, indicating the value of the actual notes.

Prior research also distinguishes between different note-taking strategies and suggests differences in the benefits of these strategies. Specifically, studies show that *conceptual notes* are more likely to lead to deeper processing and comprehension than *factual notes*. However, students are more likely to spontaneously generate factual notes [11].

In the sections that follow, we adopt these frames of reference when considering the activity of tagging. In particular, we similarly distinguish between the tagging *process* itself (in which individuals are asked to assign tags to content) and the tagging *product* (how existing tagging vocabularies affect individuals' practices and comprehension). We also distinguish between *descriptive tags* and *assessment tags*, which are conceptually similar to the notions of factual notes and conceptual notes.

Lack of Strong Empirical Evidence

The body of work we discussed above establishes a strong theoretical foundation for collaborative tagging as facilitating sensemaking within a community. However,

several recent studies that specifically examined the impact of tags on memory and sensemaking did not produce empirical evidence to support these claims. For example, Budiu *et al* [4] compared individuals' ability to remember part of written text with and without tags, also differentiating between two tagging mechanisms (typing tags, and selecting text as tags). They found that tags led to performance at best comparable with no tagging condition, but generated no improvement beyond it. Similarly, Nelson *et al* [16] investigated the impact of social tagging on making sense of novel information on the web. In their studies, tagging had a positive impact only when individuals used tags created by an expert, rather than tags of their own creation.

This seeming contradiction between a convincing theoretical foundation and the lack of strong empirical evidence calls for further investigations of collaborative tagging mechanisms. In our work we attempted to identify certain properties of collaborative tagging that may explain the lack of expected positive impact. Specifically, we hypothesized that there are substantial differences in the benefits afforded by different types of tags and tagging mechanisms: not all of them may be equally beneficial to comprehension and sensemaking. Our goal was to better understand what those properties are and what design solutions would help to maximize the positive impact.

RESEARCH QUESTIONS

Our focus on the effect of tagging in nutritional management applications led us to the following set of research questions:

1. How do individuals use social tagging mechanisms in context of nutrition management? What types of tags do they generate spontaneously?

Different applications impose different restrictions on the length and structure of tags. Some require single-word tags, others allow more complex compositions. Our first question was what style of tags individuals would choose in context of nutrition management, with no external restrictions. In addition, we were wondering what type of tags individuals would create spontaneously: descriptive or assessment. In agreement with note-taking literature, our expectation was that individuals would be more likely to spontaneously generate descriptive rather than assessment tags.

2. Does assigning tags to meals affect individuals' sensemaking in context of nutrition management?

As we mentioned earlier, sensemaking is a complex process involving a variety of components. In context of nutrition management it could involve an understanding of underlying physiological processes, different ways food affects them, and how to selectively manage diet to achieve one's nutritional goals. A fuller discussion of sensemaking in this context is beyond the scope of this paper. Here, we set to examine simpler components of sensemaking, including the following:

2.1 Does assigning tags to meals improve individuals' ability to remember these meals?

2.2 Does tagging lead to more accurate critical evaluation of food's nutritional value (as compared to the ground truth provided by a trained dietician)?

2.3 Is there a difference in the type and magnitude of impact for different tagging mechanisms (unassisted tagging and supported tagging), and different types of tags (descriptive and assessment)?

Our expectations were that collaborative tagging would positively affect both memory and evaluation. Furthermore, just as there is a difference between the effects of factual and conceptual note-taking practices, we expected that individuals who apply assessment tags would be able to better determine the nutritional value of a meal than those who merely used descriptive tags.

3. Does a socially-created tagging vocabulary have an impact on individuals' own tagging practices? For example, if individuals are presented with a vocabulary containing assessment tags, would they incorporate these into their own practices?

If in fact different types of tags have different effects on individual and collective sensemaking, the question becomes what mechanisms could be used to encourage the more beneficial practices. For example, if assessment tags have a more positive impact on memory and critical evaluation, how can designers and moderators of communities cultivate this approach to tagging? We expected that a pre-existing tagging vocabulary (e.g., one developed by the community) would influence individuals' tagging choices, leading to a convergence in practices.

METHOD

Design

To explore the research questions above, we designed a controlled laboratory experiment. The basic task of participants was to view digital images of meals with the goal of assessing nutritional value.

The study used a mixed-factorial design with repeated measures for within-group factors. The between-group factor corresponded to the type of tagging involved:

1. *Unassisted tagging*, in which individuals created free-form tags by typing them in the text box (Free, N=21),
2. *Supported tagging* with descriptive tags, in which individuals could select tags from a list as well as type them in a text-box; the list included descriptive tags ONLY (Descriptive, N=20),
3. *Supported tagging* with assessment tags, similar to the previous condition, with the list including assessment tags as well as descriptive tags (Assessment, N=17).

The within-group factor included the following conditions:

1. *Baseline*, in which individuals did not use any tagging,
2. *Intervention*, in which individuals were exposed to the 3 tagging mechanisms, depending on their between-group condition, and

3. Retention, in which all individuals were asked to use unsupported tagging to assess the residual effect of the tagging condition on individuals' own tagging approaches

The overall research questions and methods are summarized in Table 1 below:

Research questions	Methods
How do individuals use tagging in context of nutrition management	Classifying tags used by individuals in *Unassisted Tagging* condition during *Intervention* trial
Does assigning tags to meals improve individuals' ability to remember these meals?	Comparing performance on memory test between baseline (no tagging) and intervention trials for different between-subject conditions.
Does tagging lead to more accurate critical evaluation of food's nutritional value?	Comparing performance on assessment and choice tasks between baseline (no tagging) and intervention trials for different between-subject conditions
Is there a difference in the type and magnitude of impact for different tagging mechanisms and different types of tags?	Comparing performance on memory, assessment and choice tasks across different between-subject conditions.
Does a socially-created tagging vocabulary have an impact on individuals' own tagging practices?	Comparing individuals' choice of tags between intervention and retention trials for the two supported tagging conditions.

Table 1: Research questions and methods

Task

The task of participants was to assess overall nutritional value of meals based on their digital images. This type of task is becoming relatively common to the emerging health management applications. For example, several commercial services [3] and research platforms [13] allow individuals to take pictures of their meals, view them online, and share them with their peers or healthcare providers. The general premise of these applications is that viewing images of meals and their assessment by members of the community can increase individuals' nutritional literacy and help them maintain a healthy diet. In this study, we wanted to specifically investigate how these practices could be supported with social tagging mechanisms.

Procedures

Participants were seated in front of a desktop computer. Images of meals from a fictional individual were displayed on a screen one at a time. There was no restriction as to how long individuals could view the images. They were not allowed to use scrap paper or to take notes.

Participants were randomly assigned to one of the 3 conditions: Free Tagging, Descriptive Tagging, or Assessment Tagging. They were also randomly assigned to a particular starting case to avoid ordering effects. After a short, standardized introduction, the participants completed a paper-based demographic questionnaire and began the test. Each participant was asked to complete 3 trials that each included the following:

1. *Stimuli*: Individuals viewed images of meals on a computer screen (Figure 1). Each trial included 9 images that were arranged as 3 days of meals of a fictional person (breakfast, lunch, dinner, shown in that order). The participants were instructed to view the meals and think about how healthy the meals were.
2. *Evaluation*: Individuals were asked to answer a set of questions about the reviewed images; then they were presented with new images of meals (not seen before) and asked to either evaluate how healthy they were, or to choose the best meal in a set.

More specifically, the evaluation process included the following:

1. *Memory-recognition*: Participants viewed 9 images, 5 from the set they had seen before, 4 new ones, and were asked to judge whether they had seen that image before or not (see Figure 4).
2. *Evaluation-memory*: Participants were asked to assess the diet of the fictional individual, based on the 3 days of meals they viewed on a scale 1 to 5 (1–not at all healthy; 5-very healthy). They were also asked to explain their choice using free-form text
3. *Evaluation-new:* Participants viewed 3 new images of meals, one at a time (Figure 5). They were asked to decide how healthy these meals were using the same 1 to 5 scale, and to explain their decision using free-form text.
4. *Choice*: Participants viewed 2 sets of 3 new images (all 3 in a row) to decide which of these meals they would choose as the best option. They were asked to explain their choice using free-form text.

Figure 1: in a *stimuli* part of the test, individuals viewed images of meals and contemplated how healthy these meals were. In the baseline trial the individuals did not use any tagging mechanisms.

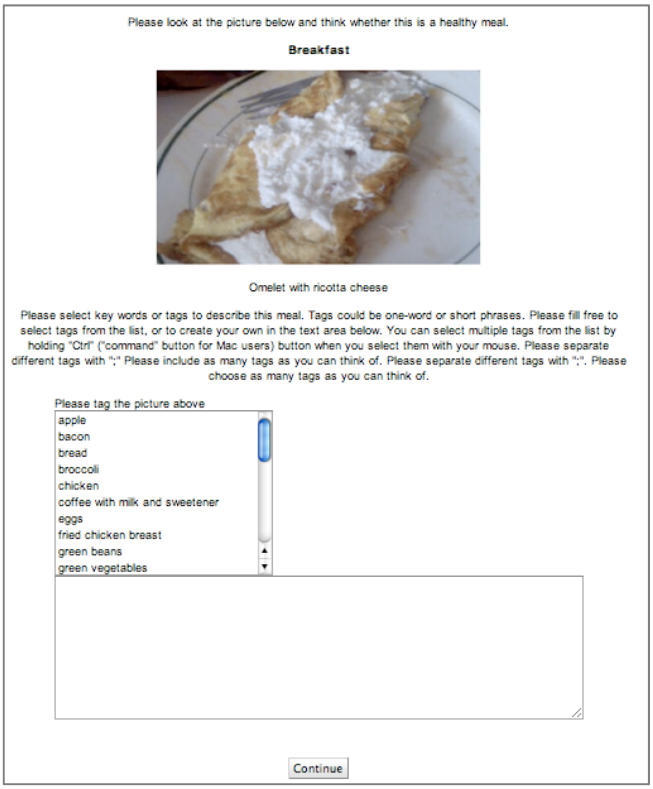

Figure 2: in the Intervention trial, individuals in both Descriptive and Assessment conditions could either enter their own tags or use tagging vocabulary that contained either descriptive tags (for Descriptive condition) or a combination of descriptive and assessment tags (for Assessment condition)

Figure 3: in the Retention trial, all participants used unsupported tags by typing them in the text box.

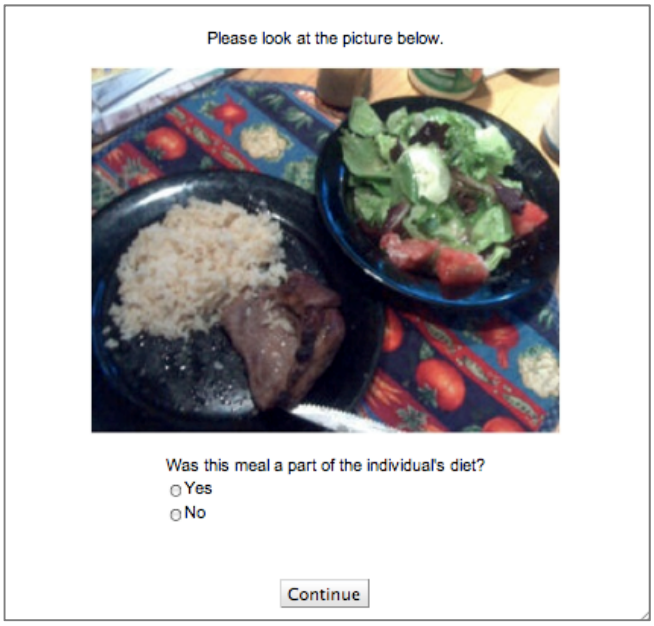

Figure 4: during the memory test, the individuals were asked to recognize whether the displayed meal was a part of the diet they viewed before.

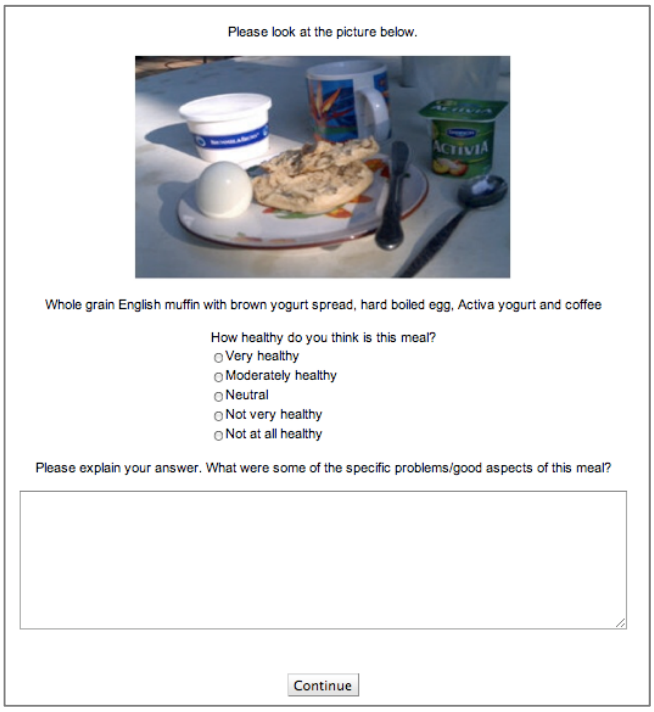

Figure 5: after reviewing the images on meals, the individuals determined how healthy they were on a scale 1-5.

Materials

The images of meals used in the experiment were collected during previous deployment studies of a mobile application for wellness management [14]. This application allowed individuals to use cell phone cameras to capture pictures of their daily meals. As a result, all the images used for the experiment are authentic images of meals consumed by real individuals, rather than stock photography.

The images for the experiment were selected from the overall pool of collected images with the assistance of a diabetes educator/certified dietician. The dietician was instructed to select images that were either clearly healthy or clearly unhealthy to avoid potential ambiguity. The dietician was also asked to rate these images on a scale of 1-5, to provide an explanation of her judgment, and to assign both descriptive and assessment tags to the images.

The results of the expert assessment were used as the ground truth for individuals' assessment tasks. Furthermore, the tags developed by the dietician, consolidated across all images, were used as the tagging vocabulary for the assisted tagging conditions.

In the Descriptive condition, participants saw only descriptive tags. In the Assessment condition, participants saw both descriptive *and* assessment tags. The reason we chose to include both types of tags in the Assessment vocabulary is to increase ecological validity of the test. In a real world situation, replacing socially created vocabulary with the one created by an expert defeats its purpose. However, one could imagine using experts' help to populate the organically created social vocabulary with the more impactful tags.

Analysis

We used the following methods to analyze our data:

1. One-Way ANOVA comparing means of summed scores across repeated measures between Free, Descriptive, and Assessment conditions for each of the following: memory, assessment, and choice. This test showed us initial difference (or lack thereof) between our three groups to ensure that our randomization was successful and that there were no systematic differences between groups.

2. Repeated Measures General Linear Model test for Group*Trial to compare the differences in means between different groups (Free, Descriptive, Assessment) across different trials (Baseline, Intervention, Retention) for each of memory, assessment, and choice. This test allowed for assessment of the impact of each of the tagging mechanisms, and allowed us to compare the differences between them.

We also collected a substantial amount of qualitative data, including tags that individuals either selected or created and their explanations for their choices, written in free-form text. We incorporated these data in our analysis using the following method.

Tags were manually coded according to whether they were descriptive or assessment in nature. We used the following guidelines:

1. Tags that referred to meal ingredients (meat, potatoes), methods of cooking (fried), and style of cooling (Italian) were coded as Descriptive

2. Tags that referred to food groups (protein, carbohydrates), or any other nutritional attributes (comment on portion size, proportions of different food groups, etc.) were coded as Assessment

We then calculated the total number of tags of each type for each individual for the combinations of conditions and trials. In each of these groups, we determined whether individuals gave preference to descriptive tags or to assessment tags. We then used Chi-Square analysis to determine differences in proportions of tags between different between-group and within-group factors.

Explanations were reviewed and manually coded by a diabetes educator/dietician on their accuracy. The dietician read all the explanations, compared them with the actual images of meals, and rated them on a scale from 1-5, with a score of 1 indicating "absolutely wrong" and a score of 5 as "perfectly correct." We then aggregated the scores into 2 categories, correct and incorrect, and used Chi-Square analysis to evaluate the difference in distributions of answers across conditions and trials.

All the coding was done with researchers and dietician blinded to participants' assignment to conditions.

Participants

We recruited 58 participants among students enrolled in an Introduction to Psychology class, using the Experimetrix experiment scheduling software. The participants varied in age between 18 and 22 (Median=20), with an even split of Male/Female participants (28 Male, 30 Female). Not surprisingly, all of them were proficient with technology: over 95% reported using the Internet, SMS, email, and Facebook several times daily. However, only a few reported previous exposure to social tagging mechanisms (less than 8%). For those who used social tagging, Flickr, Del.icio.us, and Technorati were the services participants had experience with.

RESULTS

Tags and Tagging Vocabularies

In our data, we found a range of approaches to assigning tags to pictures of meals. Close to 70% of the tags were succinct, utilizing one-word or short phrase tags. However, the remaining 30% of tags included longer descriptions, at times including several sentences. Many of these longer tags were highly judgmental, and many humorous. Below are some examples of these longer tags:

- *Healthy meal, chips not needed but doesn't take away from anything*
- *Very healthy, juice as opposed to coffee would be better*

Considering the *types* of tags used (descriptive vs. assessment), the results supported our expectations that *without guidance, individuals are more likely to generate simpler descriptive tags*. In Intervention trials of Unassisted Tagging (Free) group, the ratio of Descriptive/Assessment tags was13/7, with a ratio of 12/8 for Retention trials.

At the same time, and consistent with our expectations, we found that *exposure to a tagging vocabulary affected*

individuals' own tagging strategies: In the Descriptive Tagging condition, the majority of participants maintained the same preference for descriptive tags as the baseline group, with the ratio of Descriptive/Assessment tags of 16/4. In contrast, *those exposed to a vocabulary containing assessment tags adopted them in their own practice* (Descriptive/Assessment=3/14, difference from baseline: $\chi^2(1)=8.395$, $p<0.01$).

At the same time, this impact was not persistent: Many of those in the Assisted Tagging conditions switched back to using descriptive tags once the vocabulary was no longer visible. The ratio of tags in the Retention trial was Descriptive/Assessment=13/4, which was significantly different from Intervention trial: $\chi^2(1)=14.235$, $p<0.01$).

	Intervention	Retention
	Descriptive/Assessment	Descriptive/Assessment
Free	13/7	12/8
Descriptive	16/4	19/1
Assessment	3/14	13/4

Table 2: Proportions of Descriptive/Assessment tags for different conditions

Impact of tags

Baseline

To the impact of tags, we first needed to establish the baseline performance for the participants. A series of one-way ANOVA tests showed that the groups did not significantly vary in their ability to remember the images of meals; to accurately assess their nutritional values; or to select the best meal out of a set of 3. The participants' performance during the baseline trial did not differ significantly across the three groups.

Memory

As we expected, use of tagging mechanisms significantly and positively impacted individuals' ability to remember the meals as compared to no tagging baseline performance. As evident in Table 3, memory performance improved for all three groups (Free, Descriptive, and Assessment), creating a wsignificant effect of a within-group factor ($F=26.075$, $p<0.001$). There were no significant differences in how different types of tags impacted memory of the meals.

	Baseline	Intervention	Retention	Max
Free	8.42 (.81)	8.95 (.21)	8.76 (.53)	9
Descriptive	8.45 (.88)	8.9 (.30)	8.55 (.75)	9
Assessment	8.64 (0.49)	8.94 (.24)	8.54 (.60)	9

Table 3: Means and Standard Deviations for Recognition across Trials and Conditions

Assessment

The results of the assessment test were not as straightforward and contradicted some of our expectations. First of all, it showed a relatively low overall performance across conditions. In addition, instead of improving participants' ability to assess the meals, the introduction of tags led to a further drop in their performance, producing a significant negative impact of the within-group factor ($F=27.516$, $p<0.001$). *For the Free Tagging and Descriptive Tagging conditions, the introduction of tags significantly reduced individuals' ability to evaluate the meals.* However, the trend was different for the assisted tagging with assessment tags condition. These participants, while starting off at a slightly lower baseline, somewhat improved their scores on the assessment test for the Intervention trial, however this difference was not significant. Once the tagging vocabulary was removed this group's performance went down as well.

	Baseline	Intervention	Retention	Max
Free	2.04 (.74)	1.38 (.8)	1.33 (.65)	4
Descriptive	2.1 (.71)	1.5 (.76)	1.4 (.59)	4
Assessment	1.52 (1.06)	1.58 (.79)	1.11 (.69)	4

Table 4: Means for Assessment across Trials and Condition

Analysis of coded explanations produced similar results. The table below displays the proportions of correct/incorrect explanations for different conditions across trials.

	Baseline	Intervention	Retention
Free	94/32	87/39	99/27
Descriptive	102/18	94/26	107/13
Assessment	81/21	72/30	89/13

Table 5: Proportions of correct/incorrect explanations for answers for groups and trials

When comparing accuracy of explanations across groups and trials, we found that *using free-form tags during Intervention led to a decrease in the accuracy of explanations compared to baseline* ($\chi^2(1)=5.088$, $p<0.05$). However, *using free-form tags in Retention, after the exposure to Assisted tags, led to an increase in accuracy of explanations* ($\chi^2(1)=5.956$, $p<0.05$). No other trends were significant.

Choice

In regard to individuals' ability to select the best out of a set of 3 images, we found that the introduction of tagging mechanisms led to the drop in individuals' performance. The analysis showed a significant impact of both between-group factors ($F=1197.973$, $p<0.001$) and within-group factors ($F=28.02$, $p<0.001$). In the free-tagging Retention trial, however, all groups either returned to their baseline (Assessment Tagging condition) or exceeded it (Free Tagging and Descriptive Tagging conditions).

	Baseline	Intervention	Retention	Max
Free	1.47 (.81)	0.95 (.66)	1.61 (.49)	2
Descriptive	1.6 (.5)	1.00 (.64)	1.8 (.41)	2
Assessment	1.82 (.52)	0.94 (.65)	1.82 (.39)	2

Table 6: Means for Choice across Trials and Conditions

DISCUSSION

In our study, we set out to explore how social tagging mechanisms affect knowledge sharing communities in the

area of health and wellness management. In particular, we wanted to investigate whether these mechanisms can help individuals improve their understanding of nutrition, as well as their ability to critically assess the nutritional value of the meals captured in photographs.

Richness of Tagging Approaches

The results of the study exposed a diversity of tagging approaches. Many individuals used concise single-word tags, whereas others wrote complete sentences that captured their sentiments in regards to the meals. Based on this observation, single-word tag restrictions, common to many collaborative tagging applications, might present a challenge to users in the context of nutrition management.

Mixed Impact of Collaborative Tagging

Limited impact of Tagging

First, our study shows that *all tagging mechanisms led to significant improvements in memory recall of meals as compared to the no tagging condition*. This is an encouraging finding, because it suggests that tagging online content helps individuals remember what they've seen. However, a prolonged exposure to the stimuli that likely occurred in tagging conditions presents an alternative explanation to this finding. Further research could help to distinguish between the two.

With respect to facilitating deeper processing, comprehension, and sensemaking, we found that tagging mechanisms can, in some circumstances, be detrimental. Specifically, we found that tagging can assist individuals only when the tagging scheme is consistent with the task individuals need to accomplish. Of the three experimental conditions—unassisted tagging, supported tagging with descriptive tags, and supported tagging with assessment tags—*only supported tagging with assessment tags did not lead to the decline in individuals' assessment abilities*. In contrast to other forms of tagging, this latter condition led to a non-significant improvement in their ability to critically evaluate nutritional value of meals. Drawing on previous work in classification and categorization, we propose that a mismatch between the tagging scheme and the task at hand produces *cognitive interference* [19], and presents barriers to sensemaking. In our case, when individuals composed or selected tags that highlighted meal composition (descriptive tags), they had a harder titime accurately assessing how healthy the meals were.

Conforming to Existing Vocabulary

While the study showed that assessment tags have the most potential in positively influencing sensemaking, we also discovered that this categorization scheme did not come naturally to study participants. *Without guidance, more individuals chose descriptive tags as opposed to assessment ones.* However, it appears to be possible to *influence individuals' choice of categorization scheme with examples in the tagging vocabulary.* In both supported tagging conditions (which exposed the participants to a particular tagging vocabulary), participants adopted the tagging style consistent with that vocabulary. Individuals with access to the vocabulary composed of descriptive tags adopted them for the duration of the study. Conversely, those with access to a vocabulary composed of both descriptive and assessment tags relied on assessment tags. This finding is consistent with the results reported by Fu et al [6], however, in our study the effect was short-lived and did not translate into a long-term impact; individuals who adopted assisted tags switched back to using descriptive ones when the vocabulary was no longer visible.

Design Implications

Collectively, these findings suggest that social tagging mechanisms have a complex cognitive basis, require further exploration, and should be used with caution. A vocabulary inconsistent with individuals' tasks or goals can lead to cognitive interference and be detrimental to their performance in these tasks. In the case of nutrition management, if the tagging vocabulary is disproportionately populated with tags that refer to meal ingredients, this vocabulary may interfere with individuals' ability to focus on the nutritional value of meals.

However, our study suggests ways to assist individuals in adopting more beneficial tagging practices. Specifically, they recommend the need to engage experts in monitoring and potentially adjusting tagging vocabularies that are generated by the community. This moderation can ensure that the vocabulary is consistent with the task and does not interfere with individuals' goals.

Limitations

The described study has a number of limitations. First of all, the choice of controlled experiment as our primary research methodology limits our ability to generalize our findings to real world situations. Tagging practices are spontaneous and highly context-sensitive. The tags individuals chose in the context of our experiment may be very different from what they would have chosen in a real-world environment. Our study design allowed us to explore several questions simultaneously; however it resulted in reduced number of trials for each type of assessment (only 9 for memory, 4 for assessment and 2 for choice). This also led to somewhat small effects; while we were able to detect some statistically significant differences, the effect sizes were small and are more suggesting than confirmatory. Finally, our findings might be specific to our study population (college students), thus limiting their applicability to other user groups. For example, our participants were more familiar with social tagging mechanisms than individuals of older age who are more common participants of health and wellness communities.

CONCLUSIONS

Well-designed tools can help individuals take advantage of their collective expertise. In the context of nutrition management, an ability to learn from others can increase individuals' nutritional literacy and help them adopt healthier eating habits. Collaborative tagging mechanisms have long been lauded for their potential to assist collective

sensemaking. There is, however, a lack of empirical evidence to support these claims.

We conducted an experimental study to evaluate the impact of collaborative tagging on nutritional sensemaking, including memory and critical evaluation of meals. The results show that the composition of tagging vocabulary used by individuals has a significant impact on their ability to critically evaluate viewed digital content. This suggests a need for careful assessment and moderation of emerging tagging vocabularies.

ACKNOWLEDGEMENT

We are grateful to all our participants, to Dr. Wendy Rogers for her invaluable assistance with the design of the study and to Kliment Mamykin for assistance with the development of the experimental software.

REFERENCES

1. TuDiabetes. http://www.tudiabetes.org/.
2. PatientsLikeMe : Patients Helping Patients Live Better Every Day. http://www.patientslikeme.com/.
3. Eat.ly - Track, share, and discover great food. http://blog.eat.ly/.
4. Budiu, R., Pirolli, P., and Hong, L. Remembrance of things tagged: how tagging effort affects tag production and human memory. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 615-624.
5. Chi, E.H. and Mytkowicz, T. Understanding the efficiency of social tagging systems using information theory. *Proceedings of the ACM conference on Hypertext and hypermedia*, ACM (2008), 81-88.
6. Fu, W., Kannampallil, T., Kang, R., and He, J. Semantic imitation in social tagging. *ACM Trans. Comput.-Hum. Interact. 17*, 3 (2010), 1-37.
7. Golder, S.A. and Huberman, B.A. Usage patterns of collaborative tagging systems. *Journal of Information Science 32*, 2 (2006), 198 -208.
8. Halpin, H., Robu, V., and Shepherd, H. The complex dynamics of collaborative tagging. *Proceedings of the 16th international conference on World Wide Web*, ACM (2007), 211-220.
9. Hsieh, G., Lai, J., Hudson, S.E., and Kraut, R. Using tags to assist near-synchronous communication. *Proceeding of the SIGCHI conference on Human factors in computing systems*, ACM (2008), 223-226.
10. Kiewra, K.A. Notetaking and review: The research and its implications. *Instructional Science 16*, 3 (1987), 233-249.
11. Kiewra, K.A. A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review 1*, 2 (1989), 147-172.
12. Kiewra, K.A. and Frank, B.M. Cognitive style: Effects of structure at acquisition and testing. *ED 23*, (1985), 261-072.
13. Mamykina, L., Mynatt, E., Davidson, P., and Greenblatt, D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 477-486.
14. Mamykina, L., Mynatt, E., Davidson, P., and Greenblatt, D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 477-486.
15. Marlow, C., Naaman, M., Boyd, D., and Davis, M. HT06, tagging paper, taxonomy, Flickr, academic article, to read. *Proceedings of the seventeenth conference on Hypertext and hypermedia*, ACM (2006), 31-40.
16. Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., and Chi, E.H. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 1795-1798.
17. Oleksik, G., Wilson, M.L., Tashman, C., et al. Lightweight tagging expands information and activity management practices. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 279-288.
18. Pauen, S. Biobehavioral Development, Perception, and Action: Evidence for Knowledge–Based Category Discrimination in Infancy. *Child Development 73*, 4 (2002).
19. Sarason, I.G., Pierce, G.R., and Sarason, B.R. *Cognitive interference: theories, methods, and findings*. Psychology Press, 1996.
20. Sen, S., Lam, S.K., Rashid, A.M., et al. tagging, communities, vocabulary, evolution. *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM (2006), 181-190.
21. Shirky, C. Shirky: Ontology is Overrated -- Categories, Links, and Tags. *Clay Shirky's Writings About the Intenet*. http://www.shirky.com/writings/ontology_overrated.html.
22. Vander Wal, T. Thomas Vander Wal: When Folksonomies Supplant Taxonomies « SpicyNodes Blog. *SpicyNodes Blog*. http://www.spicynodes.org/blog/2010/09/13/thomas-vander-wal-when-folksonomies-transplant-taxonomies/.
23. Weick, K.E., Sutcliffe, K.M., and Obstfeld, D. Organizing and the Process of Sensemaking. *ORGANIZATION SCIENCE 16*, 4 (2005), 409-421.
24. Yew, J., Gibson, F., and Teasley, S. Learning by tagging: group knowledge formation in a self-organizing learning community. *Proceedings of the 7th international conference on Learning sciences*, International Society of the Learning Sciences (2006), 1010-1011.