

The Role of Explanations in Casual Observational Learning about Nutrition

Marissa Burgermaster¹, Krzysztof Z. Gajos², Patricia Davidson³, Lena Mamykina¹

¹ Columbia University
New York, NY, USA
{mb3381, om2196}
@cumc.columbia.edu

² Harvard University, School of
Engineering and Applied Sciences
Cambridge, MA, USA
kgajos@eecs.harvard.edu

³ West Chester University
West Chester, PA, USA
nutriciard@yahoo.com

ABSTRACT

The ubiquity of internet-based nutrition information sharing indicates an opportunity to use social computing platforms to promote nutrition literacy and healthy nutritional choices. We conducted a series of experiments with unpaid volunteers using an online Nutrition Knowledge Test. The test asked participants to examine pairs of photographed meals and identify meals higher in a specific macronutrient (e.g., carbohydrate). After each answer, participants received no feedback on the accuracy of their answers, viewed proportions of peers choosing each response, received correctness feedback from an expert dietitian with or without expert-generated explanations, or received correctness feedback with crowd-generated explanations. The results showed that neither viewing peer responses nor correctness feedback alone improved learning. However, correctness feedback with explanations (i.e., modeling) led to significant learning gains, with no significant difference between explanations generated by experts or peers. This suggests the importance of explanations in social computing-based casual learning about nutrition and the potential for scaling this approach via crowdsourcing.

Author Keywords

Nutrition literacy; casual learning; observational learning; crowdsourcing

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Public health “promotes and protects the health of people and the communities where they live, learn, work and play” [5] with a particular focus on prevention through adoption of healthy behaviors. Traditional public health interventions

focused on education and messaging to reach broad segments of population through mass broadcasting and health campaigns. The increasing popularity of social computing platforms creates a new opportunity to both study attitudes and social norms towards health and health behaviors, and to influence these behaviors on an unprecedented scale. There is potential to integrate the human-computer interaction (HCI) and public health fields and move toward HCI applications that support health promoting practices and norms.

The phenomenon of the food photo sharing has arisen simultaneously with the ubiquity of smartphones and social media. A recent marketing analysis found that 22% of sampled participants were “highly engaged with technology and food” [14] and people take and share photographs of their meals regularly on social media platforms, including Instagram [23] and Twitter [38]. Twitter users’ interests, demographics, and social networks have been linked to their food-related tweets [1]. Sharma & DeChoudhury [33] analyzed over 1 million food-related Instagram posts for calorie content and then used these data to link posts about foods with moderate calorie content with more and more positive community response than those with very low or very high calorie content. These findings suggest the potential for large-scale observational learning about nutrition in the social computing context. However, previous research on food photo sharing with social media focused primarily on existing attitudes and photo sharing practices, rather than on examining whether photo sharing presents opportunities for learning about nutrition and for promoting healthy eating choices. Determining if people can increase their nutrition literacy through food photos on social media and if so, identifying mechanisms by which this type of learning occurs, could make a significant contribution to public health research seeking to promote healthy eating.

In this research, we are particularly interested in examining the possibility of enabling observational nutritional learning with social computing platforms. Observational learning—learning that happens through observing behaviors of others (models) rather than through formal instruction or one’s own experience [7]—has many advantages over other types of learning mechanisms and is particularly scalable in the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4655-9/17/05...\$15.00

DOI: <http://dx.doi.org/10.1145/3025453.3025874>

context of online communities and social media platforms as it relies on resources available within these communities, rather than experts or expert-created instructional materials. Previous research investigated the plausibility of observational learning about nutrition with simple social computing features, such as collaborative tagging [26] and within communities of paid crowd workers [27]. The results of the work with paid workers showed that on tasks that required nutritional knowledge, observing aggregated solutions provided by others helped workers on Amazon Mechanical Turk (MTurk) improve their understanding of nutrition, and led to increased accuracy in their own solutions [27]. Building upon this prior work, the questions investigated in this study examined whether these findings hold for communities of volunteers interested in casual learning about nutrition, and for more realistic nutritional tasks consistent with nutritional choices made by regular individuals. If this study returned positive results, it could suggest that observational learning is a plausible mechanism for promoting nutrition literacy among intrinsically motivated individuals and in the context of their nutritional choices.

We conducted two experiments to examine the effectiveness of observational learning about nutrition in the social computing context, and to compare it with more traditional mechanisms for facilitating learning, such as expert-generated correctness feedback. We conducted these experiments with a Nutrition Knowledge Test (NKT) on a high-volume, web-based research platform that hosts multiple behavioral experiments drawing intrinsically motivated, unpaid online volunteers. Using this platform, volunteers participate in engaging activities and consent to provide the data they generate to researchers in exchange for information about their personality or behavior. These informal and intrinsically motivated activities align well with the concept of casual learning [21].

Using the NKT, we assessed the influence of different observational learning interventions on performance of a nutrition literacy task. We asked participants to perform a realistic, but complex task: Comparing photographs of mixed-ingredient meals and identifying and evaluating the ingredients that contribute to macronutrient (i.e., carbohydrate, protein, fat, fiber) content. Because of the high volume of participation, the research platform presented an opportunity to compare the effects of different types of learning mechanisms on task performance. These mechanisms included 1) traditional correctness feedback obtained by comparing an individual's answer to the answer generated by a trained registered dietitian nutritionist (RDN), 2) correctness feedback accompanied by an explanation for how the correct answer was achieved, which served as a model for promoting observational learning, and 3) more direct observational learning through comparison of an individual's answer with distribution of answers provided by others who took the same test. We

also examined if feedback and explanations generated by peers and experts had similar effects on learning.

The study showed that people participating in social computing-based casual learning significantly improved their accuracy between the pre- and post-tests when we provided them with correctness feedback together with explanations. We further established the potential for scaling this approach through crowdsourcing by demonstrating that the learning gains yielded by peer-generated explanations were not statistically different from those achieved with expert-generated explanations.

BACKGROUND AND THEORETICAL FOUNDATIONS

Nutrition Literacy

Health literacy, “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions,” [30] has been linked to health outcomes [8], health system navigation [28], and primary and secondary prevention of disease [32]. The more specific *nutrition literacy* has been less well studied [10], but nutrition knowledge has been related to healthier eating patterns [34]. Although it is well established that knowledge alone is not sufficient to change nutrition behavior [12], nutrition knowledge is certainly an important part of dietary decision making.

Nutrition literacy has been linked to “nutrition information-seeking behaviors” [40]. In a 2011 survey of randomly sampled Americans, 46% reported that they actively engaged in nutrition information seeking; though 67% reported receiving information about nutrition from television, 40% reported receiving it from the internet, with higher reporting among those aged 18-44 [4]. Nutrition literacy typically focuses on increasing nutrition knowledge and skills, like reading nutrition facts labels [10, 20]. Its application is illustrated by Gao, Constanza, and schraefel's [17] examination of participants' ratings of the healthiness of meal photographs; they found that macronutrients were the most frequently considered attribute in assigning a rating. Researchers have called for improved nutrition literacy tools and applications [20]. Given the increase in nutrition information seeking that occurs on the internet, online and social computing-based learning activities may be a particularly appropriate solution. Furthermore, providing learning activities in the context of real meals may have a stronger link to dietary choices.

Observational Learning

Observational learning, an essential component of Bandura's Social Cognitive Theory, is based on the human phenomenon of vicarious learning by observing others' actions and their consequences [7]. Observational learning has been implicated in aggressive behavior [6], online purchasing behavior [11], and attitudinal diffusion through a cultural group [31]. Observational learning requires: 1) attention to the behavior being modeled, 2) retention of the

salient aspects of the modeled behavior, 3) ability to enact the behavior, and 4) motivation to engage in the modeled behavior [7]. Importantly, motivation typically arises when one observes models being rewarded for the modeled behavior or when models are authority figures [7].

Observational learning has been applied in research on social computing platforms, specifically in the area of crowdsourcing, working under the assumption that collective intelligence [37] may improve the quality of modeling. In peer-to-peer lending, observational learning led to more rational assessment of borrower creditworthiness [39]. In nutrition, Mamykina, Smyth, Dimond and Gajos [27] examined how paid crowd workers' accuracy in identifying the nutritional composition of different meals was affected by expert feedback, peer-generated direct feedback on their solutions, and comparison of their solutions with solutions provided by others. They found that workers' accuracy improved both when they received expert feedback as well as when they were asked to compare their responses to others' responses, but not when they received direct feedback from peers. These findings suggest that exposing individuals to crowdsourced models that explain the correct response to the task can facilitate observational learning. This has important implications for scaling learning via observational learning. But what about when learning is not directly motivated by financial gain or loss?

Casual Learning

As early as 1932, casual learning was mentioned in educational research in the context of learning outside of the education system and as synonymous to observational learning [21]. More recently, casual learning has become a buzzword in business where it is applied to employee training by encouraging employees to learn about what is interesting and relevant to them, fits their needs, happens in a social environment where and when people have time or need for it. Casual learning has also been associated with informal education (e.g., self-help books, seminars) and with online education platforms for continuing education, like MOOCs and businesses like Udemy, which provide courses across a wide range of topics and interests. Casual learning is characterized by *intrinsic motivation*, the "motivation based in people's natural interest in various activities that provide novelty and challenge" [13]. Duolingo [36] is an example of casual learning that has been demonstrated to improve Spanish language ability. Notably, those who used Duolingo to learn Spanish for travel (and were therefore assumed to be more intrinsically motivated) had significantly greater learning gains than those who used it for school or personal interest [35]. In nutrition, Epstein, Cordeiro, Fogarty, Hsieh and Munson [16] found that social computing-based observational learning was a potential mechanism for improving nutrition behavior via daily nutrition challenges, like "eat a food high in fiber." During their "Crumbs" intervention, participants noted that they got ideas for completing the challenges by

viewing photographs of how other participants met the challenges.

Taken together, these theoretical foundations point to the potential for casual observational learning to occur when people are intrinsically motivated to participate in learning activities to improve nutrition literacy on a social computing platform. Here we report on the results of two experiments using NKT to better understand if observational learning can improve nutrition literacy in a casual learning social computing context.

Conducting behavioral research with unpaid online volunteers

Several web-based platforms (e.g., TestMyBrain.org, Project Implicit¹, GamesWithWords.org, LabintheWild.org) have emerged for conducting behavioral research with unpaid online volunteers. Instead of offering monetary compensation, these platforms motivate participation by promising participants that they will receive their results immediately at the end of the study and will be able to compare their own performance to others. Several validation studies demonstrated that data collected on those platforms are not statistically different from the data collected in conventional laboratory settings [18, 19, 29]. Despite the fact that participants self-select to take part in studies on these platforms, the samples appear more diverse (and thus more representative of the general population) than the samples that participate in studies in conventional academic laboratories or populations recruited via MTurk [29]. We launched our experiment on LabintheWild.org.

ENABLING CASUAL OBSERVATIONAL LEARNING IN NUTRITION

We studied how people learn about nutrition from others in an online context using an internet platform where volunteers participate in engaging online experiments in exchange for personalized feedback about their personality or behavior, i.e., casual observational learning.

Nutrition Knowledge Test

We designed, implemented, and launched a Nutrition Knowledge Test (NKT). The test asked participants to perform a realistic, but complex task: Comparing complicated mixed-ingredient meals and identifying and evaluating the ingredients that contribute to macronutrient content. NKT participants were shown a series of realistic meal photograph pairs. Each pair had one difference in ingredients that led to one of the meals in the pair containing a higher amount of a specific macronutrient (i.e., carbohydrate, fiber, fat, or protein). Participants were asked to view the photographs and textual descriptions of pairs of meals and to choose which meal had more of a given macronutrient (pretest). Immediately after each response, participants received a message with feedback about their response in the form corresponding to one of the feedback conditions, including the no-feedback condition. Later in

¹ <https://implicit.harvard.edu/implicit/>

	Study 1	Study 2	Peer Explanation Generation
Number of participants	1315	556	148
Age in years (mean±SD)	30±13	32±14	32±14
Female	63%	68%	55%
Education			
Pre-high school	1%	2%	1%
High school	16%	11%	17%
College	70%	62%	72%
PhD	13%	8%	10%
Computer usage			
Once a week or less	1%	1%	1%
A few times a week	4%	6%	4%
A couple of hours most days	29%	31%	24%
Many hours on most days	66%	62%	66%
Self-reported nutrition knowledge compared to peers			
Much less	3%	2%	4%
A little less	6%	6%	7%
About the same	27%	29%	28%
A little more	45%	38%	40%
Much more	20%	25%	22%

Table 1: Participant Characteristics by Study

the test, participants were asked to examine a second pair of photographs with the same key ingredient in different meals and were again asked to identify which meal contained more of a given macronutrient (posttest). The study materials were prepared in a way to minimize carryover effect between feedback conditions.

The meal photographs presented to participants during the NKT were randomly selected from 19 topic sets. Each topic set included 2 pairs of photographs, one pair presented at pretest and one at posttest. In each pair, there was a specific ingredient that contributed significantly to the difference in macronutrient content of otherwise similar meals (e.g. milk in latte versus cream in coffee with cream). Although the meal in which the key ingredient was presented was different at pre- and posttest, the key ingredient was the same (e.g. milk in latte versus cream in coffee in pretest, and milk in oatmeal versus water in oatmeal in posttest with milk as the key ingredient contributing to the difference in carbohydrate in each of these pairs). We ensured each participant was presented with an equal number of topic sets per macronutrient, but each topic set of pre- and posttest pairs was otherwise presented randomly and independently. This minimized the possible impact of order in which ingredients were presented. Key ingredients were unique to each topic set; therefore, knowledge was not generalizable across topic sets or feedback conditions (e.g., knowing that milk is higher in carbohydrate than cream or

water does not help to learn that soybeans are higher in protein than pinto beans).

Our outcome of interest, learning gain, was defined as pretest accuracy (the accuracy of the participants' response to the first pair of meals with a given ingredient, before receiving any feedback) subtracted from posttest accuracy (the accuracy of participants' response to the second pair of meals with the same ingredient, after reviewing feedback). The ingredient topic was assigned from the 19 topic sets and the type of feedback provided was randomly assigned to each ingredient topic. Each participant received all types of feedback of interest during the study.

In the design of the NKT, we built on prior work on conducting studies with unpaid online volunteers to motivate participation and ensure validity of the data [18, 19, 29]. Specifically, we motivated participation by promising participants that they would receive feedback on their performance and would be able to compare their own performance with that of other people. We also made all demographic questions optional (to reduce chance of untruthful responses) and we gave participants opportunities to report if they had taken the test before, if they had cheated in any way or if they had encountered any technical difficulties.

Participants arrived at the study site primarily through word-of-mouth referrals or a referral at the end of another

study on the site. The landing page for the study included study title and a brief description of the study, including the type of feedback the participant would receive at the end and the anticipated duration of the study (10 minutes). Next, participants were presented with an informed consent statement followed by a brief demographic questionnaire. Next, participants saw brief instructions and then began the actual test. After completing the test, but before seeing their results, participants were offered a chance to indicate whether they cheated in any way or encountered technical difficulties. Finally, they were presented with the results page, where they saw the percentage of the questions they answered correctly, the average score of other participants, their own score broken down by macronutrient as well as the correct answers and expert explanations for all the questions they got wrong. Participant characteristics for each of the studies are reported in Table 1.

Study 1: Feedback, Peer Comparison, or Explanations?

First we studied if correctness feedback generated by experts improved learning in the NKT, if exposure to the

distribution of answers provided by peers led to learning gains (i.e., peer modeling), if correctness feedback accompanied by an explanation offering further detail about why the correct answer contained more of the macronutrient further improved learning, and if there was difference in learning between correctness feedback generated by experts and comparison with solutions generated by peers.

Participants

We collected data from 1315 NKT participants from December 2015 to March 2016. Detailed information about the participants is provided in Table 1.

Tasks

We created four variants of the intervention to be shown to the participants immediately after they provided their response to a question, summarized below and depicted in Figure 1.

Control: No immediate feedback to the participant, but let participants know they would receive feedback later.

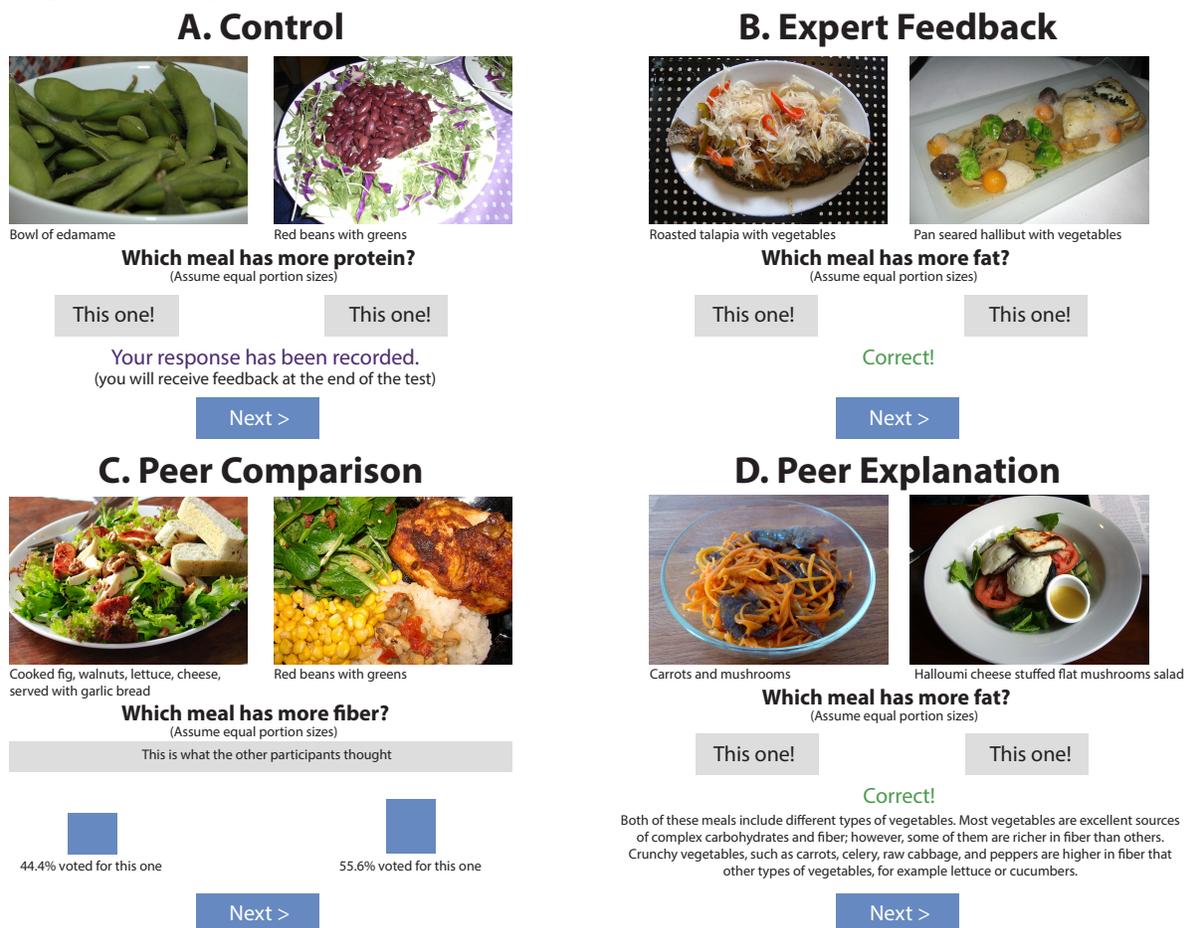


Figure 1: Examples of Study 1 exposure conditions.

Images: A1 – by Jon 'ShakataGaNaI' Davis licensed under CC BY-SA 3.0 via Wikimedia Commons; A2 – by Niupoundswea licensed under CC BY-SA 3.0 via Wikimedia Commons; B1 – by Ramon F Velasquez licensed under CC 1.0 via Wikimedia Commons; B2 – by Arnold Gatilao licensed under CC BY 2.0 via Flickr; C1 – by Triv.rao licensed under CC BY-SA 4.0 via Wikimedia Commons; C2 – by Manfred&Barbara Aulbach licensed under CC BY-SA 3.0 via Wikimedia Commons; D1 – by Off-shell licensed under CC BY-SA 4.0 via Wikimedia Commons D2 – by Ewan Munro licensed under CC BY-SA 2.0 via Wikimedia Commons

Expert correctness feedback: Simple correct (“Correct!”) or incorrect (“Not Quite.”) feedback immediately after the participant submitted their response.

Peer comparison: Visually depicted proportions of other participants who chose each response option. Peer feedback was calculated based on responses of participants from a previous version of NKT. It is important to note that because these responses were generated from the crowd, there could have been meal pairs for which more peers chose the incorrect response.

Expert correctness feedback with explanation: Correctness feedback accompanied by a short paragraph written by an RDN describing why the correct meal had more of the indicated macronutrient. Each explanation highlighted the specific ingredient that made the difference in macronutrient content and quantified the difference in macronutrient for participants. This condition combined elements of both correctness feedback and observational learning by modeling expert reasoning through their explanation.

Design and analysis

We used a within-subjects repeated measures design. Each participant randomly received no feedback (control; 25%) immediately following her response to a question, expert correctness feedback (25%), peer comparison (26%), or expert correctness feedback with expert explanation (24%) for each of the 24 items.

We calculated learning gain (the difference between posttest accuracy and pretest accuracy) and determined if learning had occurred in each of the four conditions using one-sample t-tests adjusted for multiple comparisons with Holm’s sequentially rejective correction [22]. Then, we used a linear mixed-effects model fit by maximum likelihood to compare the effect of each exposure on posttest accuracy across groups, controlling for pretest accuracy. This model accounted for the within-subjects design by including the participant as a random effect on exposure. We used Tukey’s HSD posthoc procedure with single-step adjustment to assess differences between groups.

Results

Pretest accuracy was similar across conditions. Accuracy increased at posttest in both the expert explanation and

expert feedback groups, but not in the control or peer comparison groups. Learning gains occurred in the expert feedback and expert explanation conditions, but only expert explanations caused a learning gain that was significantly greater than zero when correcting for multiple comparisons. These results are reported in table 2.

Controlling for pretest accuracy, we observed a significant main effect of feedback condition on posttest performance ($F_{2,372} = 9.183, p = .0004$). Posthoc comparisons indicated that, controlling for pretest accuracy and within-subjects design, **expert explanations led to significantly higher posttest accuracy than any other condition**. Expert feedback led to significantly better posttest accuracy than control, while peer comparison did not, but the difference between expert and peer comparison was not significant. These relationships are depicted in Figure 2.

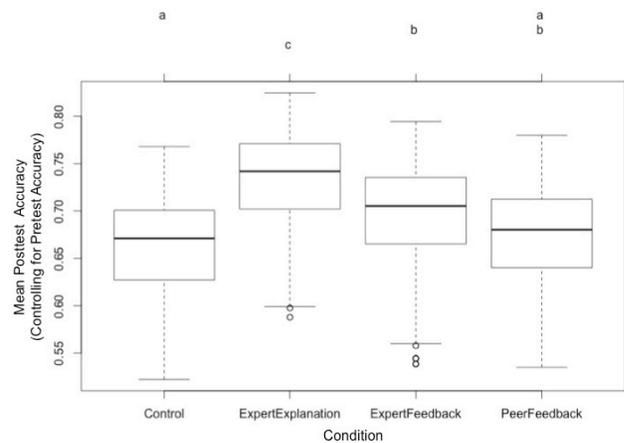


Figure 2: Tukey’s Posthoc Comparison of Exposure Groups. Based on results of a repeated-measures ANOVA. Letters indicate significant differences among groups.

Study 2: Expert or Peer Explanations?

Study 1 revealed that expert explanations yielded significant learning gains as well as significantly better posttest accuracy than the other exposure conditions studied. These results suggest that for a nutrition test that involves complex multi-ingredient meals, simple correctness feedback and peer comparison are not sufficient to facilitate learning. However, when this feedback was

	Control	Expert Feedback	Expert Explanation	Peer Comparison
Learning gain (mean±SD)	0.01±0.6 t=0.6 Cohen’s d = 0.02	0.03±0.6 t=2.4 Cohen’s d = 0.08	0.06±0.6*** t=5.6 Cohen’s d = 0.2	0.0±0.6 t=0.2 Cohen’s d = 0.04
Pretest accuracy (mean±SD)	66±47%	67±47%	67±47%	67±47%
Posttest accuracy (mean±SD)	66±47%	70±46%	74±44%	67±47%

Table 2: Summary of learning and accuracy in each exposure group. Learning gain greater than 0 assessed with one-sample t-test; significance indicated, * p<.0001**

accompanied by an explanation that drew participants' attention to the sources of difference between the meals and specifically pointed to which ingredient was responsible for the difference and why, individuals achieved higher accuracy on follow-up questions with the same key ingredient, thus exhibiting learning. While these results were encouraging, composing such explanations is a labor-intensive undertaking that relies on availability of experts, and is not scalable in the context of large volunteer-based communities. In the second study, we examined the potential of using the community itself to generate effective explanations, and compared learning gains with explanations carefully crafted by experts and explanations generated by test-takers. Notably, in this study we continued to rely on experts for generating correctness feedback.

Participants

We collected data from 556 NKT participants from July-September 2016. Detailed information about the participants is provided in Table 1.

Tasks

We created three variants of the intervention to be shown to the participants immediately after they provided their response to a question, summarized below and depicted in Figure 3. In addition to collecting participant responses, we also asked them to provide information about the confidence they had in their responses using a slider. Instead of clicking a button indicating binary "This one" as in Study 1, participants moved a slider along a color gradient continuum from neutral center towards "I am confident this is the right answer." As they moved the slider closer to maximum confidence, the outline around the photograph became darker.

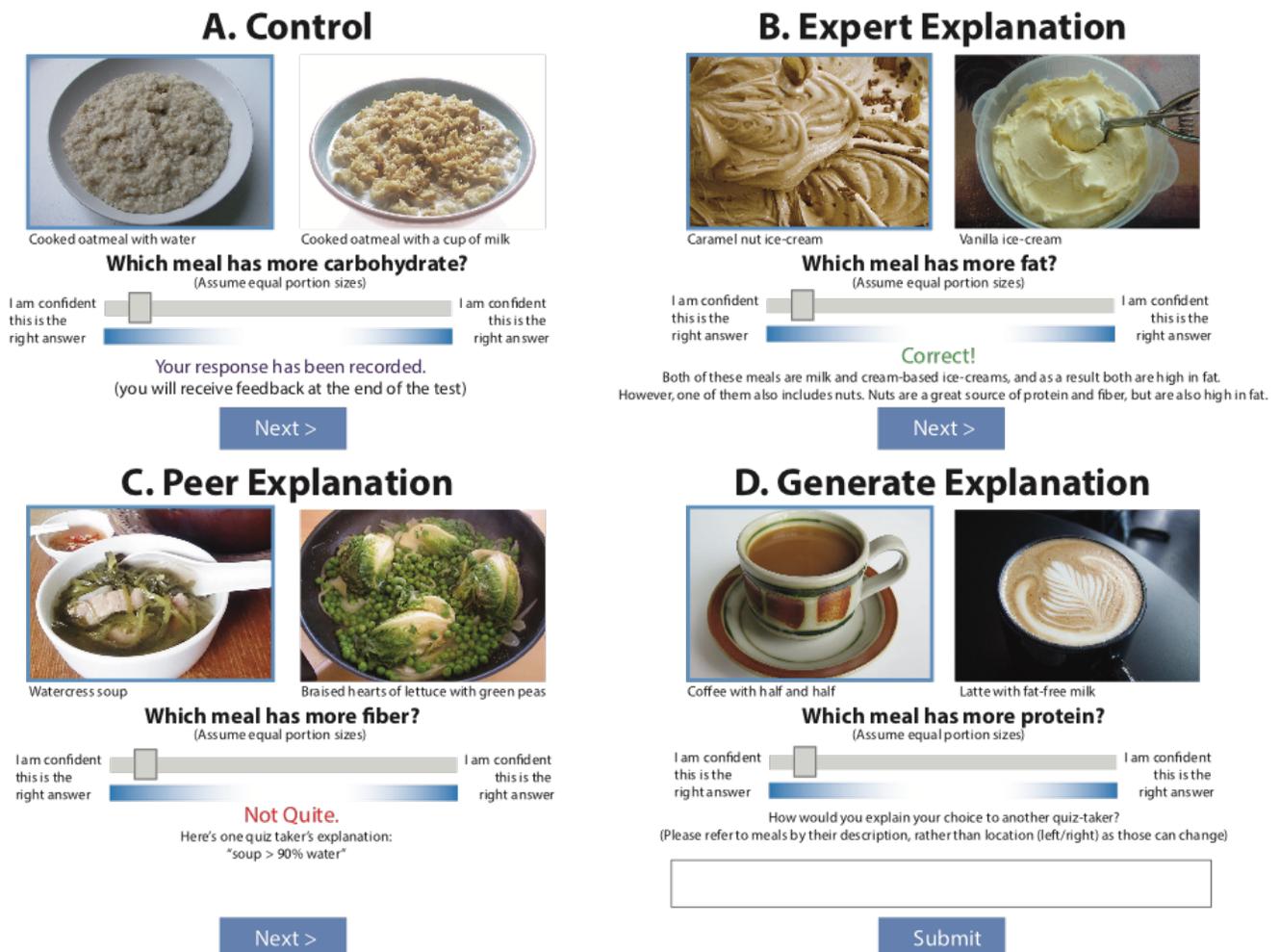


Figure 3: Example conditions for Study 2.

Images: A1 – by Nillerdk licensed under CC BY 3.0 via Wikimedia Commons; A2 – by Renee Comet for National Cancer Institute public domain via Wikimedia Commons; B1 – by Lotus Head licensed under CC BY-SA 3.0 via Wikimedia Commons; B2 – by I, ElinorD licensed under CC BY-SA 3.0 via Wikimedia Commons; C1 – by ProjectManhattan licensed under CC BY-SA 3.0 via Wikimedia Commons; C2 – by Kochtopf licensed under CC BY 2.0 via Wikimedia Commons; D1 – by Douglas Paul Perkins licensed under CC BY 3.0 via Wikimedia Commons D2 – by Juan de Dios Santander Vela licensed under CC BY-SA 2.0 via Wikimedia Commons

Control: No immediate feedback to the participant, but let participants know they would receive feedback later.

Expert correctness feedback with expert explanation: Like Study 1 described above, correctness feedback in this condition was accompanied by a short paragraph written by an RDN describing why the correct meal had more of the indicated macronutrient. Each explanation highlighted the specific ingredient that made the difference in macronutrient content and quantified the difference in macronutrient for participants.

Expert correctness feedback with peer explanation: This condition used the same correctness feedback as condition 2 above; however, instead of pairing it with expert-generated explanation, it was accompanied by explanations generated by other individuals who took the same test, *peer explanations*. Peer explanations (n=308) were collected prior to the launch of this experiment. NKT participants were asked, “How would you explain your choice to another quiz-taker?” as depicted in Figure 3. A single participant was asked for no more than four explanations and providing an explanation was optional (they could proceed while leaving the explanation field blank). Information about the participants who provided explanations (n=148) is presented in Table 1. Explanations were filtered manually, using two distinct criteria. First, we excluded explanations accompanying incorrect responses (i.e., the meal selected as higher in carbohydrate was, in fact, lower in carbohydrate); this led to exclusion of 41% of explanations (n=185) that came with incorrect responses. Then, we applied exclusion criteria that could be reasonably automated in future applications. Specifically, we excluded explanations that used swear words, had fewer than three words, used only symbols, or were confusing in their reference to the images (e.g., used “image on the left” or “on the right” instead of referring to specific ingredients). Notably, correctness of the explanation itself was not assessed as it would be difficult to replicate with computational methods. “Spilt pea soup incorporates a green vegetable while black bean soup has no green vegetables,” and “Peas are high carb” are examples of typical peer explanations.

Design and analysis

We used a within-subjects repeated measures design. Each participant randomly received no feedback (control; 17%), expert feedback with expert explanation (42%), or expert feedback with peer explanation (42%) for each of 24 items. To optimize participant engagement, we set randomization into the control group to happen half as often as either of the exposure conditions. We analyzed these data using the same methods as in Study 1.

Results

Pretest accuracy was similar across exposure conditions. Accuracy increased at posttest in both the expert explanation and peer explanation condition, but not in the control. Significant learning gains occurred in both the

expert explanation and peer explanation conditions when correcting for multiple comparisons. These results are summarized in Table 3.

Controlling for pretest accuracy, we observed a significant main effect of feedback condition on posttest performance ($F_{2,1110} = 18.906, p < .0001$). Posthoc comparisons indicated that, controlling for pretest accuracy and within-subjects design, *expert and peer explanations both led to significantly higher posttest accuracy than control and there was no significant difference between expert and peer explanations*. These relationships are depicted in Figure 4.

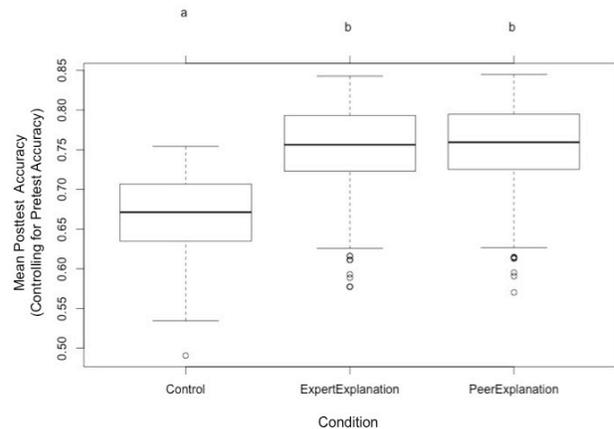


Figure 4: Tukey's Posthoc Comparison of Exposure Groups on Posttest Accuracy. Based on results of a repeated-measures ANOVA. Letters indicate significant differences among groups.

Pretest confidence was similar across study conditions. Confidence decreased significantly in all study conditions, including control. Adjusting for within-subjects design, posttest confidence was significantly greater in both expert and peer explanation exposure groups than control. There was no significant difference in posttest confidence between expert and peer explanation exposure groups. Confidence was a significant predictor of posttest accuracy in a model controlling for pretest accuracy and repeated measures, but when it was included in the ANOVA model, it did not change the results.

DISCUSSION

In this study, we examined how different types of observational learning influenced performance on a nutrition literacy task in the context of casual learning on a social computing platform. Specifically, we examined learning gains achieved from expert correctness feedback, the impact of explanations on individuals' learning, and comparative effectiveness of explanations generated by experts and peers. We examined these questions in the context of nutritional learning about complex multi-ingredient meals and tasks that required participants to compare pairs of meals on their nutritional content and choose which meal was higher in a specific macronutrient.

	Control	Expert Explanation	Peer Explanation
Learning gain (mean±SD)	0.001±0.6 t=0.1 Cohen's d = 0.006	0.1±0.6*** t=8.1 Cohen's d = 0.3	0.1±0.6*** t=8.4 Cohen's d = 0.3
Pretest accuracy (mean±SD)	67±47%	66±47%	66±47%
Posttest accuracy (mean±SD)	67±47%	75±43%	76±43%
Confidence gain (mean±SD)	-0.4±0.3*** t=-5.2 Cohen's d = -0.3	-0.2±0.3** t=-3.6 Cohen's d = -0.1	-0.3±0.3*** t=-5.3 Cohen's d = -0.2
Pretest confidence (mean±SD)	65±30%	65±30%	65±30%
Posttest confidence (mean±SD)	60±30%	63±30%	63±30%

Table 3: Summary of learning, accuracy, and confidence in each exposure group. Learning gain greater than 0 assessed with one-sample t-test; significance indicated, ** p<.001 * p<.0001**

In the first of our studies, we found that only expert-generated explanations of their reasoning about meals, when combined with correctness feedback, significantly improved learning. In contrast, neither correctness feedback from experts nor comparison with solutions provided by peers was sufficient to facilitate learning. This suggests that in cases when nutritional experts are available for generating correctness feedback and explanations for their answers, these mechanisms can be effectively used to facilitate learning. In the second study, we further examined different approaches to constructing explanations, with a focus on enabling scalable solutions that minimized involvement of experts. The study showed that explanations generated by peers, when combined with correctness feedback, were just as effective in leading to learning gains as explanations generated by experts.

These findings are at least partially consistent with existing theories of observational learning and with prior work in using to facilitate learning in crowdsourcing communities. The effectiveness of explanations in improving learning could be attributed to their ability to attune participants to the key ingredient contributing to the difference in macronutrient content between the meals. According to Bandura [7], attuning to important characteristics of modeled behaviors is one of the critical processes of observational learning. That peer modeling through explanations could match learning gains from expert modeling is consistent with observational learning which does not necessarily privilege expert information [7] and with previous work in paid workers [27]. Observational learning seems to have translated to intrinsically motivated volunteers and more complex, realistic tasks, suggesting that with some computational assistance in applying exclusion criteria, similar learning gains may be possible without experts.

At the same time, lack of learning gain after comparing one's own choices with choices of others is inconsistent with previous research for similar nutritional judgment

tasks [27]. This could potentially be explained by difference in the complexity of the task. Bloom's taxonomy [9] categorizes learning objectives along a continuum of concrete to abstract, and emphasizes that knowledge is a prerequisite for all higher-level abstractions. From this perspective, it is reasonable to consider the tasks presented in this study required learners to analyze and apply knowledge, whereas the tasks presented to learners in [27] would require recall of knowledge. Another explanation could be that on some occasions, aggregating peer solutions resulted in an incorrect solution receiving majority vote. This could have been related to information cascade wherein participants followed the pattern established by peer modeling instead of employing their own knowledge [3, 24]. In crowdsourcing, this has been referred to as the "foolishness of crowds" [25].

This identification of observational learning mechanisms that improved nutrition knowledge in a casual learning context could assist in the development of new applications and tools for improving nutrition literacy. On the other hand, these mechanisms could also explain in part the proliferation of misinformation in social computing contexts. In these studies, we have demonstrated that assuming we have ground truth, we can teach people about nutrition in a realistic performance task by providing them explanations, including explanations provided by non-expert peers, that model the thinking behind the correct response. The large number of people who participated in NKT without compensation provides further evidence for the potential of larger scale applications based on these findings: People are motivated by self-discovery, in particular about their own nutrition knowledge. This has implications for scaling this approach in a social computing platform for public health benefit.

In this study, we increased the relevance of learning gains in nutrition literacy by presenting ingredients in the context of complex meals, as participants might experience them in everyday life. We focused on how peer and expert feedback

and explanations help participants better assess the nutritional content of a meal. For many health conditions, nutrition literacy in the form of knowledge about macronutrients is important for dietary decision making, justifying nutrition knowledge as a reasonable target for improved learning. However, it remains important to acknowledge that nutrition behavior is influenced by multiple factors and knowing the macronutrient composition of different foods is only part of the picture. Other research has considered how crowdsourced and friendsourced plans might improve individuals' behavior change efforts [2].

The results of this study lead to many questions for future research. First, while one can reasonably expect expert explanations to be relatively consistent, we found great variability in quality and helpfulness of peer-generated explanations that could have had an impact on their ability to lead to learning gain. The question becomes what tools and mechanisms can help to identify more effective explanations that could more consistently inspire learning. One important step towards this goal will be to optimize scenarios in which ground truth exists but is not provided. In the study presented here, peer modeling alone did not lead to learning, but peer explanations did. This could have been attributed to the complexity of the task or to the limitations of crowdsourcing. Do peer explanations mitigate the lack of learning gains from peer feedback? Identifying and leveraging the features of effective explanations would be an important aspect of this research. Recently, techniques have been developed to promote the accuracy of crowdsourced responses through identification of complex explanations and engaging crowd workers in assessment of logical reasoning [15]. A second line of research might explore what happens when there is no ground truth. Nutrition and health is a particularly interesting area in which to study this because of nutrition's reputation of being ever-changing; the experts themselves don't always agree. In this study, we examined macronutrient content specifically because that information that is relatively objective. Applying these methods in a more subjective context has potential to contribute to knowledge about how observational learning about nutrition happens in social computing context.

While strengths of this research included the large sample sizes and randomized, controlled design, it also had several limitations that are important to acknowledge. First, it is important to note that because of the nature of the platform on which the study was conducted, only people who were motivated to learn about their own nutrition knowledge participated. Second, NKT was only administered in English, so it is impossible to account for cultural differences of non-English speakers. Third, because we examined predefined meals, it is also impossible to extrapolate how increases in nutrition literacy during NKT translate to participants' ability to apply this literacy to their own typical diet, nor do we know how long the effects of

the intervention might last. Finally, because this study specifically examined nutrition knowledge, we do not know if the increases in nutrition knowledge demonstrated would have an influence on nutrition behavior in the form of dietary choices.

CONCLUSIONS

In this paper, we assessed the impact of peer- and expert-modeling on uncompensated participants' accuracy in assessing the macronutrient content of meals in an online casual learning context. Volunteers who chose to take the NKT were asked to compare pairs of realistic meal photographs and evaluate which one had more of the indicated macronutrient. For some meal photograph pairs, participants received no feedback, for others they received correctness feedback generated by an expert RDN, explanations by the RDN about why the correct choice was correct, a visual representation of the percentage of other participants who had chosen each answer, or an explanation of the correct response provided by another participant. In exchange, they received personalized feedback about their nutrition knowledge relative to other participants after the activity. Our studies demonstrated that when learning involved objective information (i.e., a true answer exists), 1) explanations were most effective in eliciting learning and 2) peer- and expert-generated explanations were equally effective. We conclude that potential exists for casual observational learning about nutrition in a social computing context.

ACKNOWLEDGMENTS

The authors acknowledge funding from NSF Smart and Connected Health (Grant No. 1551708, PI Mamykina). Dr. Burgermaster's involvement in the preparation of this manuscript was supported by an NIH training grant (T15LM007079).

REFERENCES

1. Sofiane Abbar, Yelena Mejova, Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through Twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3197-3206. <http://doi.acm.org/10.1145/2702123.2702153>
2. Elena Agapie, Lucas Colusso, Sean A. Munson and Gary Hsieh. 2016. Plansourcing: Generating behavior change plans with friends and crowds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 119-133. <http://doi.acm.org/10.1145/2818048.2819943>
3. Lisa R. Anderson and Charles A. Holt. 2008. Information cascade experiments. *Handbook of Experimental Economics Results*, 335-343. [http://dx.doi.org/10.1016/S1574-0722\(07\)00039-X](http://dx.doi.org/10.1016/S1574-0722(07)00039-X)
4. American Dietetic Association. 2011. Nutrition and you: Trends 2011 survey, public opinion on food and

- nutrition: 20 years of insights. Retrieved September 15, 2016 from <http://www.eatright.org/nutritiontrends>
5. American Public Health Association. 2016. What is Public Health? Retrieved September 15, 2016 from www.apha.org
 6. Albert Bandura Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1: 1-26.
 7. Albert Bandura, Richard H. Walters. 1977. *Social learning theory*. Prentice Hall, Oxford, England.
 8. Nancy D. Berkman, Stacey L. Sheridan, Katrina E. Donahue, David J. Halpern and Karen Crotty Low. 2011. health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155, 2: 97-107.
 9. Benjamin Samuel Bloom. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: D. McKay Co., Inc.
 10. Elena T. Carbone and Jamie M. Zoellner. 2012. Nutrition and health literacy: A systematic review to inform nutrition research and practice. *Journal of the Academy of Nutrition and Dietetics*, 112, 2: 254-265.
 11. Yubo Chen, Qi Wang and Jinhong Xie. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, 48, 2: 238-254.
 12. Isobel R. Contento. 2007. *Nutrition Education: Linking Research, Theory, and Practice*. Burlington, MA: Jones & Bartlett Learning.
 13. Edward L. Deci and Richard M. Ryan. 2010. Intrinsic motivation. In *Corsini Encyclopedia of Psychology* (4th ed.), Irving B. Weiner and W. Edward Craighead (eds.). Hoboken, NJ: John Wiley and Sons, 1-2.
 14. Allison E. Doub, Aron Levin, Charles Edward Heath and Kristie LeVangie. 2015. Mobile app-etite: Consumer attitudes towards and use of mobile technology in the context of eating behaviour. *Journal of Direct, Data and Digital Marketing Practice*, 17, 2: 114-129.
 15. Ryan Drapeau, Lydia B. Chilton, Jonathan Bragg and Daniel S. Weld. 2016. Microtalk: Using argumentation to improve crowdsourcing accuracy. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*.
 16. Daniel A. Epstein, Felicia Cordeiro, James Fogarty, Gary Hsieh and Sean A. Munson. 2016. Crumbs: Lightweight daily food challenges to promote engagement and mindfulness. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*.
 17. Feng Gao, Enrico Costanza and mc schraefel. 2012. Honey= sugar means unhealthy: Investigating how people apply knowledge to rate food's healthiness. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 71-80. <http://doi.acm.org/10.1145/2370216.2370228>
 18. Laura Germine, Ken Nakayama, Bradley C. Duchaine, Christopher F. Chabris, Garga Chatterjee and Jeremy B. Wilmer. 2012. Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, 19, 5: 847-857.
 19. Laura Germine, Ken Nakayama, Eric Loken, Bradley Duchaine, Christopher Chabris, Garga Chatterjee and Jeremy Wilmer. 2010. Downloadable science: Comparing data from internet and lab-based psychology experiments. *Journal of Vision*, 10, 7: 682-682.
 20. Heather Gibbs and Karen Chapman-Novakofski. 2012. A review of health literacy and its relationship to nutrition education. *Topics in Clinical Nutrition*, 27, 4: 325-333.
 21. Ralph Haefner. 1932. Casual learning of word meanings. *The Journal of Educational Research*, 25, 4-5: 267-277.
 22. Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*: 65-70.
 23. Yuheng Hu, Lydia Manikonda and Subbarao Kambhampati. 2014. What we instagram: A first analysis of instagram photo content and user types. In *Proceedings of the International AAAI Conference on Web and Social Media*.
 24. Angela A. Hung and Charles R. Plott. 2001. Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *The American economic review*, 91, 5: 1508-1520.
 25. Jean-Fabrice Lebraty and Katia Lobre-Lebraty. The dangers of crowdsourcing. 2013. The dangers of crowdsourcing. In *Crowdsourcing*. Hoboken, NJ: John Wiley & Sons, 97-100.
 26. Lena Mamykina, Andrew D. Miller, Yevgeniy Medynsky, Catherine Grevet, Patricia R. Davidson, Michael A. Terry, Elizabeth D. Mynatt. 2011. Examining the Impact of Social Tagging on Sensemaking in Nutrition Management, in the Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2011 <http://doi.acm.org/10.1145/1978942.1979037>
 27. Lena Mamykina, Thomas N. Smyth, Jill P. Dimond and Krzysztof Z. Gajos. 2016. Learning from the crowd: Observational learning in crowdsourcing communities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2635-2644. <http://doi.acm.org/10.1145/2858036.2858560>

28. Michael K. Paasche-Orlow and Michael S. Wolf. 2007. The causal pathways linking health literacy to health outcomes. *American Journal of Health Behavior*, 31, Supplement 1: S19-S26.
29. Katharina Reinecke and Krzysztof Z. Gajos. Labyrinthwild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1364-1378. <http://doi.acm.org/10.1145/2675133.2675246>
30. S. C. Ratzan and R. M. Parker. 2000. Introduction. In *National Library of Medicine Current Bibliographies in Medicine: Health Literacy*, National Institutes of Health, U.S. Department of Health and Human Services., Bethesda, MD.
31. D. Schacter, D. Gilbert, D. Wegner and M.K. Nock. 2014. *Psychology*. New York: Worth Publishers.
32. Tracy L. Scott, Julie A. Gazmararian, Mark V. Williams and David W. Baker. 2002. Health literacy and preventive health care use among medicare enrollees in a managed care organization. *Medical Care*, 40, 5: 395-404.
33. Sanket Sharma and Munmun De Choudhury. 2015. Detecting and characterizing nutritional information of food and ingestion content in instagram. In *Proceedings of the 24th International Conference on World Wide Web*, 115-116. <http://doi.acm.org/10.1145/2740908.2742754>
34. Inge Spronk, Charina Kullen, Catriona Burdon and Helen O'Connor. 2014. Relationship between nutrition knowledge and dietary intake. *British Journal of Nutrition*, 111, 10: 1713-1726.
35. Roumen Vesselinov and John Grego. 2012. Duolingo effectiveness study. City University of New York,
36. Luis von Ahn. 2013. Duolingo: Learn a language for free while helping to translate the web. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, 1-2. <http://doi.acm.org/10.1145/2449396.2449398>
37. Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi and Thomas W. Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330, 6004: 686-688.
38. Keiji Yanai and Yoshiyuki Kawano. 2014. Twitter food photo mining and analysis for one hundred kinds of foods. In *Advances in Multimedia Information Processing – PCM 2014: 15th Pacific-Rim Conference on Multimedia, Kuching, Malaysia, December 1-4, 2014, Proceedings*, 22-32. http://dx.doi.org/10.1007/978-3-319-13168-9_3
39. Juanjuan Zhang and Peng Liu. 2012. Rational herding in microloan markets. *Management science*, 58, 5: 892-912.
40. Jamie Zoellner, Carol Connell, Wendy Bounds, LaShaundra Crook and Kathy Yadrick. 2009. Peer reviewed: Nutrition literacy status and preferred nutrition communication channels among adults in the lower mississippi delta. *Preventing Chronic Disease*, 6, 4: A128.