Leveraging the Contributory Potential of User Feedback

Mikhil Masli
IBM
1 New Orchard Road, Armonk, NY, USA
mnmasli@us.ibm.com

Loren Terveen
GroupLens Research
University of Minnesota, Minneapolis, MN, USA
terveen@cs.umn.edu

ABSTRACT

Under contribution is an important problem in online social production communities: important tasks don't get done, and only a small minority of participants are active contributors. How can we remedy this situation? We explore the feasibility of using the act of *consuming* information as a gateway to *contributing* information; specifically, we investigate semi-automated means to extract useful information from standard types of *user feedback*. We explore this approach in the context of a geographic wiki and route-planning system for bicyclists. We analyzed naturally occurring textual route feedback, finding that the feedback was rich in information such as bikeability ratings, tags and notes that are useful to improve the system's route finding and navigational assistance capabilities. We also present a technique to extract such information by engaging users in dialogue immediately after they obtain a route. We believe that our results and ideas are applicable to a broad class of social production systems.

Author Keywords

User feedback; under-contribution; increasing participation; social production; online communities

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Human Factors, Design

INTRODUCTION

In social production communities, large numbers of people freely collaborate to produce common goods [2]. In recent times, online social production communities have become a part and parcel of our daily lives. For example, breaking news now appears on Twitter and Wikipedia, household items are sold and exchanged on Craigslist and eBay, and available parking spaces, cycling and running routes are shared via Google Maps-based mash-ups.

However, *under contribution* is a major problem for such communities. For example, 1,661 Wikipedia articles were

marked as needing expansion in April 2013 alone[1], and 87.4% of collaboratively-made Flash games and animations on the site Newgrounds[2] are incomplete [14].

Consumers of socially produced knowledge resources greatly outnumber contributors to these resources: for example, readers of Wikipedia are estimated to outnumber Wikipedia editors by a ratio of 10,000:1 [10]. Research on other online groups reports that about 90% of the total membership of the online group never makes a contribution [12, 18].

Our fundamental hypothesis is that we can harness information consumption activities as a gateway to elicit contributions and perhaps develop new contributors. We explore this hypothesis in the context of Cyclopath (`cyclopath.org`), a bicycle route-planner and geographic wiki for the Minneapolis/St. Paul metropolitan area of the United States.

In the remainder of this paper: we survey relevant related work and provide a brief introduction to Cyclopath; we then report on a study that shows that naturally occuring user feedback on routes contains potentially useful information; in a second study, we introduce a semi-automated technique for obtaining useful information from another common type of user feedback, visual route modification; we conclude by discussing the generality, benefits, and limitations of our work, and directions for future work.

RELATED WORK

Nature of Online Social Production

A majority of contribution in online social production communities is done by a few contributors [25]. Most participants are information consumers. Many will never contribute, while others may be learning about the community [21] and may transition into contribution and organization-related activities as they gain more experience [24].

An important motivation to contribute is to fill gaps or fix problems; in other words, to improve information consumption. For example, Bryant et al. found that new users primarily use Wikipedia for information gathering, and identify problems and mistakes in passing and fix them [3]. In the context of open source software development, Hertel et al. discovered that one reason people contributed code was to make the software meet their own needs [11]. Prior research on Cyclopath revealed that a major reason for users to edit the map and rate road segments is to get Cyclopath to compute the route they desire [23]. However, there are barriers to contributing information, including: (a) a perception that one has

---

[1] `en.wikipedia.org/wiki/Category:Articles_to_be_expanded_from_April_2013`

[2] `www.newgrounds.com`

no information to contribute, and (b) the combination that contribution is not required and that it is perceived to be a high-cost activity [22].

A related body of research has studied ways to motivate users to contribute information to public resources. Intelligent Task Routing systems match users with tasks they are likely to be willing and able to perform; these systems have been shown to boost contributions [5, 26]. However, matching tasks to users accurately is challenging. *Gamification* is a popular technique that attempts to make activities fun and game-like; the ESP Game is the seminal example [33]. Finally, techniques that manipulate users into participating and contributing information may succeed in the short-term but might cause longer-term harm, because users tend to recognize the manipulations and may consider them unfair [6, 17].

Nature of User Feedback

Several instances of prior research and experience have highlighted the utility of user feedback. First, user feedback can contain useful content. It is well known in the field of user interface design, that users are better able to communicate interface requirements when presented with prototypes, even if they are low-fidelity. Various research studies and practical use cases have demonstrated the utility of employing the *recognition-over-recall* principle for effectively capturing user requirements and reactions to interfaces [20].

Second, user feedback on the output of a system can be used to improve the system itself. Chen and Pu [4] suggest that a combination of system-generated and user-driven critiquing is the best way for a recommender system to incorporate user feedback and improve its recommendations. An evaluation of a simple tool on Wikipedia demonstrated that consumers can be persuaded to make small contributions in a low-cost, low-risk way [10] through the medium of providing feedback on the article they are reading. Several articifial intelligence and machine learning techniques rely on user feedback to improve their internal models and performance, including the well-known Winston Learning Algorithm which does so through examples and counter-examples [34].

Third, the cognitive processes behind generation of feedback are often automatic and hence, low-cost for people. When presented with a stimulus, natural human tendency is to develop an evaluation about it. To some extent, this construction of evaluation is automatic and happens without one's knowledge. This is due to the use of associative/heuristic-based cognitive information processing [7, 29]—one of the two channels described by dual-processing models of cognition. Evaluations that are more "intuitive" or affective, involving how one subjectively feels about the stimulus, appear to be more associatively driven, compared to more analytic, rational judgments such as those about causation [8]. Further, it seems likely that more richly detailed, specific stimuli are better cues for responses from the associative system [7].

Prior work suggests our opportunity: a pathway from naturally-occurring evaluation and feedback processes to contributions of information has the potential to overcome barriers to contribution and shortcomings of previous work elicitation techniques.

THIS RESEARCH: CONTEXT AND OUTLINE

Cyclopath is a route-planning system for bicyclists that generates personalized, bicycle-friendly routes. As of May 25, 2013, Cyclopath had 5,128 registered users, visits from 196,263 different IP addresses and 128,330 routes requested. However, unlike traditional route-planning systems, Cyclopath is also a geographic wiki—a map that anyone can edit—and relies on community knowledge to compute these routes. Cyclopath went live in the summer of 2008.

Like any wiki, users contribute information to Cyclopath in sets of edits called revisions[3]. Each revision may contain edits to points, roads, regions, notes and tags. Roads and trails (e.g. *Kennilworth Trail*) are useful for route-finding, points help define route endpoints (e.g., from *CS Building* to *Lagoon Theater*), ratings enable smart and personalized routes, tags enable more customized route requests (e.g., a route that is more *scenic*, but less *bumpy*) [31], and notes supplement road segments included in a route with subjective information useful for evaluating the route (e.g., "icy during the winter"). In order to maintain quality, all revisions are publicly visible and revertible (via the Recent Changes list). As of May 25, 2013, the Cyclopath map has seen 17,559 revisions (13,892 by 999 registered users and 3,667 anonymously). Prior research has demonstrated that user-contributed information has helped improve route-finding [26] as well as transportation planning tasks [15].

We instantiate our idea of a pathway from naturally occurring feedback to contribution in Cyclopath as follows. After obtaining a route, user can submit feedback about the route—including specific segments—while their evaluation of the route is fresh in their minds. We present two studies as evidence for this method.

Study 1. We analyzed naturally occurring textual route feedback, finding that it contains contributory potential—information that, through appropriate techniques, may be assimilated into Cyclopath's database.

Study 2. We studied one such technique, user modification of a computed route by "dragging" the route. We showed that machine learning algorithms can identify automatically part of the contributory potential; this in turn makes it easier for users to translate their route evaluation into concrete contributions such as ratings, tags and notes.

STUDY 1: ANALYZING TEXTUAL FEEDBACK

Aim: To determine whether naturally occurring route feedback contains contributory potential

Data

From May 2009 to January 2012, Cyclopath included a simple route feedback tool. When a user obtained a route from

---

[3]See [27] for a more in-depth description.

Question (Does this feedback...)	% Yes	N
...include positive evaluations about any roads/areas?	24%	394
...include negative evaluations about any roads/areas?	48%	392
...include any objective facts about any roads, their surroundings and/or vehicles on them?	51%	399
...suggest any alternative roads or areas to take?	39%	440

Table 1. Findings from our textual route feedback analysis: the majority of naturally occuring route feedback contains contributory potential. Combining the first two rows, 57% of comments contained either a positive or negative evaluation. The N varies across questions and is less than 488 because we ignored responses that did not meet our 67% agreement threshold for that question.

Cyclopath, a button was displayed next to the route details, inviting feedback about the route. When the user clicked this button, a simple form popped up that asked them to tell us how satisfied they were with route on a 5-point Likert scale (very dissatisfied to very satisfied) and optionally, express in text, what they thought of the route.

While this feature was active, 688 instances of route feedback were submitted, 488 of which had textual comments. The median comment length was 137 characters ($mean = 169.020, sd = 132.256$). 120 comments were submitted by 90 registered users, and 368 comments were submitted by anonymous users.

Method
We developed a coding scheme to evaluate the contributory potential contained in the route feedback comments. For each comment we asked: *Does this feedback...*

1. ...include positive evaluations about any roads/areas?

2. ...include negative evaluations about any roads/areas?

3. ...include any objective facts about any roads, their surroundings and/or vehicles on them?

4. ...suggest any alternative roads or areas to take?

These questions let us identify specific types of information useful for Cyclopath: positive/negative evaluations correspond to bikeability ratings of road and trail segments, and objective information corresponds translate to tags and notes. We also looked for the occurrence of alternative route suggestions since this is an intuitive way of expressing route feedback used on popular mapping sites like Google Maps.

We coded the route feedback via crowdsourcing, using CrowdFlower[4] to deploy the coding task on Amazon Mechnical Turk. CrowdFlower provides a service that augments Mechanical Turk's platform with survey-building interfaces, quality-control mechanisms and reporting tools. As a quality-control measure, we created 32 artifical route route feedback comments with known answers ("gold standard items") and added them to the set of 488 actual comments to be coded, resulting in a dataset of 520 total items[5]. We omitted responses submitted by workers whose accuracy on the gold standard items (their "trust score") was less than 70%.

Our goal was 3 responses per question per item, but due to the way Crowdflower manages the process, we received an average of 3.14 responses per item. We only considered responses

where there was at least 67% agreement between workers, where responses were weighted by the workers' trust scores. For example, if three workers had the same trust scores, then 67% agreement would mean a simple 2-out-of-3 majority. See the CrowdFlower documentation[6] for more detail on their quality control mechanisms.

Findings
Naturally occurring route feedback contains evaluations of roads/areas. 24% of comments contained positive evaluations about roads/areas, 48% contained negative evaluations, and 57% contained one or the other or both.

> "Totally smooth, pleasant route, not much interaction with cars, felt very safe." (positive evaluation)

> "Biking straight down Ramsey Hill from Summit Ave to Grand Ave is an awful idea and I would never bike that way. It is a very steep hill with a stop light at the bottom. No way." (negative evaluation)

Naturally occurring route feedback contains objective facts about roads, etc. 51% of comments contained such information.

> "The part of the route from 3.65 to 5.29 is wretched. The bicycle path (in the directions, it's called "unnamed bicycle path") is downright dangerous, too. The path crosses two sets of railroad tracks and the seams are dangerously deep and wide. In addition, the path sweeps down onto a driveway to an industrial site, and that driveway is filled with grit and rubble. I take my chances on Country Road C, which is not great, either, but at least I avoid the wheel-swallowing railroad tracks."

Naturally occurring route feedback contains descriptions of alternative routes. 39% of comments contained such information.

> "A better route is getting on E River Road to Franklin Ave Bridge, cross the River, get on West River Road, tkae that to Greenway Trail and follow all the way to The Depot. Too much traffic on Cedar Ave."

Non-contributors offer feedback. We wondered whether most route feedback came from people who already contributed to Cyclopath; this would be disappointing, as it would decrease the chance of attracting new contributors through a route-feedback-based mechanism. However, this was not the case. First, as already noted, 368 of 488 route feedback comments came from anonymous users. And

---

[4] **crowdflower.com**

[5] The recommended amount of quality control items is about 5% of the data set size.

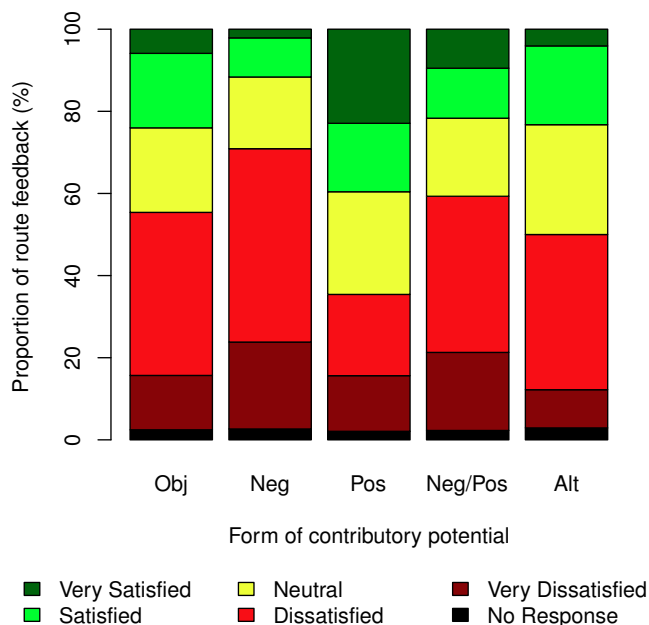[6] **crowdflower.com/solutions/self-service/faqs**

Figure 1. Relationship between users' satisfaction with routes and the contributory potential of their route feedback. Obj = Objective Information, Neg = Negative Evaluation, Pos = Positive Evaluation, Alt = Alternative Route Suggestion. It is not only the dissatisfied who submit useful feedback.

anonymous users account for a small proportion of contribution to Cyclopath (3,667 out of 17,559 revisions, or 21%). Therefore, we can infer that most of the anonymous route feedback came from non-contributors. Second, route feedback comments from anonymous users were more likely to include contributory potential than comments from registered users: negative evaluations (52% vs 37%, $p = .012$), objective information (55% vs 39%, $p = .007$), alternative routes (43% vs 28%, $p = .010$)[7]. Third, only half of the 90 registered users who submitted comments had contributed to Cyclopath. Therefore, route feedback offers a significant opportunity to obtain contributions from new sources.

Not only dissatisfied users provide useful feedback. Most useful content came from users who said they were dissatisfied with their routes. However, users who were satisfied or very satisfied with their routes accounted for 22% of comments with positive or negative evaluations of roads, 23% of comments with alternative route suggestions, and 24% of comments with objective information (see Figure 1).

Discussion and Implications
We found that naturally occurring route feedback is rich in several kinds of useful information. How can that information be mapped on to Cyclopath constructs?

Evaluations → Ratings. Evaluations of roads can be expressed as bikeability ratings. A simple ratings widget could be displayed along with a route, enabling users to quickly identify segments they want to rate higher or lower.

---

[7]To compute statistical significance, we used the 2-sample test for equality of proportions with continuity correction.

Objective Facts → Tags, Notes. Cyclopath represents factual information as tags (short phrases) and notes (longer, freeform text). As with ratings, an appropriate widget could be displayed to enable easy annotation of road segments. Further, techniques like Tag Expression [32] have been shown to be effective in eliciting tags and tag preferences.

Alternative Routes → *???*. When a user specifies an alternative route, there is no direct correspondence to specific Cyclopath constructs. Indeed, various things could be changed in order for Cyclopath to produce a different route: bikeability ratings, tags, or the details of the routing algorithm (e.g., how various parameters are weighted). Instead, an alternative route suggestion—along with the rejection of the original route—constitutes an opportunity for Cyclopath to learn. *How* that learning might be done is the topic of the next study.

STUDY 2: ANALYZING VISUAL FEEDBACK
Aim: To investigate a semi-automated technique for learning from visual route feedback

We have shown that naturally occurring route feedback often includes a suggestion of an alternative route. Further, an intuitive means for users to specify an alternative route is by "dragging" the route, as on Google Maps and Cyclopath. There are two ways a system could respond to this learning opportunity: (1) Automatically modify the route planner's internal models and/or algorithms, thus producing better routes in the future [36, 37, 19]; (2) Engage the user in a dialogue to elicit reasons for preferring the new route over the old route [28]; these reasons then can be analyzed to produce new system knowledge.

We explored the second approach for the following reasons: (1) Richer types of information can be gathered, e.g., notes and tags in addition to ratings. Indeed, in the context of recommender systems, researchers have recommended a hybrid system-and-user-generated critiquing approach to improving system performance [4]. (2) The consumer is introduced to and nudged towards contribution through simple, easy steps. This is consistent with theories such as Legitimate Peripheral Participation [13] and the Reader-to-Leader framework [24] that model the transition from consumption to contribution. These approaches have been demonstrated to have value in the context of Wikipedia [10]. Thus, the result of a semi-automated, dialogue approach may be not just a *contribution*, but also a new *contributor*.

The following use case illustrates the type of dialogue we have in mind.

1. A user requests a route.

2. Cyclopath computes the best path based on available information.

3. The user is not satisfied with the route, and corrects it by dragging the route line to follow their preferred path.

4. Cyclopath automatically identifies segments along the original and new routes that it thinks caused the user to modify the route, highlights the segments, and offers the

user the opportunity to provide ratings, tags, and notes for the segments. The user responds as he/she desires.

This use case is plausible under three conditions: (1) The user's route preference is based on preferences for individual segments of the route. (2) The system can identify the appropriate route segments with sufficient accuracy. (3) The user will provide information about these segments. We consider each of these in turn.

Users might prefer one route over another because of properties of individual segments of the routes: for example, they might want to avoid hilly or high traffic segments and instead ride on quiet, scenic off-road facilities. Alternatively, users might prefer one route over another for more holistic reasons, for example, to ride near certain resources (say, places to stop for a drink). Given the way our route-finder works, the former preference is more directly useful. We use the A* graph search algorithm, and at each step in the search process, the system considers only the properties of the possible next segments that could be added to the route (and not global properties like "make sure some part of the route passes within a kilometer of a place to get a drink"). Thus, we conducted a study to explore the reasons why users prefer one route over another, specifically, the extent to which they do so because of properties of specific segments (Study 2A).

If users do tend to prefer one route over another because of the properties of specific segments, then the next challenge is for the system to identify these segments. We therefore developed a machine learning classifier to predict such segments and evaluated its accuracy (Study 2B).

But even if the system can identify these segments—segments in an old route that a user is apt to dislike, segments in the user's suggested route that the user is apt to like—why would a user go to the trouble of providing knowledge (ratings, tags, notes) about them? Besides the likelihood that the user will have this information fresh in their mind, prior work offers compelling answers. As we noted above, fixing problems and improving one's own experience are common reasons people begin contributing to a social production system; Panciera et al. found that fixing problems was the most common reason for Cyclopath users to begin editing [23]. Moreover, prior work on Cyclopath has shown that focusing users' attention on specific geographic areas makes them more likely to contribute information [26]. Given this prior work, we did not study user motivation further in this research.

Data and Method

We modified the Cyclopath interface to elicit visual route feedback. When a user generated a route using Cyclopath's route planner, a button was displayed in the route details panel labeled "Suggest Different Route". Clicking this button led to the user being guided through a two-step procedure.

In step 1, users were asked to drag the route to meet their preferences. (Route dragging was an existing feature of Cyclopath, which works similarly to other mapping applications such as Google Maps: clicking on the route line introduces an intermediate waypoint at the location of the click and dragging it changes the route line in such a way that it is forced to pass through the waypoint.)

In step 2, users were asked to provide specific feedback on the original route and their suggested route, including identifying specific segments of either route that most influenced their preference. These segments constituted a data set that we used to train and evaluate a machine-learning classifier. Users also were asked to estimate how much of an improvement their suggested route was over the route Cyclopath computed. Note that we implemented step 2 solely for the purpose of data gathering—in the wild, users would only be required to drag their routes for Cyclopath to suggest road segments to annotate.

We received 85 instances of route feedback between October 25, 2012 and April 16, 2013. Each feedback instance consisted of the original route, the suggested route, one or more waypoints, zero or more marked segments, optional reasons for changing the original route and suggesting the alternative and an evaluation of how much of an improvement the suggested route was over the original.

In 38 of these 85 cases, users had marked specific segment(s) as reasons for modifying the route. In most cases (36 of 38) where users marked segments, they also provided an explanation for modifying the route. An additional 17 feedback instances carried explanations (for changing the original or for suggesting the alternative) but no marked segments, making a total of 53 feedback instances with explanations.

As in Study 1, most of the feedback was submitted anonymously: 23 instances were submitted by 20 registered users and the remaining 62 anonymously. Anonymous users also provided more feedback than registered ones, accounting for 24 of 36 instances where liked/disliked segments, and 33 of 53 instances where comments were submitted. However, 18 of the 20 registered users had made at least 1 contribution.

Study 2A: Why did the user modify the route?

Aim: To identify why users modified their routes, and specifically, to what extent did they do so because of the segments they marked

Method

We manually coded the comments users specified along with their visual route feedback to to understand why users chose to modify their routes. For the first question, coders assigned one or more of the following sets of codes to the reasons users gave for rejecting the original route and for suggesting the alternative:

- Longer vs. Shorter

- Unsafe vs. Safe

- No dedicated bike facilities vs. Dedicated bike facilities

- Hills vs. No hills

- Bad surface vs. Good surface
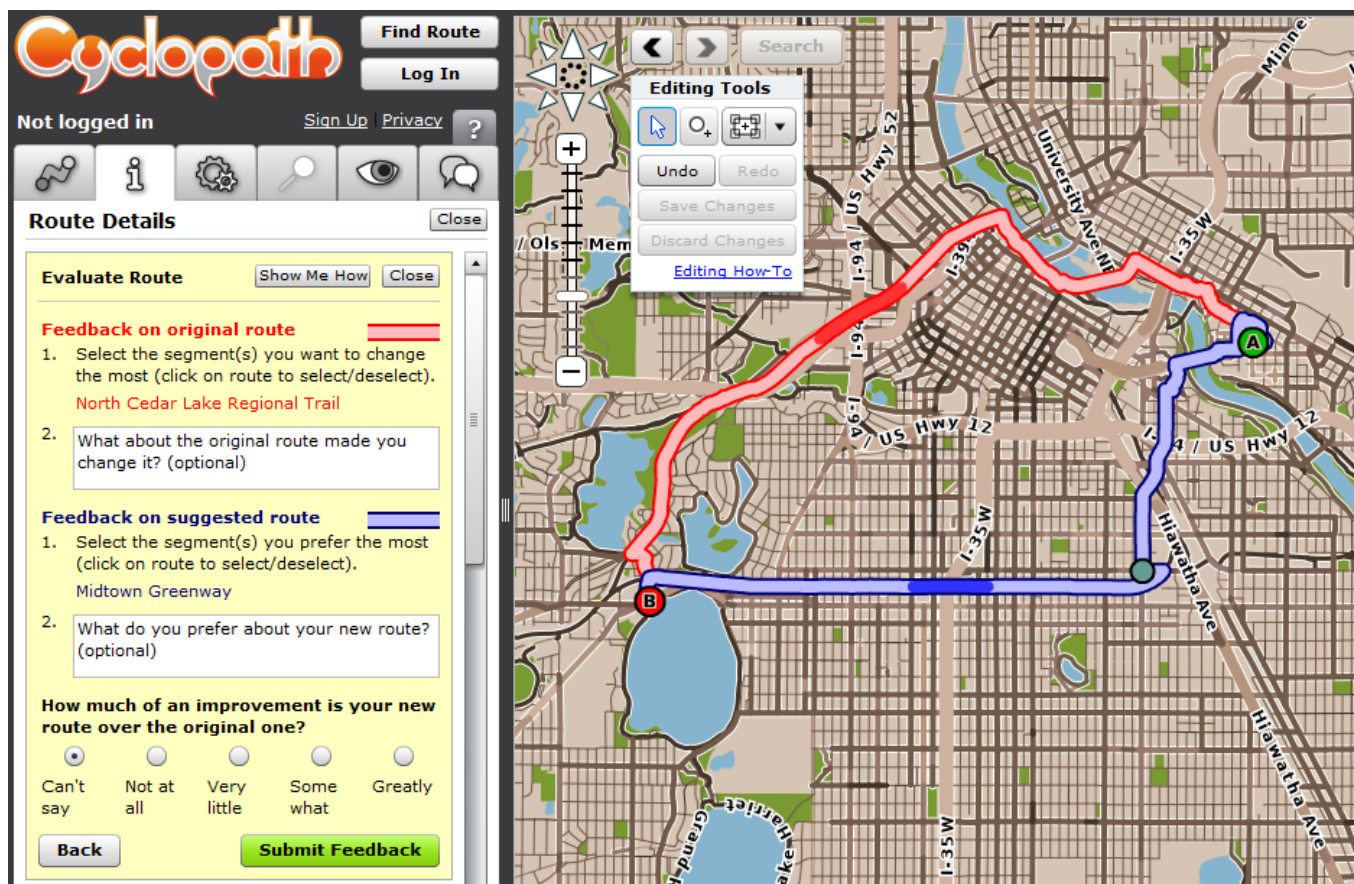
- High traffic vs. Low traffic

Figure 2. The Cyclopath interface to collect visual route feedback through dragging. This images is showing step 2 of the 2-step process where the user is being asked to mark road segment(s) that caused them to modify their route. The "Show Me How" button led users to a tutorial video describing the process.

- Noisy vs. Quiet

- Many stoppages vs. Few stoppages

- Bad intersection vs. Good intersection

One of the authors developed these codes by identifying prominent patterns within the feedback comments. These patterns are consistent with those used by prior research on bicycling route choice [30].

For the second question, we used the following ordered set of rules to code the comments. That is, if Rule 1 let us conclude that the user modified route because of the segment(s) they marked, then rules 2 and 3 were not considered. *We conclude that the user modified the route because of the segments they marked if...*

Rule 1. ...the user mentions a road segment by name in the feedback and has also marked a road segment with the same name. For example, one comment said *"Big hill on North St, too much traffic on University, freeway on/off traffic on 12th St"* and the user marked segments of North St. or 12th St.

Rule 2. ...the feedback content is highly synonymous with the road type and the contents of the tags and notes

attached to the marked segments. To help make this determination, we augmented each feedback instance with road types, tags and notes attached to the corresponding marked segments. For example, a comment said *"This is a long, steep hill."* and the "hill" tag was applied to one of the marked segments.

Rule 3. ...we find factual, objective information with the feedback that we can verify applies to the marked segments using aerial imagery or by browsing around the area. For example, a comment said *"I don't wanrt [sic] to ride through the U [University of Minnesota]. Don't like the traffic and vibe."* and browsing the map around the marked segment reveals that the segments were within the University of Minnesota neighborhood.

Thus, for this coding task, coders assigned one of two codes to user comments: "Yes" (by Rules 1, 2 or 3) or "No evidence". Three independent coders coded the feedback instances for the both the coding tasks.

Findings

Why did users modify the route? We found evidence for several important reasons. The most common reasons for rejecting the original route were *high traffic* (17/53; inter-rater

Attribute	Entire Section				Marked Segments Only			
	Original	Suggested	p	sig	Original	Suggested	p	sig
Avg. annual daily automotive traffic	6240	5250	$< .001$	$* * *$	8030	4300	$< .001$	$* * *$
Shoulder width (m)	0.831	1.16	$< .001$	$* * *$	0.856	1.12	.309	
Outer lane width (ft)	9.91	12.1	$< .001$	$* * *$	12.0	13.5	.0502	$\diamond$
Speed limit (mph)	31.6	31.6	.79		32.0	31.0	.011	$*$
Avg. user bikeability rating	4.09	4.73	$< .001$	$* * *$	2.97	2.96	.943	

Table 2. Differences between the original and suggested route sections. All values are means. Suggested sections had lower automotive traffic, wider shoulders and outer lanes, and higher user-submitted bikeability ratings. All road attributes were loaded into Cyclopath when map data was initially imported from the Minnesota Department of Transportation in 2008 and except average annual daily traffic, the others can be modified and kept up to date by users. P-values and statistical significance codes (sig) are computed using the Welch 2-sample t-test ($\diamond : < .1$, $* : < .05$, $** : < .01$, $*** : < .001$).

agreement for this category using Fleiss' Kappa, $\kappa = 0.85$) and *hilliness* (10/53; $\kappa = 0.91$), whereas those for suggesting the alternative route were *presence of dedicated bicycle facilities* (12/53; $\kappa = 0.79$) and *low traffic* (11/53; $\kappa = 0.79$).

These findings are consistent with prior research on commuter bicyclists: The most important factors in choosing a route are: travel time, the presence of a bicycle facility (bike lane or separate path on a bridge), the level of automobile traffic, quality of the pavement or riding surface [30]. This is also consistent with the data we collected. Table 2 shows quantitative differences between the original and suggested sections of routes. We see that on an average, the suggested section had much lower automotive traffic, wider shoulders and outer lanes, and a higher average user-submitted bikeability rating.

Did users modify the route because of the segment(s) they marked? Yes, in many cases. Marking segments was an optional step of the data collection process—users could successfully submit route feedback by simply dragging the route and doing nothing else. Despite this, in 29 out of 85 cases, users marked segment(s) as a response to the prompt "Select the segment(s) you want to change the most" and in 26 out of 85 cases, users marked segment(s) as a response to the prompt "Select the segment(s) you prefer the most". In sum, in 38 out of 85 cases (45%), users indicated at least one segment as a reason why they modified their route.

Further, combining "Yes (by Rule 1)", "Yes (by Rule 2)" and "Yes (by Rule 3)" into a single code "Yes", we found evidence that the marked segments were a part of the reason why users modified the route in 23 of the 36 instances (64%) where users had marked segments ($\kappa = 0.50$). Note that this is a lower-bound: in the remaining 34% cases, we could not reliably conclude from our manual coding whether the feedback was about the segments marked.

Finally, as a response to the question, "How much of an improvement is your new route over the original one?", we obtained the following results: *can't say* (34 cases), *very little* (3 cases), *somewhat* (26 cases) and *greatly* (22 cases). This suggests that obtaining information about why users preferred one route over another, and then mapping it into a form usable by the route finder will result in demonstrably better routes.

Study 2B: Predicting liked/disliked segments
Aim: Given the original and suggested sections, to predict the segments that the users marked

Method
We adopted a machine learning approach for this task. For our choice of classifier to use, we chose the Random Forest classifier due to its sensitivity towards handling data sets with imbalanced classes. We trained and evaluated this classifier using the Weka data mining software package [35]. We evaluated the classifier using 10-fold cross-validation.

We used several attributes of the road segments in question, the routes (original and suggested versions) to which they belonged, and the user that requested the routes (and submitted the feedback, whenever available):

- **Road attributes:** Type of the road (local road, highway, bike trail, etc.), lane width, lane count, shoulder width, speed limit, average annual daily automotive traffic, number of user-recorded bikeability ratings, average user bikeability rating, Cyclopath-estimated bikeability rating (calculated based on road properties), presence of bike lane (or similar tags such as bike trail), presence of hills, presence of roughness (or similar tags such as unpaved), presence of traffic (or similar tags such as busy).

- **Route attributes:** Length (in metres), number of road segments, average bikeability rating.

- **User properties:** Number of revisions made, number of routes requested, number of map views done, number of days since registering for Cyclopath.

We used three primary metrics in reporting the performance of our machine learning classifiers: sensitivity, specificity, and area under the Receiver-Operator-Characteristic (ROC) curve (also called area-under-curve or AUC) [9]. Our metrics may be interpreted as:

- **Sensitivity:** proportion of marked road segments that are correctly classified

- **Specificity:** proportion of unmarked road segments that are correctly classified

- **AUC:** a single scalar value representing the overall performance of the classifier.

We used the 0-R classifier to provide a baseline model as a frame of reference for interpreting our results. A 0-R algorithm always predicts the most commonly occurring class. Due to the design of our data gathering procedure (the instructions we gave to the users), there will be many fewer marked than unmarked road segments in any given old/new

(a) Given the original section, predict the road segment(s) the user marked.

Attributes Used	Sensitivity	Specificity	AUC
Road	0.346	0.978	0.814
Road + User	0.365	0.987	0.857
Road + Route	0.738	0.984	0.951
Road + User + Route	0.757	0.985	0.962

(b) Given the suggested section, predict the road segment(s) the user marked.

Attributes Used	Sensitivity	Specificity	AUC
Road	0.339	0.983	0.870
Road + User	0.347	0.984	0.883
Road + Route	0.771	0.989	0.964
Road + User + Route	0.792	0.989	0.974

Table 3. Classifier performance. Using all three sets of attributes (road, user and route), we can detect about 76% of marked segment(s) for the original section of the route, and about 79% for the suggested section. The values are average of 10-fold cross-validation. User attributes had a small effect on sensitivity because a majority of the feedback was submitted anonymously.

route pair. So, given an old (or new) route section, the 0-R algorithm will always predict that none of its road segments were marked: sensitivity = 0, specificity = 1, AUC = 0.5. Note that our baseline outperforms random guessing, which would converge to an overall performance of sensitivity = 0.5 and specificity = 0.5.

Findings
Our classifier outperformed the baseline 0-R in all our cross-validation tests. On an average, given the original section of the route, the marked road segments within it were correctly identified in about 76% (sensitivity = 0.757) of the cases when road, route and user attributes were all utilized for prediction (see Table 3(a)).

Similarly, given the suggested section of the route, the marked segments were correctly identified in about 79% (sensitivity = 0.792) of the cases on an average by using all three sets of attributes (see Table 3(b)).

Discussion, Implications and Limitations
In this study, we have presented a technique to leverage the intuitive action of correcting a route when dissatisfied with it to capture the content of the underlying route evaluation. We now address the generality, utility and limitations of this study.

Generality of Results and Ideas
Social production communities such as Wikipedia and Cyclopath offer contributors tasks across a wide spectrum of expertise and difficulty: some tasks are quite easy and do not require topic or work-type expertise (e.g. slightly modifying a route in Cyclopath, correcting a typo in Wikipedia) whereas other tasks can require both (e.g. fixing road segment connectivity around a complex intersection, or composing and editing an entire article in Wikipedia). Prior research suggests that consumers take their first steps into becoming a contributor by performing the simpler tasks [13, 24], a pattern that is observed in both Wikipedia [3] and Cyclopath [16]. Fixing

or tweaking the output that the system computes and presents is one such simple task that consumers have a propensity to do: after all, it brings them direct benefit. Consequently, in Study 2, we explored how route dragging can be used to engage the consumer in a dialogue that may help translate some naturally-occurring route evaluations into concrete contributions such as ratings, tags and notes.

This basic idea is applicable in a wide spectrum of social production systems. For example, when a recommender system presents the consumer with a list of recommended items, the consumer could correct it by rating the items listed or specifying different items. To some extent, Amazon already does this with the "Fix this recommendation" feature. Similarly, another domain where this idea is applicable is the collaborative construction of the semantic web. It may be easier to identify the inaccuracy and correct the output of a information query on a structured knowledge graph, than to identify the holes by browsing through the graph itself.

Estimation of Utility
How useful might the contributions resulting from the system-user dialogue explored in Study 2 be? Study 2A highlighted several reasons users mentioned for modifying their routes. Many of the reasons can be expressed as tags: e.g., hilliness can be represented by the *hill* tag and dedicated bicycle facilities can be recorded using the *bikelane* tag. By automatically predicting a part of users' feedback in the form of especially liked/disliked segments, we can reduce the cost of contribution to just that involved in submitting ratings, tags, and notes.

A more interesting approach is to estimate the potential "conversion rate", i.e., the proportion of route-modification actions that might be leveraged to obtain a contribution. It is precisely those times when users modify computed routes (by dragging) when they are dissatisfied with the route; thus, these consitute opportunities for Cyclopath to learn. Our results let us estimate the conversion rate. To do so, must enumerate and analyze the various steps involved in the process. See Figure 3 for a back-of-the-envelope version of this analysis. Recall that for a route-modification action to be "converted" to a contribution, two things must happen:

1. *The modification must be due to the user liking/disliking specific segments.* Study 2A showed that in 45% of cases when users described an alternative route, they marked segments they liked/disliked; further, we have evidence that at least 64% of the time, segments were marked because of issues with the segments themselves, as opposed to issues with their surroundings. Thus, we that when a user modifies a computed route, the probability that specific road or trail segments caused the modification is 45% × 64% = 29%.

2. *Cyclopath should be able to successfully predict these segments.* In Study 2B, we built a predictor that could identify these segments with a sensitivity of about 76%. In other words, when users modified a route because of specific segments, the classifier had a 76% probability of predicting the segments correctly.
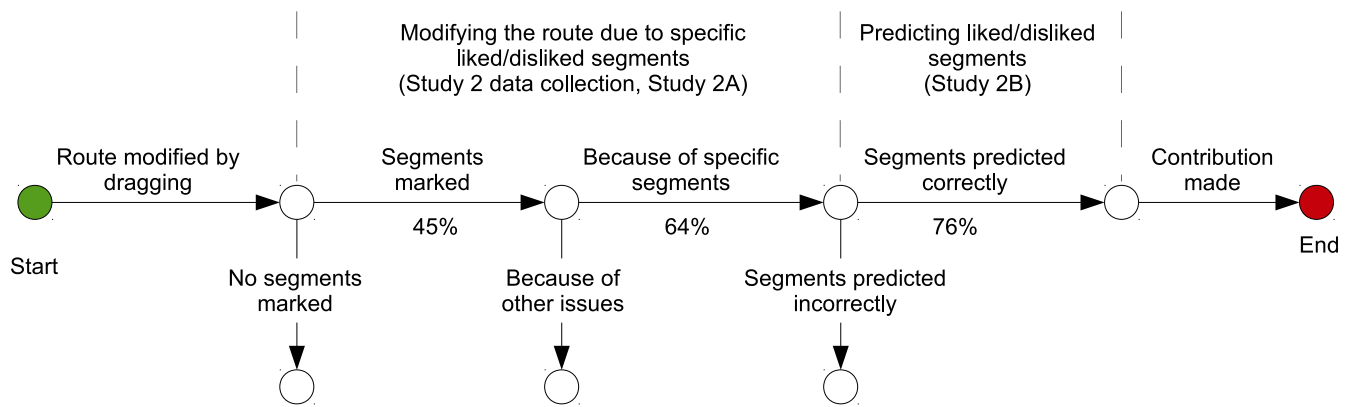
Figure 3. Flowchart illustrating the back-of-the-envelope calculation used to estimate the probability that a user would make a contribution after dragging a route through the technique explored in Study 2. Note that in a real-life use-case, the user would drag their route and then directly contribute information about the predicted segments—the intermediate steps would happen internally (either within the user's mind or within the system).

Thus, if each route-modification action with correctly-predicted segments yielded one contribution, we would get a route-modification-to-contribution conversion rate of about $29\% \times 76\% = 22\%$ (22 contributions per 100 route-modification actions). This may be an underestimate, as prior research has shown that users may contribute beyond what they are asked directly [26]. A controlled field study is required to provide more concrete evidence in this regard.

How sustainable is this process for eliciting contributions? Since users modify routes for self-benefit, we think the first part of the process will continue to be "triggered" until the route finder is perfect for all users... which likely never will happen. And we already have presented an argument why users are likely to provide simple explanations (in the form of ratings, tags, and notes) for why they modified a route.

Moreover, the utility extends beyond getting more *contributions*: we may be able to get more *contributors* too. Indeed, research has shown that having consumers do simple contribution tasks, such as providing feedback, may convert them into regular contributors [10]. This is valuable in and of itself, because an increased inflow of contributors into the community is an important force to combat attrition and increase the diversity and quality of the community resource [1].

Limitations
An important limitation of the process we have outlined for eliciting contributions from a route modification is that it captures only one possible type of preference, i.e., preference for (or against) individual route segments. However, as we noted, it is possible that users may have more holistic preferences. For example, one user gave the following reason for modifying their route: "It brings me by Freewheel Cycle."

Another limitation of this process is that dragging-based feedback depends on the road network being connected correctly. For example, intersection faults (road segments that visually cross but are not connected in the database) can cause routing to work incorrectly. While collecting visual route feedback, we observed several cases where users who had the motivation and skills to make geographic edits, made the required

connections in the road network in order to enable them to drag their route. On the other hand, some users who had the motivation but not the skills wrote to us complaining that they were not able to drag the route on paths they wanted. After investigating why, we found out that some roads were not connected appropriately. We connected these roads and the users were then able to record their feedback correctly.

Some users mentioned this limitation as a part of their feedback comments.

> "Shorter distance, wide shoulders, and new path on east side of Flying Cloud between ValleyView and Technology. Couldn't make the route follow Flying Cloud there..."

> "System doesn't seem to know that both Sibley and Jackson connect to Shepard Rd"

With a modification to the route-dragging mechanism, we believe we can convert this limitation to an advantage. For example, if we made the route-dragging permissive enough to allow the route to treat intersection faults like regular connections, we may be able to either automatically create connections and fix the faults, or at least flag them for user attention.

SUMMARY AND FUTURE WORK
In this research, we have explored the possibility of harnessing a commonly-found information consumption behavior to drive contribution into a social production system. We have done so in the context of Cyclopath, a route-planner and geographic wiki, shown that route feedback contains contributory potential, and outlined a process for focusing it on specific road segments.

Important future work is to investigate further how users choose to express their feedback in the context of a modified route. We might do this by inviting the user to modify their route (like our Study 2) and if the user accepts the invitation, employ our new classifier to automatically predict and highlight road segments that might be the reason the user modified the route. Then, for each highlighted segment, we could ask the user to express feedback in one or more of three

different ways: ratings, tags and notes. We may then be able to study any association between the route modification and the existence and nature of the information elicited.

REFERENCES
1. Arazy, O., Morgan, W., and Patterson, R. Wisdom of the crowds: Decentralized knowledge construction in wikipedia. In *16th Annual Workshop on Information Technologies & Systems (WITS) Paper* (2006).

2. Benkler, Y. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006.

3. Bryant, S. L., Forte, A., and Bruckman, A. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, ACM (2005), 1–10.

4. Chen, L., and Pu, P. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction 22*, 1-2 (2012), 125–150.

5. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, ACM (2007), 32–41.

6. Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., and Riedl, J. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2003), 585–592.

7. Epstein, S. Cognitive-experiential self-theory of personality. *Handbook of psychology* (2003).

8. Epstein, S., Lipson, A., Holstein, C., and Huh, E. Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology 62*, 2 (1992), 328.

9. Fawcett, T. ROC graphs: Notes and practical considerations for researchers. *Machine Learning 31* (2004), 1–38.

10. Halfaker, A., Keyes, O., and Taraborelli, D. Making peripheral participation legitimate: reader engagement experiments in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, ACM (2013), 849–860.

11. Hertel, G., Niedner, S., and Herrmann, S. Motivation of software developers in open source projects: an internet-based survey of contributors to the linux kernel. *Research policy 32*, 7 (2003), 1159–1177.

12. Katz, J. Luring the lurkers, slashdot, 1998.

13. Lave, J., and Wenger, E. *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.

14. Luther, K., Caine, K., Ziegler, K., and Bruckman, A. Why it works (when it works): success factors in online creative collaboration. In *Proceedings of the 16th ACM international conference on Supporting group work*, ACM (2010), 1–10.

15. Masli, M., Bouma, L., Owen, A., and Terveen, L. Geowiki + route analysis = improved transportation planning. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, ACM (2013), 213–218.

16. Masli, M., Priedhorsky, R., and Terveen, L. Task specialization in social production communities: The case of geographic volunteer work. *Proc. ICWSM* (2011).

17. Masli, M., and Terveen, L. Evaluating compliance-without-pressure techniques for increasing participation in online communities. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM (2012), 2915–2924.

18. Mason, B. Issues in virtual ethnography. *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities* (1999), 61–69.

19. McGinty, L., and Smyth, B. Personalised route planning: A case-based approach. In *Advances in Case-Based Reasoning*. Springer, 2000, 431–443.

20. Nielsen, J., and Hackos, J. T. *Usability engineering*, vol. 125184069. Academic press Boston, 1993.

21. Nonnecke, B., and Preece, J. Why lurkers lurk. In *Americas Conference on Information Systems* (2001), 1–10.

22. Nonnecke, B., Preece, J., and Andrews, D. What lurkers and posters think of each other [online community]. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, IEEE (2004), 9–pp.

23. Panciera, K., Masli, M., and Terveen, L. How should I go from _ to _ without getting killed?: motivation and benefits in open collaboration. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, ACM (2011), 183–192.

24. Preece, J., and Shneiderman, B. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction 1*, 1 (2009), 13–32.

25. Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, ACM (2007), 259–268.

26. Priedhorsky, R., Masli, M., and Terveen, L. Eliciting and focusing geographic volunteer work. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 61–70.

27. Priedhorsky, R., and Terveen, L. Wiki grows up: arbitrary data models, access control, and beyond. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, ACM (2011), 63–71.

28. Rogers, S., Fiechter, C.-N., and Langley, P. An adaptive interactive agent for route advice. In *Proceedings of the third annual conference on Autonomous Agents*, ACM (1999), 198–205.

29. Smith, E. R., and DeCoster, J. Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review 4*, 2 (2000), 108–131.

30. Stinson, M. A., and Bhat, C. R. Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation Research Record: Journal of the Transportation Research Board 1828*, 1 (2003), 107–115.

31. Torre, F., Sheppard, S. A., Priedhorsky, R., and Terveen, L. bumpy, caution with merging: An exploration of tagging in a geowiki. In *Proceedings of the 16th ACM international conference on Supporting group work*, ACM (2010), 155–164.

32. Vig, J., Soukup, M., Sen, S., and Riedl, J. Tag expression: tagging with feeling. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 323–332.

33. Von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of the ACM 51*, 8 (2008), 58–67.

34. Winston, P. H. Learning structural descriptions from examples. Tech. rep., DTIC Document, 1970.

35. Witten, I. H., and Frank, E. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

36. Ziebart, B. D., Maas, A. L., Dey, A. K., and Bagnell, J. A. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In *Proceedings of the 10th international conference on Ubiquitous computing*, ACM (2008), 322–331.

37. Zita, K., Jonathan, H., Shewchuk, R., and Veloso, M. M. Exploiting domain geometry in analogical route planning. *Journal of Experimental & Theoretical Artificial Intelligence 9*, 4 (1997), 509–541.