The Seventh AAAI Conference on Human
Computation and Crowdsourcing (HCOMP-19)

AI-Based Request Augmentation to Increase Crowdsourcing Participation

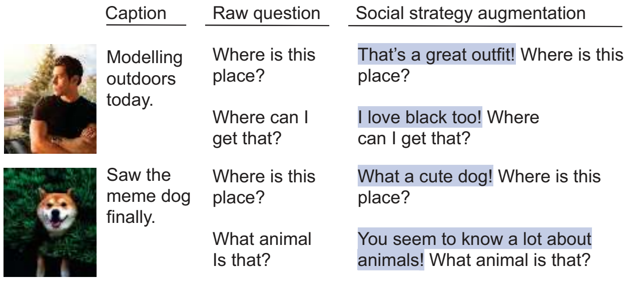
Junwon Park, Ranjay Krishna, Pranav Khadpe, Li Fei-Fei, Michael Bernstein
Stanford University
{junwon, ranjaykrishna, pkhadpe, feifeili, msb}@cs.stanford.edu

Abstract

To support the massive data requirements of modern supervised machine learning (ML) algorithms, crowdsourcing systems match volunteer contributors to appropriate tasks. Such systems learn *what* types of tasks contributors are interested to complete. In this paper, instead of focusing on *what* to ask, we focus on learning *how* to ask: how to make relevant and interesting requests to encourage crowdsourcing participation. We introduce a new technique that augments questions with ML-based request strategies drawn from social psychology. We also introduce a contextual bandit algorithm to select which strategy to apply for a given task and contributor. We deploy our approach to collect volunteer data from Instagram for the task of visual question answering (VQA), an important task in computer vision and natural language processing that has enabled numerous human-computer interaction applications. For example, when encountering a user's Instagram post that contains the ornate Trevi Fountain in Rome, our approach learns to augment its original raw question "Where is this place?" with image-relevant compliments such as "What a great statue!" or with travel-relevant justifications such as "I would like to visit this place", increasing the user's likelihood of answering the question and thus providing a label. We deploy our agent on Instagram to ask questions about social media images, finding that the response rate improves from 15.8% with unaugmented questions to 30.54% with baseline rule-based strategies and to 58.1% with ML-based strategies.

Introduction

Modern supervised machine learing (ML) systems in domains such as computer vision are reliant on mountains of human-labeled training data. These labeled images, for example the fourteen million images in ImageNet (Deng et al. 2009), require basic human knowledge such as whether an image contains a chair. Unfortunately, this knowledge is both so simple that it is extremely tedious for humans to label, and also so tacit that the human annotators are required. In response, crowdsourcing efforts often recruit volunteers to help create labels via intrinsic interest, curiosity or gamification (Lintott et al. 2008; Law et al. 2016; Willis et al. 2017; von Ahn and Dabbish 2004a).



Caption	Raw question	Social strategy augmentation
Modelling outdoors today.	Where is this place?	That's a great outfit! Where is this place?
	Where can I get that?	I love black too! Where can I get that?
Saw the meme dog finally.	Where is this place?	What a cute dog! Where is this place?
	What animal Is that?	You seem to know a lot about animals! What animal is that?

Figure 1: We introduce an approach that increases crowdsourcing participation rates by learning to augment requests with image- and text-relevant question asking strategies drawn from social psychology. Given a social media image post and a question, our approach selects a strategy and generates a natural language phrase to augment the question.

The general approach of these crowdsourcing efforts is to focus on *what* to ask each contributor. Specifically, from a large set of possible tasks, many systems formalize an approach to route or recommend tasks to specific contributors (Geiger and Schader 2014; Lin, Kamar, and Horvitz 2014; Ambati, Vogel, and Carbonell 2011; Difallah, Demartini, and Cudré-Mauroux 2013). Unfortunately, many of these volunteer efforts are restricted to labels for which contributions can be motivated, leaving incomplete any task that is uninteresting to contributors (Reich, Murnane, and Willett 2012; Hill 2013; Healy and Schussman 2003; Warncke-Wang et al. 2015).

Our paper specifically studies an instantiation of this common ailment in the context of visual question answering (VQA). VQA generalizes numerous computer vision tasks, including object detection (Deng et al. 2009), relationship prediction (Lu et al. 2016), and action prediction (Niebles, Wang, and Fei-Fei 2008). Progress in VQA supports the development of many human-computer interaction systems, including VizWiz (Bigham et al. 2010), TapTapSee, BeMyEyes, and CamFind[1]. VQA is a data-hungry machine

[1] Applications can be found at https://taptapsee.com/, https://www.bemyeyes.com/, and https://camfindapp.com/

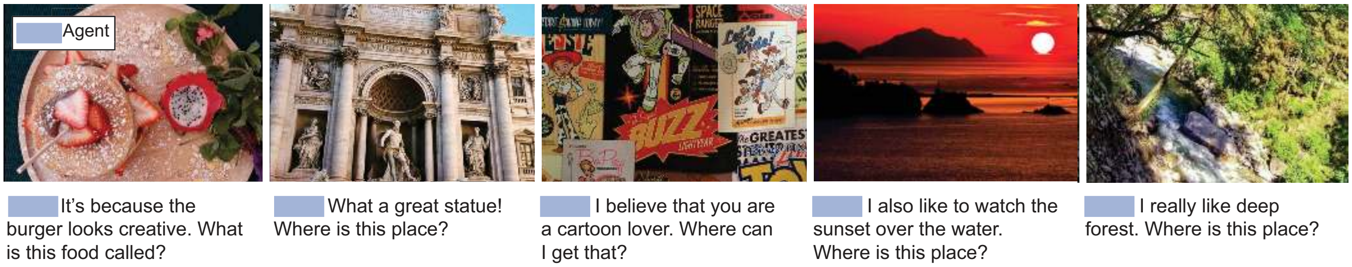

Agent
It's because the burger looks creative. What is this food called?
What a great statue! Where is this place?
I believe that you are a cartoon lover. Where can I get that?
I also like to watch the sunset over the water. Where is this place?
I really like deep forest. Where is this place?

Figure 2: Our agent chooses appropriate social strategies and contextualizes questions to maximize crowdsourcing participation.

learning task that is challenging to motivate contributors. Existing VQA crowdsourcing strategies have suggested using social media to incentivize online participants to answer visual questions for assistive users (Bigham et al. 2010; Brady, Morris, and Bigham 2015), but many such questions remain unanswered (Brady et al. 2013).

To meet the needs of modern ML systems, we argue that crowdsourcing systems can automatically generate plans not just for *what* to ask about, but also for *how* to make that request. Social psychology and social computing research have made clear that how a request is structured can have substantial effects on resulting contribution rates (Kraut and Resnick 2011; Yang and Kraut 2017). However, while it is feasible to manually design a single request such as one email message to all users in an online community, or one motivational message on all web pages on Wikipedia, in real life (as in VQA) there exist a wide variety of situations that must each be approached differently. Supporting this variety in *how* a request is made has remained out of reach; in this paper, we contribute algorithms to achieve it.

Consider, for example, that we are building a dataset of images with their tagged geolocations (Figure 1). When we encounter an image of a person wearing a black shirt next to a beautiful scenery, existing machine learning systems can generate questions such as "where is this place?". However, prior work reports that such requests seem mechanical, resulting in lower response rates (Brady et al. 2013). In our approach, requests might be augmented by `content compliment` strategies (Robert 1984) reactive to the image content, such as "What a great statue!" or "That's a beautiful building!", or by `interest matching` strategies (Cialdini 2016) reactive to the image content, such as "I love visiting statues!" or "I love seeing old buildings!"

Augmenting requests with social strategies requires (1) defining a set of possible social strategies, (2) developing a method to generate content for each strategy conditioned on an image, and (3) choosing the appropriate strategy to maximize response conditioned on the user and their post. In this paper, we tackle these three challenges. First, we adopt a set of social strategies that social psychologists have demonstrated to be successful in human-human communication (Cialdini 2016; Robert 1984; Langer, Blank, and Chanowitz 1978; Taylor and Thomas 2008; Hoffman 1981). While our set is not exhaustive, it represents a diverse list of strategies — some that augment questions conditioned on the image and others conditioned on the user's language. While previous work has explored the use of ML models to generate image-conditioned natural language fragments, for generating captions and questions, ours is the first method that employs these techniques to generate strategies that increase worker participation.

To test the efficacy of our approach, we deploy our system on Instagram, a social media image-sharing platform. We collect datasets and develop machine learning-based models that use a convolutional neural network (CNN) to encode the image contents and a long short-term memory network (LSTM) to generate each social strategy across a large set of different kinds of images. We compare our ML strategies against baseline rule-based strategies using linguistic features extracted from the user's post (Li et al. 2010). We show a sample of augmented questions in Figure 2. We find that choosing appropriate strategies and augmenting requests leads to a significant absolute participation increase of 42.36% over no strategy when using ML strategies and a 14.78% increase when using rule-based strategies. We also find that no specific strategy is the universal best choice, implying that knowing when to use a strategy is important. While we specifically focus on VQA and Instagram, our approach generalizes to other crowdsourcing systems that support language-based interaction with contributors.

Related Work

Our work is motivated by research in crowdsourcing, peer production and social computing that increase contributors' levels of intrinsic motivation. We thread this work together with advances in natural language generation technologies to contribute generative algorithms that modulating the form of the requests to increase contribution rates.

Crowdsourcing strategies. The HCI community has investigated different ways to incentivise people to participate in data-labeling tasks (Hill 2013; Healy and Schussman 2003; Reich, Murnane, and Willett 2012). Designing for curiosity, for example, increases crowdsourcing participation (Law et al. 2016). Citizen science projects like GalaxyZoo mobilize volunteers by motivating them to work on a domain that aligns with their interests (Lintott et al. 2008). Unlike the tasks typically explored by such methods, image-labeling is not typically an intrinsically motivated task, and is instead completed by paid ghost work (Gray and Suri 2019). To improve image-labeling, the ESP Game harnessed game design to solve annotation tasks as by-products of entertainment activities (von Ahn and Dabbish 2004b). However, games result in limited kinds of labels, and need to be designed specifically to attain certain types of labels. Instead, we ask directed questions through conversations to label data and use social strategies to motive participation.

Interaction through conversations. The use of natural language as a medium for interaction has galvanized many systems (Huang, Chang, and Bigham 2018; Lasecki et al. 2013). Natural language has been proposed as a medium to gather new data from online participants (Bigham et al. 2010) or guide users through workflows (Fast et al. 2018). Conversational agents have also been deployed through products like Apple's Siri, Amazon's Echo, and Microsoft's Cortana. Substantial effort has been placed on teaching people how to talk to such assistants. Noticing this limitation, more robust crowd-powered conversational systems have been created by hiring professionals, as in the case of Facebook M (Hempel 2015), or crowd workers (Lasecki et al. 2013; Bohus and Rudnicky 2009). Unlike these approaches where people have a goal and invoke a passive conversational agent, we build active agents reach out to people with questions that increase humans participation.

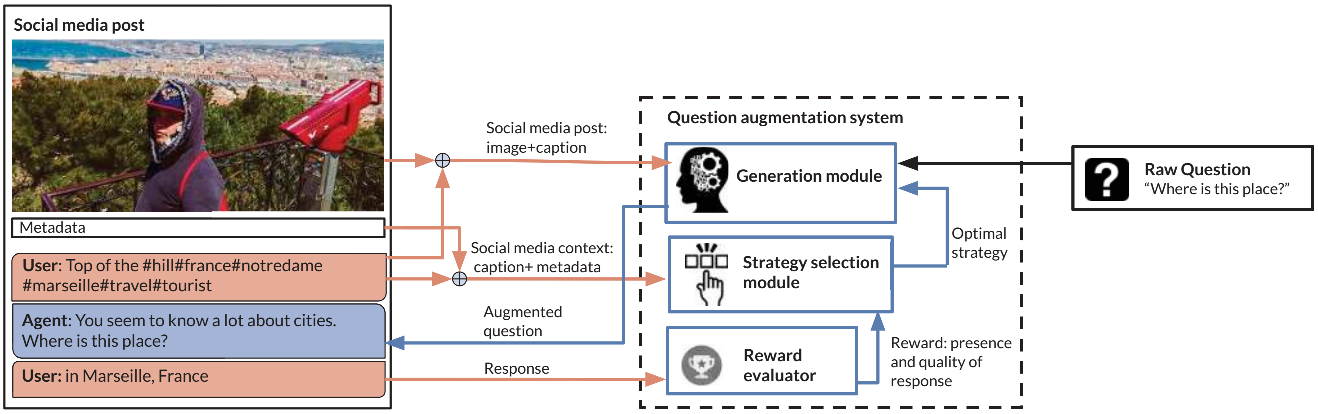
Social interaction with machines. To design an agent capable of eliciting a user's help, we need to understand how a user views the interaction. The Media Equation proposes that people adhere to similar social norms in their interactions with computers as they do in interactions with other people (Reeves and Nass 1996). It shoes that agents that seem more human-like, in terms of behaviour and gestures, provoke users to treat them similar to a person (Cassell and Thórisson 1999; Cerrato and Ekeklint 2002; Nass and Brave 2007). Consistent with these observations, prior work has also shown that people are more likely to resolve misunderstandings with more human-like agents (Corti and Gillespie 2016). This leads us to question whether a human-like conversational agent can encourage more online participation from online contributors. Prior work on interactions with machines investigates social norms that a machine can mimic in a binary capacity — either it respects the norm correctly or violates it with negligence (Sardar et al. 2012; Chidambaram, Chiang, and Mutlu 2012). Instead, we project social interaction on a spectrum — some social strategies are more successful than others in a given context — and learn a selection strategy that maximizes participation.

Structuring requests to enhance motivation. There have been many proposed social strategies to enhance the motivation to contribute in online communities (Kraut and Resnick 2011). For example, asking a specific question rather than making a statement or asking an open-ended question increases the likelihood of getting a response (Burke, Kraut, and Joyce 2014). Requests succeed significantly more often when contributors are addressed by name (Markey 2000). Emergencies receive more responses than requests without time constraints (Darley and Latané 1968). Prior work has shown that factors that increase the contributor's affinity for the requester increase the persuasive power of the message on online crowdfunding sites (Yang and Kraut 2017). It has also been observed that different behaviour elicits different kind of support from online support groups with self disclosure eliciting emotional support and questioning resulting in informational support (Wang, Kraut, and Levine 2015). The severity of the outcome of responding to a request can also influence motivation (Chaiken 1989). Our work incorporates some of these established social strategies and leverages language generation algorithms to build an agent that can deploy them across a wide variety of different requests.

Social strategies

The goal of our system is to draw on theories of how people ask other people for help and favors, then learn how to emulate those strategies. Drawing on prior work, we sampled a diverse set of nine social strategies. While the set of nine social strategies we explore are not an exhaustive set, we believe it represents a wide enough range of possible strategies to demonstrate the method and effects of teaching social strategies to machines. The social strategies we explore are:

1. `Content compliment`: Compliment the image or an object in the image before asking the question. This increases the liking between the agent and the contributor, making them more likely to reciprocate with the request (Robert 1984).
2. `Expertise compliment`: Compliment the knowledge of the contributor who posted the image. This commits the contributor as an "expert", resulting in a thoughtful response (Robert 1984).
3. `Interest matching`: Show interest in the topic of the contributor's post. This creates a sense of unity between the agent and contributor (Cialdini 2016).
4. `Valence matching`: Match the valence of the contributor based on their image's caption. People evolved to act kindly to others who exhibit behaviors from a similar culture (Taylor and Thomas 2008).
5. `Answer attempt`: Guess an answer and ask for a validation. Recognizing whether a shown answer is correct or not is cognitively an easier task for the listener than recalling the correct answer (Gillund and Shiffrin 1984).
6. `Time scarcity`: Specify an arbitrary deadline for the response. People are more likely to act if the opportunity is deemed to expire, even if they neither need nor want the opportunity (Robert 1984).
7. `Help request`: Explicitly request the contributor's help. People are naturally inclined to help others when they are asked and able to do so (Hoffman 1981).
8. `Logical justification`: Give a logical reason for asking the question to persuade the contributor at a cognitive level (Langer, Blank, and Chanowitz 1978).
9. `Random justification`: Give a random reason for asking the question. People are more likely to help if a justification is provided, even if it does not actually entail the request (Langer, Blank, and Chanowitz 1978).


Social media post
Metadata
Social media post: image+caption
Social media context: caption+ metadata
Augmented question
Response
Question augmentation system
Generation module
Strategy selection module
Reward evaluator
Optimal strategy
Reward: presence and quality of response
Raw Question
"Where is this place?"

User: Top of the #hill#france#notredame #marseille#travel#tourist

Agent: You seem to know a lot about cities. Where is this place?

User: in Marseille, France

Figure 3: Given a social media post and a question we want to ask, we augment the question with a social strategy. Our system contains two components. First, a selection component featurizes the post and user and chooses a social strategy. Second, a generation component creates a natural language augmentation for the question given the image and the chosen strategy. The contributor's response or silence is used to generate a feedback reward for the selection module.

System Design

In this section, we describe our approach for augmenting requests with social strategies (see Figure 3). Our approach is divided into two components: generation and selection. Given a social media post, we featurize the post metadata, question, and caption, then send them to the selection component. The selection component chooses an effective strategy to use for the given post. This strategy, along with a generated question to ask (Krishna, Bernstein, and Fei-Fei 2019), and the social media post are sent to the generation component, which augments the question by generating a natural language phrase for the chosen social strategy. The augmented request is then shared with the contributor. The selection module gathers feedback, positive if the contributor responds in an informative manner. Uninformative responses or no response are counted as a negative feedback.

Selection: Choosing a social strategy

We model our selection component as a contextual bandit. Contextual bandits are a common reinforcement learning technique for efficiently exploring different options and exploiting the best choices over time, generalizing from previous trials to uncommonly observed situations (Li et al. 2010). The component receives a feature vector and outputs its choice of an arm (option) that it expects to result in the highest expected reward.

Each social media post is represented as a feature vector that encodes information about the user, the post, and the caption. User features include- number of posts the user has posted, number of followers, number of accounts the user is following, number of other users tagged in their posts, filters and AR effects the user uses frequently on the platform, user's engagement with videos, whether the user is a verified business or an influencer, user's privacy settings, the engagement with Instagram features such as highlight reels and resharing, and sentiment analysis on their biography. Post features include the number of users who like the post and the number of users who commented on the post. User and post features are drawn from Instagram's API and featurized as bag of words or one-hot vectors. Lastly, caption features are extracted from sentiment using Vader (Hutto and Gilbert 2014), and the hashtags extracted using regular expressions.

We train a contextual bandit model to choose a social strategy given the extracted features, conditioned on the success of each social strategy used on similar social media posts in the past. The arms that the contextual bandit considers represent each of the nine social strategies that the system can use. If a chosen social strategy receives a response, we parse and check if the response contains an answer (Devlin et al. 2018). If so, the model receives a positive reward for choosing the social strategy. If a chosen social strategy does not receive a response, or if the response does not contain an answer, the model receives a negative reward.

Our implementation of contextual bandit uses the adaptive greedy algorithm for balancing the trade-off between exploration and exploitation. During training, the algorithm chooses an option that the model associates with a high uncertainty of reward. If there is no option with a high uncertainty, the algorithm chooses a random option to explore. The threshold for uncertainty decreases as the model is exposed to more data. During inference, the model predicts the social strategy with highest expected reward (Zhang 2004).

Generation: Augmenting questions

The generation component receives the social media post (an image and a caption) and a raw question automatically generated by existing visual question generation algorithms (e.g., "Where is this place?"). It produces a natural language contextualization of the question using one of the nine social strategies chosen by the selection component.

We build nine independent natural language generation systems that each receive a social media post as input and produce a comment using the corresponding social strategy as output. Four of the social strategies require knowl-

		Machine learning based				Rule based				
Image	Raw question	Content compliment	Expert compliment	Interest matching	Logical justification	Valence matching	Answer attempt	Help request	Time scarcity	Random justification
	Where is this place?	What a cute elephant! Where is this place?	You must really love animals. Where is this place?	I like the design. Where is this place?	Where is this place? It's because I would love to visit a vineyard!	Thank you for posting this. Where is this place?	Where is this place? in Bangladesh, right? I've been there myself	I need your help to find out the name of this place! Could you help me?	Where is this place? I need to know for a homework due tomorrow midnight	Where is this place? Because I am a doctor.
	Where is this place?	The detail in that building is gorgeous! Where is this place?	You seem to know a lot about architecture. Where is this place?	I like stairs too. Where is this place?	Where is this place? It's because I want to see the view	Thank you for posting this. Where is this place?	Where is this place? in Middle East right? I've been there myself	Please help me find the name of this place! Could you tell me?	Where is this place? I need to know for a homework due tomorrow midnight	Where is this place? Because I want to make some myself.
	Where is this place?	Such a beautiful sunset. Where is this place?	You seem to know a lot about oceans. Where is this place?	I really love the beach too. Where is this place?	Where is this place? It's because I love the mountains	Thank you for posting this. Where is this place?	Where is this place? in Russia right? I've been there myself	Please help me find the name of this place! Could you tell me?	Where is this place? Could you please tell me before end of day tomorrow?	Where is this place? Because I am in France right now.

Figure 4: Example augmentations generated by each of our social strategies.

edge about the content of images, and are implemented using machine learning-based models. These strategies cannot be templatized, as there is substantial variation in the kinds of images found online and the approaches much be personalized to the content of the image. We use the other five social strategies as baseline strategies that only require knowledge about the speaking style of the social media user, and are implemented as rule-based expert systems in conjunction with natural language processing techniques. We discuss these two types of models below.

Machine learning-based social strategies. To generate sentences specific to the image of each post, we train one machine learning model for each of the four social strategies that require knowledge about the image: expert compliment, content compliment, interest matching, and logical justification.

We build a dataset of 10k social media posts alongside examples of questions that use each of the four social social strategies, with the help of crowd workers on Amazon Mechanical Turk. This process results in a dataset of 40k questions, each with social strategy augmentations. The posts are randomly selected by polling Instagram for images with one of the top 100 most popular hashtags on Instagram and filter for those that refer to visual content, such as #animal, #travel, #shopping, #food, etc. Crowdworkers are designated to one of the four strategy categories and trained using examples and a qualifying task, which we manually evaluate. Each task contains 10 social media posts (images and captions) and the generated questions. Workers are asked to submit a natural language sentence that can be pre- and post-pended to the question while adhering to the social strategy they are trained to emulate. The workers are paid a compensation that is equivalent to $12 an hour for their work.[2]

We adopt a traditional image-to-sequence machine learning model to generate the sentence for each strategy. Each model encodes the social media image using a convolutional neural network (CNN) (Krizhevsky, Sutskever, and Hinton 2012) and generates a social strategy sentence, conditioned on image features, using a long short term memory (LSTM) network (Hochreiter and Schmidhuber 1997). We train each model using the dataset of 10k posts dedicated to its assigned strategy using stochastic gradient descent with a learning rate of $1e - 3$ for 15 epochs.

Baseline rule-based social strategies. To generate social strategy sentences that are relevant to the caption of each social media post, we create a rule-based expert system for each of the five social strategies: valence matching, answer attempt, help request, time scarcity, and random justification. While these algorithms use statistical machine learning approaches for natural language processing, we call them rule-based systems to clarify that the generation, itself, is a deterministic process unlike the machine learning strategies.

Valence matching detects the emotional valence of the caption through punctuation parsing and sentiment analysis using an the Vader algorithm (Hutto and Gilbert 2014). The algorithm generates a sentence with emotional valence that is approximately equal to valence of the caption by matching type and number of punctuations and adding appropriate exclamations like "Wow!" or "Aw". Answer attempt guesses a probable answer for the input post based on the raw question and hashtags of the post. To guess a probable answer, we manually curate a set of likely answers for problem domains and words from caption and randomly choose one from the set. For example, when asking where we could buy the same item on a post that references the word "jean" in the "#shopping" domain, the set of probable answers are a list of brands that sell jeans to consumers. Deployments of this strategy does not have to rely on a curated list and can instead use existing answering models (Antol et al. 2015). Help request augments the agent's question with variations of words and sentence structures that humans use to request help from one another. Time scarcity augments the agent's question with vari-

[2] The dataset of social media posts and social strategies for training the reinforcement learning model, as well as the trained contextual bandit model, is publicly available at http://cs.stanford.edu/people/ranjaykrishna/socialstrategies.

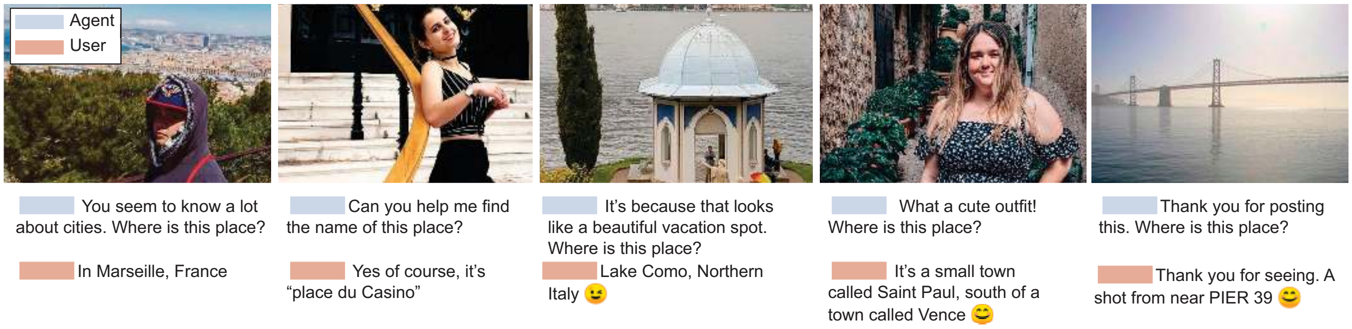

Agent
User
You seem to know a lot about cities. Where is this place?
In Marseille, France
Can you help me find the name of this place?
Yes of course, it's "place du Casino"
It's because that looks like a beautiful vacation spot. Where is this place?
Lake Como, Northern Italy 😉
What a cute outfit! Where is this place?
It's a small town called Saint Paul, south of a town called Vence 😊
Thank you for posting this. Where is this place?
Thank you for seeing. A shot from near PIER 39 😊

Figure 5: Example responses to `expertise compliment`, `help request`, `logical justification`, `content compliment` and `valence matching` in the travel domain.

ations of a sentence that requests the answer to be provided within 24 hours. `Random justification` augments the agent's question with a justification that is chosen irrespective of the social media post. Specifically, we store a list of justification sentences generated from the logical justification system for other posts, and retrieve one at random. Figure 4 visualizes example augmentations generated by each of our nine strategies, conditioned on the post.

Experiments

We evaluate the utility of augmenting questions with social strategies through a real-world deploying on Instagram. Our aim is to increase online crowdsourcing participation from Instagram users when we ask them questions about their image contents. We begin our experiments by first describing the experimental setup, the metrics used, the baselines, and strategies surveyed. Next, we study how generated social strategies impact participation. Finally, we study the importance of selecting the correct social strategy.

Experimental setup

We poll images from Instagram, featurize the post, select a social strategy, and generate the question augmentation. We post the augmented question and wait for a response.

Images and raw questions. We source images from Instagram across 4 domains: travel, animals, shopping and food. Images from each domain are polled by searching for posts with hashtags: #travel, #animals, #shopping, and #food. Images in these four domains consitute an upper bound of 7.06% of all images posted with one of the top 100 popular hashtags that represent visual content. Since we are studying the impact of using different social strategies by directly interacting with real users on Instagram, we can not post multiple questions, each augmented with a different strategy, to the same image post. Ideally, in online crowdsourcing deployments, the raw questions generated would be conditioned on the post or image. In our case, however, we use only one question per domain so that all users are exposed to the same basic question. For each domain, we hold the raw question constant. For example, "Where is this place?" for travel, "What animal is that?" for animals, "Where can I get that?" for shopping, and "What is this food?" for food.

Table 1: Response rates achieved by different strategies on posts in the source and target domains. The bottom of the table shows a comparison between average performance of ML based strategies, average performance of rule-based strategies and baseline un-augmented questions

	Source domain (%)	Target domain (%)
Expertise compliment	**72.90**	29.55
Content compliment	59.11	68.96
Interest matching	45.31	**85.38**
Logical justification	55.17	19.7
Answer attempt	41.37	42.69
Help request	31.52	32.84
Valence matching	37.43	36.12
Time scarcity	24.63	26.27
Random justification	17.73	32.84
ML based strategies	**58.12**	**50.89**
Rule based strategies	30.54	34.15
No strategy	15.76	13.13

Metrics. To measure the improvements in crowdsourcing participation, we report the percentage of informative responses. After a question is posted on Instagram, we wait 24 hours to check if a response was received. If the question results in no response or if the response doesn't answer the question or the user appears confused (e.g. "huh?" or "I don't understand"), the interaction is not counted as an informative response. To verify if a response is informative, we send all responses to Amazon Mechanical Turk (AMT) workers to report whether the question was actually answered with gold standard responses to guarantee quality.

Strategies surveyed. We use all nine strategies described earlier and add a baseline and an oracle strategy. The baseline case posts the raw question with no augmentation. The oracle method asks AMT workers to modify the question to maximize the chances of receiving the answer. They don't have to follow any of our outlined social strategies.

Dataset of online interactions. To study the impact of using social strategies, we collect a dataset of 10k posts for each of the 4 ML social strategies, resulting in a dataset

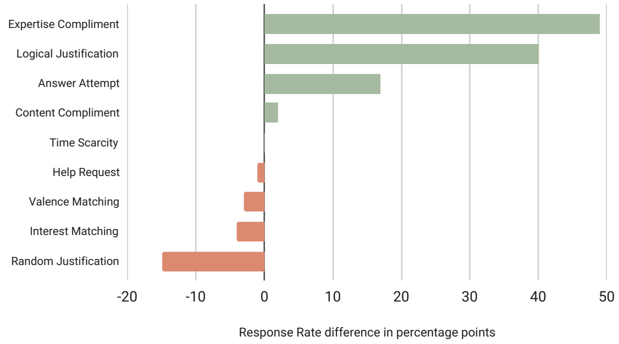


Figure 6: Difference between response rate of the agent and humans for each social strategy. Green indicates the agent is better than people and red indicates the opposite.

of 40k questions with augmentations. The 5 rule strategies don't require any training data. Once trained, we post 100 questions per strategy to Instagram, resulting in 1100 total posts. To further study the scalability and transfer of strategies learned in one domain and applied to another, we train augmentation models using data from a "source" domain and test its effect on posts from "target" domains. For example, we train models using data collected from the #travel source domain and test on the rest as target domains.

To train the selection model, we gather 10k posts from Instagram and generate augmentations with each of the social strategies. Each post, with all the augmentated questions, is sent to AMT workers, who are asked to pick the strategies that would be appropriate to use. We choose to train the selection model using AMT instead of Instagram as it allows us to quickly collect large amounts of training data and negate the impact of other confounds. Each AMT task included 10 social media posts. One out of the ten posts contained an attention checker in the question to verify that the workers were actually reading the questions. Workers were compensated at a rate of $\$12$ per hour.

Augmenting questions with social strategies

Our goal in the first set of experiments is to study the effect of using social strategies to augment questions.

Informative responses. Before we inspect the effects of social strategies, we first report the quality of responses from Instagram users. We manually annotate all our responses and find that 93.01% of questions are both relevant as well as answerable. Out of the relevant questions, 95.52% of responses were informative, i.e. the responses contained the correct answer to the question. Figure 5 visualizes a set of example responses for different posts with different social strategies in the travel domain. While all social strategies outperformed the baseline in receiving responses, the quality of the responses differed across strategies.

Effect of social strategies. Table 1 reports the informative response rate across all the social strategies. We find that, compared to the baseline case, where no strategy is used, rule-based strategies improve participation by 14.78 percent points. An unpaired t-test confirms that participation increases by designing appropriate rule-based social strategies ($t(900) = 3.05$, $p < 0.01$). When social strategy data is collected and used to train ML strategies, performance increases by 42.36 percent points and 27.58 percent points when compared against un-augmented ($t(900) = 8.17$, $p < 0.001$) and rule-based strategies ($t(900) = 8.96$, $p < 0.001$) and confirmed by unpaired t-tests. Overall, we find that `expertise compliment` and `logical justification` performed strongly in shopping domain, but weakly in animals and food domains.

To test the scalability of our strategies across image domains, we train models on a source domain and deploy them on a target domain. We find that `expertise compliment` drops in performance while `interest matching` improves. The drop implies that machine learning models that heavily depend on example data points used in training process are not robust in new domains. Therefore, while machine learning strategies are the most effective, they require strategy data collected for the domain in which they are deployed. The drop in performance, however, still results in improvements in response rate, demonstrating that machine learning strategies scale across domains but their impact reduces as the distribution of image content increases from the source domain. The increase in performance of `interest matching` indicates that different domains might have different dominating social strategies, i.e. no single dominant strategy exists across all domains and that a selection component is necessary.

Agent versus human augmentations. We compare the augmentations generated by our agent against those created by crowdworkers. We report the difference in response rate between the agent and the human augmentations across the different strategies in Figure 6. A two-way ANOVA finds that the strategy used has a significant effect on the response rate ($F(8, 900) = 12.99$, $p < 0.001$) but the poster has no significant effect on the response rate ($F(1, 900) = 1.82$, $p = 0.17$). The ANOVA also found a significant interaction effect between the strategy and the poster on response rate ($F(1, 900) = 2.09$, $p = 0.03$). A posthoc Tukey test indicates that the agent using the machine learning strategies is significantly increases response rate than the agent using rule-based ($p < 0.05$) or humans using rule-based strategies ($p < 0.05$). This demonstrates that a machine learning model that has witnessed examples of social strategies can outperform rule-based systems. However, there is no significant difference between the agent using machine learning strategies versus humans using the same social strategies.

Learning to select a social strategy

In our previous experiment we established that different domains have different strategies that perform best. Now, we evaluate how well our selection component performs at selecting the most effective strategy. Specifically, we test how well our selection model performs (1) against a random strategy, (2) against the most effective strategy (`expertise compliment`) from the previous experi-

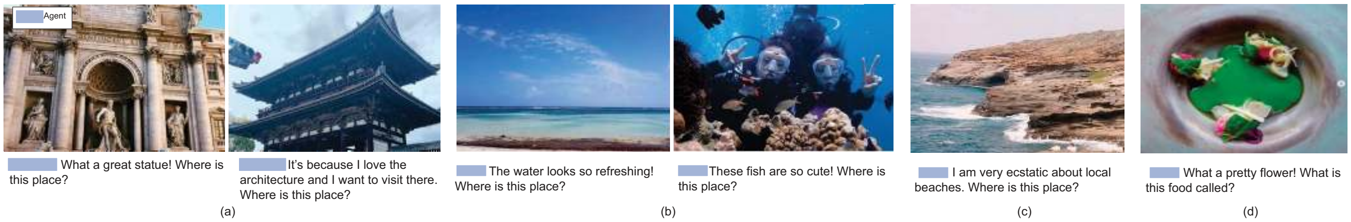

Agent
What a great statue! Where is this place?
It's because I love the architecture and I want to visit there. Where is this place?
(a)
The water looks so refreshing! Where is this place?
These fish are so cute! Where is this place?
(b)
I am very ecstatic about local beaches. Where is this place?
(c)
What a pretty flower! What is this food called?
(d)

Figure 7: Example strategy selection and augmentations in the travel domain. (a) Our system learns to focus on different aspects of the image. (b) The system is able to discern between very similar images and understand that the same objects can have different connotations. (c, d) Example failure case when objects were misclassified.

ment, and (3) against the oracle strategy generated by crowdworkers. Recall that the oracle strategy does not constrain workers to use any particular strategy.

Since this test needs be able to test multiple strategies on the same post, we perform our evaluation on AMT. Workers are shown two strategies for a given post and asked to choose which strategy is most likely to receive a response. We perform pairwise comparisons between our selection model against a random strategy across 11k posts, against `expertise compliment` across 549 posts and against open-ended human questions across 689 posts.

Effect of selection. A binomial test indicates that our selection method was chosen 54.12% more often than a random strategy $B(N = 11,844, p < 0.001)$. It was chosen 58.28% more often than `expertise compliment` $B(N = 549, p < 0.001)$. And finally, it was chosen 75.61% more often than the oracle human generated questions $B(N = 689, p < 0.001)$. We conclude that our selection model outperforms all existing baselines.

Qualitative analysis. Figure 7(a) shows that the agent can choose to focus on different visual aspects even when the subject of the image is roughly the same. In one, the agent compliments the statue, which is the most salient feature of the old European building shown in the image. In the other, it shows appreciation for the overall architecture of the old Asian building, which does not have a single defining feature like a statue. Figure 7(b) shows two images that are both contain water and has similar color composition. In one, the agent compliments the water seen on the beach as refreshing and in the other, the fish seen underwater as cute. Referring to a fish in a beach photo would have been incorrect as would have been describing water as refreshing in an underwater photo. Though social strategies are useful, they can also lead to new errors. Figure 7(c, d) showcases an example questions where the agent fails to recognize mountains and food and generates phrases referring to beaches and flowers.

Discussion

Intended use. This work demonstrates that it is possible to train an AI agent to use social strategies that are found in human-to-human interaction contexts to increase the likelihood of a human crowdsourcing respondent. Such responses suggest a future in which supervised ML models can be trained on authentic online data that are provided by willing helpers than from paid workers. We expect that such strategies can lead to adaptive ML systems that can learn during their deployment, by asking their users whenever they are uncertain about their environment. Unlike existing paid crowdsourcing techniques that grow linearly in cost as the number of annotations increases, our method is a fixed cost solution where social strategies need to be collected for a specific domain and then deployed to encourage volunteers.

Negative usage. We pause to note the potential negative implications of computing research, and how they can be addressed. The psychology techniques that our work relies on have been used in negotiations and marketing campaigns for decades. Automating such techniques can also lead to influencing emotions or behavior at a magnitude greater than single human-human interaction (Kramer, Guillory, and Hancock 2014; Ferrara et al. 2016). When using natural language techniques, we advocate that agents continue to self-identify as bots. Online communities should establish a standard acceptable use of such techniques and inform contributors about the intentions behind an agent's request.

Limitations and future work. Our social strategies are not an exhaustive list. Future research could directly learn to emulate strategies by observing human-human interactions. Currently, our requests involve exactly one dialogue turn, and we do not yet explore multi-turn conversations. This can be important: for example, the answer attempt strategy may be more effective at getting an answer now, but might also decrease the contributor's likeliness to continue cooperating in the future. Future work can explore how to guide conversations to enable more complex labeling schemes.

Conclusion

Our work: (1) identifies social strategies that can be repurposed to improve crowdsourcing requests for visual question answering, (2) trains and deploys machine learning and rule-based models that deploy these strategies to increase crowdsourcing participation, and (3) demonstrates that these models significantly improve participation on Instagram, that no single strategy is optimal, and that a selection model can chooses the appropriate strategy.

Acknowledgments. This work was partially funded by the Brown Institute of Media Innovation and by Toyota Research Institute ("TRI"). This article solely reflects the opinions and conclusions of its authors and not any Toyota entity.

References

Ambati, V.; Vogel, S.; and Carbonell, J. 2011. Towards task recommendation in micro-task markets.

Antol, S.; Agrawal, A.; Lu, J.; Mitchell, M.; Batra, D.; Lawrence Zitnick, C.; and Parikh, D. 2015. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, 2425–2433.

Bigham, J. P.; Jayant, C.; Ji, H.; Little, G.; Miller, A.; Miller, R. C.; Miller, R.; Tatarowicz, A.; White, B.; White, S.; et al. 2010. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 333–342. ACM.

Bohus, D., and Rudnicky, A. I. 2009. The ravenclaw dialog management framework: Architecture and systems. *Computer Speech & Language* 23(3):332–361.

Brady, E. L.; Zhong, Y.; Morris, M. R.; and Bigham, J. P. 2013. Investigating the appropriateness of social network question asking as a resource for blind users. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 1225–1236. ACM.

Brady, E.; Morris, M. R.; and Bigham, J. P. 2015. Gauging receptiveness to social microvolunteering. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 1055–1064. New York, NY, USA: ACM.

Burke, M.; Kraut, R.; and Joyce, E. 2014. Membership claims and requests: Some newcomer socialization strategies in online communities. *Small Group Research*.

Cassell, J., and Thórisson, K. R. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13:519–538.

Cerrato, L., and Ekeklint, S. 2002. Different ways of ending human-machine dialogues.

Chaiken, S. 1989. Heuristic and systematic information processing within and beyond the persuasion context. *Unintended thought* 212–252.

Chidambaram, V.; Chiang, Y.-H.; and Mutlu, B. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 293–300. ACM.

Cialdini, R. 2016. *Pre-Suasion: A revolutionary way to influence and persuade*. Simon and Schuster.

Corti, K., and Gillespie, A. 2016. Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior* 58:431 – 442.

Darley, J. M., and Latané, B. 1968. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of personality and social psychology* 8(4p1):377.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Difallah, D. E.; Demartini, G.; and Cudré-Mauroux, P. 2013. Pick-a-crowd: Tell me what you like, and i'll tell you what to do. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, 367–374. New York, NY, USA: ACM.

Fast, E.; Chen, B.; Mendelsohn, J.; Bassen, J.; and Bernstein, M. S. 2018. Iris: A conversational agent for complex tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 473. ACM.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. *Communications of the ACM* 59(7):96–104.

Geiger, D., and Schader, M. 2014. Personalized task recommendation in crowdsourcing information systems – current state of the art. *Decision Support Systems* 65:3 – 16. Crowdsourcing and Social Networks Analysis.

Gillund, G., and Shiffrin, R. M. 1984. A retrieval model for both recognition and recall. *Psychological review* 91(1):1.

Gray, M., and Suri, S. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Eamon Dolan.

Healy, K., and Schussman, A. 2003. The ecology of open-source software development. Technical report, Technical report, University of Arizona, USA.

Hempel, J. 2015. Facebook launches m, its bold answer to siri and cortana. *Wired. Retrieved January* 1:2017.

Hill, B. M. 2013. Almost wikipedia: Eight early encyclopedia projects and the mechanisms of collective action. *Massachusetts Institute of Technology* 1–38.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Hoffman, M. L. 1981. Is altruism part of human nature? *Journal of Personality and social Psychology* 40(1):121.

Huang, T.-H. K.; Chang, J.; and Bigham, J. 2018. Evorus: A crowd-powered conversational assistant built to automate itself over time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 295. ACM.

Hutto, C. J., and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Kramer, A. D.; Guillory, J. E.; and Hancock, J. T. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24):8788–8790.

Kraut, R. E., and Resnick, P. 2011. Encouraging contribution to online communities. *Building successful online communities: Evidence-based social design* 21–76.

Krishna, R.; Bernstein, M.; and Fei-Fei, L. 2019. Information maximizing visual question generation. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In Pereira, F.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. 1097–1105.

Langer, E. J.; Blank, A.; and Chanowitz, B. 1978. The mindlessness of ostensibly thoughtful action: The role of" placebic" information in interpersonal interaction. *Journal of personality and social psychology* 36(6):635.

Lasecki, W. S.; Wesley, R.; Nichols, J.; Kulkarni, A.; Allen, J. F.; and Bigham, J. P. 2013. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, 151–162. ACM.

Law, E.; Yin, M.; Goh, J.; Chen, K.; Terry, M. A.; and Gajos, K. Z. 2016. Curiosity killed the cat, but makes crowdwork better. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4098–4110. ACM.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, 661–670. ACM.

Lin, C.; Kamar, E.; and Horvitz, E. 2014. Signals in the silence: Models of implicit feedback in a recommendation system for crowdsourcing.

Lintott, C. J.; Schawinski, K.; Slosar, A.; Land, K.; Bamford, S.; Thomas, D.; Raddick, M. J.; Nichol, R. C.; Szalay, A.; Andreescu, D.; et al. 2008. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society* 389(3):1179–1189.

Lu, C.; Krishna, R.; Bernstein, M.; and Fei-Fei, L. 2016. Visual relationship detection with language priors. In *European Conference on Computer Vision*, 852–869. Springer.

Markey, P. M. 2000. Bystander intervention in computer-mediated communication. *Computers in Human Behavior* 16(2):183–188.

Nass, C., and Brave, S. 2007. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.

Niebles, J. C.; Wang, H.; and Fei-Fei, L. 2008. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision* 79(3):299–318.

Reeves, B., and Nass, C. I. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press.

Reich, J.; Murnane, R.; and Willett, J. 2012. The state of wiki usage in us k–12 schools: Leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher* 41(1):7–15.

Robert, C. 1984. Influence: The psychology of persuasion. *William Morrow and Company, Nowy Jork.*

Sardar, A.; Joosse, M.; Weiss, A.; and Evers, V. 2012. Don't stand so close to me: users' attitudinal and behavioral responses to personal space invasion by robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 229–230. ACM.

Taylor, P. J., and Thomas, S. 2008. Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research* 1(3):263–281.

von Ahn, L., and Dabbish, L. 2004a. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326. ACM.

von Ahn, L., and Dabbish, L. 2004b. Labeling images with a computer game. 319–326.

Wang, Y.-C.; Kraut, R. E.; and Levine, J. M. 2015. Eliciting and receiving online support: Using computer-aided content analysis to examine the dynamics of online social support. *J Med Internet Res* 17(4):e99.

Warncke-Wang, M.; Ranjan, V.; Terveen, L.; and Hecht, B. 2015. Misalignment between supply and demand of quality content in peer production communities. In *Ninth International AAAI Conference on Web and Social Media*.

Willis, C. G.; Law, E.; Williams, A. C.; Franzone, B. F.; Bernardos, R.; Bruno, L.; Hopkins, C.; Schorn, C.; Weber, E.; Park, D. S.; et al. 2017. Crowdcurio: an online crowdsourcing platform to facilitate climate change studies using herbarium specimens. *New Phytologist* 215(1):479–488.

Yang, D., and Kraut, R. E. 2017. Persuading teammates to give: Systematic versus heuristic cues for soliciting loans. *Proc. ACM Hum.-Comput. Interact.* 1(CSCW):114:1–114:21.

Zhang, T. 2004. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the twenty-first international conference on Machine learning*, 116. ACM.