Framing Effect: Choice of Slogans Used to Advertise Online Experiments Can Boost Recruitment and Lead to Sample Biases

TAL AUGUST, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA
NIGINI OLIVEIRA, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA
CHENHAO TAN, Department of Computer Science, University of Colorado Boulder, USA
NOAH A. SMITH, Paul G. Allen School of Computer Science & Engineering, University of Washington & Allen Institute for Artificial Intelligence, USA
KATHARINA REINECKE, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA

Online experimentation with volunteers relies on participants' non-financial motivations to complete a study, such as to altruistically support science or to compare oneself to others. Researchers rely on these motivations to attract study participants and often use incentives, like performance comparisons, to encourage participation. Often, these study incentives are advertised using a slogan (e.g., *"What is your thinking style?"*). Research on framing effects suggests that advertisement slogans attract people with varying demographics and motivations. Could the slogan advertisements for studies risk attracting only specific users? To investigate the existence of potential sample biases, we measured how different slogan frames affected which participants self-selected into studies. We found that slogan frames impact recruitment significantly; changing the slogan frame from a 'supporting science' frame to a 'comparing oneself to others' frame lead to a 9% increase in recruitment for some studies. Additionally, slogans framed as learning more about oneself attract participants significantly more motivated by boredom compared to other slogan frames. We discuss design implications for using frames to improve recruitment and mitigate sources of sample bias in online research with volunteers.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**;

Additional Key Words and Phrases: framing; study recruitment; volunteer-based online experimentation

ACM Reference Format:
Tal August, Nigini Oliveira, Chenhao Tan, Noah A. Smith, and Katharina Reinecke. 2018. Framing Effect: Choice of Slogans Used to Advertise Online Experiments Can Boost Recruitment and Lead to Sample Biases . *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 22 (November 2018), 19 pages. https://doi.org/10.1145/3274291

Authors' addresses: Tal August, Paul G. Allen School of Computer Science & Engineering, University of Washington, 185 E Stevens Way NE, Seattle, WA, 98195-2350, USA, taugust@cs.washington.edu; Nigini Oliveira, Paul G. Allen School of Computer Science & Engineering, University of Washington, 185 E Stevens Way NE, Seattle, WA, 98195-2350, USA, nigini@cs.washington.edu; Chenhao Tan, Department of Computer Science, University of Colorado Boulder, 430 UCB, Boulder, CO, 80309, USA, chenhao@chenhaot.com; Noah A. Smith, Paul G. Allen School of Computer Science & Engineering, University of Washington, 185 E Stevens Way NE, Seattle, WA, 98195-2350, & Allen Institute for Artificial Intelligence, USA, nasmith@cs.washington.edu; Katharina Reinecke, Paul G. Allen School of Computer Science & Engineering, University of Washington, 185 E Stevens Way NE, Seattle, WA, 98195-2350, USA, reinecke@cs.washingont.edu.

2573-0142/2018/11-ART22 $15.00
https://doi.org/10.1145/3274291

1 INTRODUCTION

Researchers are increasingly conducting online studies with volunteers, such as on the platforms VolunteerScience [62], GamesWithWords [23], LabintheWild [47], and TestMybrain [19], as well as by simply posting study links on online social networks. Unlike compensated studies, which motivate participants through financial gain, studies with volunteers incentivize participation through intrinsic motivations, such as the desire to help science or to learn more about oneself. To leverage these motivations and attract participants, researchers often advertise studies using catchy slogans, such as *"Shine a light on vision research!"* Prior work in volunteer recruiting across a variety of domains shows that the use of slogan *framing*—where the way information is presented leads to differences in how it is perceived [13, 60]—can lead to surprisingly large differences in how many and which people are attracted [22, 33, 56, 60].

Similar effects of slogan framing could be expected in volunteer-based online studies. In fact, these studies often collect widely varied sample sizes [35], with markedly different makeups in terms of demographics and motivations to take studies (e.g., interest in supporting science versus interested in comparing to others) [29, 47]. These *sample biases*, i.e., recruiting a participant sample not representative of the entire population, can threaten study validity and generalizability [6].

The known differences in sample size and makeup of people that studies attract, coupled with framing literature showing that slogans can attract differently depending on the individual, suggest that how a study is advertised (i.e., what slogan it uses) is important in collecting adequate and representative samples of participants. Are some slogan frames best at attracting participants? Do certain types of participants prefer specific slogan frames more than others? If so, what sorts of slogans will attract whom?

The goal of this paper is to (1) identify how using slogan frames affects recruitment in online studies with volunteers and (2) determine whether different slogan frames recruit participants differently based on motivation and demographics. To find out how slogan preferences, participant motivations and demographics relate to one another, we collected preferences for different slogan frames and motivations for volunteering from 59 participants on LabintheWild, a volunteer-based online experimentation site [47]. We then explored how these preferences could affect sample size and composition in online studies by deploying these slogans on five LabintheWild studies, measuring how each affected subject recruitment with 2,670 participants.

Our analyses contribute two main findings:

(1) *Slogans offering the incentive of participants learning about themselves or comparing themselves to others are preferred by significantly more participants than other frames.* This preference leads to marked differences in recruitment: using a slogan that offers participants the benefit of learning something new about themselves, such as personalized feedback or a social comparison, can lead to almost a 9% increase in recruitment over a slogan appealing to participants' altruism to support science.
(2) *Slogans focusing on participants learning something new about themselves recruited participants who were significantly more interested to take studies out of boredom than other frames.* This influence of participant motivation on slogan preference can lead to higher drop out rates and less attentive participants [29], impacting sample composition and threatening results.
(3) *There were no significant differences between the age or gender of participants recruited by each frame.* While past work has shown differences in the the age and gender of people attracted to different slogan advertisements, we did not find a similar result.

Our results suggest that researchers can use frames to boost recruitment in online studies, but that it is important to recognize the possible sample biases a researcher introduces by using a specific frame. For example, advertising studies as a way for participants to learn something new about themselves attracts participants more motivated by boredom than other frames. Participants motivated by boredom have previously been found to be more inclined to drop out or not pay attention to study instructions [29]. Based on our results, we recommend cycling through multiple different slogan frames to advertise an online study in order to both collect adequate numbers of participants while also ensuring participants with diverse motivations.

2 RELATED WORK

Related to our research questions mentioned in the introduction is prior work on: (1) motivations to participate in online experiments and community-driven projects, (2) framing effects and target marketing, and (3) framing for recruitment of volunteers.

2.1 Motivation for Taking Online Studies and Contributing to Community-Driven Online Projects

People take part in online studies around the world, yet their motivations for doing so differ. For example, people on Amazon Mechanical Turk (MTurk)—a large, general-purpose crowd-sourcing site that researchers use to collect participant samples by offering small financial compensation [38, 50]—complete tasks (whether an experiment or general crowd-sourcing task) for a variety of motivations other than the financial, such as to have fun, kill time, or seek a sense of purpose [2, 30]. However, studies on MTurk and similar sites rely on financial compensation: studies posted on MTurk that offer no compensation are significantly less effective than compensated ones at attracting and retaining participants [21, 38, 49]. Hsieh and Kocielnik [26] showed that Turkers' values influence the type of financial reward they prefer (e.g., a lottery or fixed reward) which led to differences in who self-selected into studies. These differences in self-selection affected task outcome: a lottery reward attracted participants who provided more ideas in a brainstorming task than participants recruited through a fixed reward [26]. This paper explores enhancing recruitment through leveraging intrinsic motivations, which are often more varied and lack the financial component [29].

Volunteer-based online experimentation has become increasingly popular in a number of disciplines, including linguistics (e.g., GamesWithWords [23]), psychology (e.g., Project Implicit [45], TestMyBrain [19]), and social science and HCI research (e.g., LabintheWild.org [47]). These sites have been shown to attract large numbers of participants. LabintheWild [47], for example, engages an average of 1,000 participants a day, with over 4 million participants across more than 230 countries since its 2012 inception. As of 2017, TestmyBrain has hosted 1.7 million participants for 140+ research studies representing over 240 countries [59].

Unlike compensated studies, participant motivations for volunteer-based online studies are more diverse [29]. An analysis of tweets about LabintheWild, for example, revealed four distinct motivators for participants to share their results on social media: (1) comparison to others, (2) curiosity about themselves, (3) desire to know if their own characteristics could be predicted, and (4) improvement of particular skills [47]. Jun et al. [29] later surveyed 7,674 participants for three LabintheWild studies on their motivations to participate and identified five unique categories: (1) interest in supporting science, (2) interest in learning about oneself, (3) interest in comparing oneself to others, (4) boredom, and (5) desire to have fun [29]. They found that the degree of these motivations differed significantly by study, and that the type of motivation affected participant rating consistency and accuracy in studies, as well as the likelihood a participant would drop out of a study [29]. The findings from Jun et al. [29] suggest that the samples of different studies are skewed in terms of motivations (i.e., some studies collect more participants interested in comparing themselves to others relative to other studies); however, it is unclear whether this is due to how the study is advertised. We extend this related work by analyzing whether the framing of slogans is the reason for skewed study samples.

Another body of work relevant to intrinsic motivations to volunteer is the theory of psychological empowerment (PE). PE refers to perceived personal control and often is exemplified by an understanding and confidence in effecting change within one's environment [64]. Research has shown that encouraging factors of PE, such as self-efficacy, can motivate greater citizen participation in online community projects [20]. Additionally, volunteering in community efforts has been shown to promote aspects of PE, such as self-efficacy and a sense of community [43].

Participants who volunteer in community-driven projects, such as Wikipedia [63] or GalaxyZoo [46], also display a similarly diverse set of motivations. In GalaxyZoo, a citizen science site where users classify galaxies from astronomy images, Raddick et al. [46] identified 12 motivations for volunteering, the most popular being an interest in contributing to science, astronomy and discovering new galaxies [46]. Participants on Foldit, a science-game site that addresses protein folding, exhibited the strongest interests in contributing to science and interacting with others in a community [14]. For the citizen science project stardust@home, the collective motivations (i.e., working towards a project's goals) and intrinsic motivations (i.e., enjoyment from participating in the activity itself) were strongest [41]. Other studies have also explored the motivation to take part in pro-social behaviour. Liu et al. [36] categorized motivations for Kiva users, a peer-to-peer micro-lending site, into 10 categories, including religious duty, empathy, altruism, and reciprocity.

2.2 Framing Effects

Recruiting participants for online studies can be inspired by a large body of work on marketing research and framing effects. *Framing* refers to highlighting certain details (e.g., through wording or use of visuals) to promote a particular interpretation of information [13, 16, 60].

There are numerous examples of framing in political science, marketing literature, and psychology [13, 22, 34, 60]. A classic example of framing in psychology is Tversky and Kahneman [60]'s Asian Disease study, where participants were asked to choose between one of two possible medicines for a rare disease outbreak: option A would save 200 out of 600 lives, guaranteed; option B had a $\frac{2}{3}$ chance of saving no one, and a $\frac{1}{3}$ chance of saving all 600 people. 72% of participants selected option A. However, when the same choices were framed negatively (i.e., 400 people will die for option A), 78% of participants selected option B. The experiment showed how framing choices positively or negatively could result in marked differences in risky behavior; people are more risk averse when choices are positively framed (200 people will be saved) verses negatively framed (400 will die) [60]. In an example of framing in consumer choice, consumers rated beef almost 20% more tasty (a difference of 1.26 on a 7 point scale) when it was framed as 75% lean versus 25% fat [34].

In the political context, frames are used to decide how a controversial issue is defined or presented, emphasizing some aspects of an issue while ignoring others to influence public opinion. Often these frames become so widespread that a simple phrase or slogan can evoke them (e.g., *The war on terror*) [11, 13]. Sniderman and Theriault [55] found that 85% of survey respondents were in favor of allowing a hate group to hold rallies when the question was prefaced with a free speech reference, whereas only 45% were in favor of it when introduced with a warning about violence [55]. Other examples of diverse and competing political frames are gun control vs. personal freedom in the American gun debate, traditional morality vs. egalitarianism in the same-sex marriage debate, and media violence vs. weak gun laws in the gun safety and school shootings debate [9, 13, 22]. Previous work has identified framing dimensions that cut across political issues–such as morality, security and defense, and cultural identity–implying that political frames relate to one another in non-issue-specific ways [8, 11].

The recent development of online platforms has also enabled researchers to study framing effects online [5, 58]. Berger and Milkman [5] found that news stories evoking more positive emotions tended to be shared more often than negative stories, and that stories with high emotional arousal were more likely to go viral. Tan et al. [58] explored the effect of wording in message propagation on Twitter, finding that tweets (from the same author and on the same topic) that were worded to match the language style of a community and the author's prior messages, or that mimicked news headlines, were shared more often. Additionally, the language in job postings can have substantial impacts on who applies [56]. Significantly more men apply to job posts that use words or phrases like *"tackle"* or *"prove that"* while significantly more women apply to jobs advertised with words such as *"meaningfully"* or *"passion for learning"* [56].

2.2.1 Framing in Pro-Social Activities. Framing has also been shown to be a useful strategy in promoting *pro-social activities*, such as charitable giving, organ donation and voting [10, 12, 48]. In one instance, respondents reported significantly higher intentions to volunteer and donate to child poverty initiatives after being shown a negatively framed slogan (*"There are no silver spoons for children born into poverty. Without your donation, their life would be hopeless."*) compared to a positively framed slogan (*"If only every child was born with a silver spoon. With your donation, their life could become hopeful."*) [12]. A study on voter turnout found that voter survey questions framed to invoke the personal identity of being a voter, as opposed to the behavior of voting, significantly increased turnout [10]. In addition, Bond et al. [7] showed that framing voting in a social context (i.e., by showing friends who voted) attracted an estimated 340,000 more voters (0.14% of the voting age population in 2010) than a simple ad. Other factors that have been shown to affect participation in pro-social activities, like blood donation and donating data to research, are perceived need, organization reputation, and social signals [3, 37, 42]. Although prior findings strongly indicate that framing is an effective way to recruit volunteers, and therefore could be effective at recruiting participants for online studies, frames are unique across situations, and those for one context do not necessarily apply to other scenarios [13, 16]. This paper seeks to identify what frames will be the most effective in the context of online studies with volunteers.

In the context of citizen science, Lee et al. [32] applied framing in recruiting for Zooniverse, basing frames on previously identified motivations to participate [32]. They found that framing recruitment messages in terms of contributing to science was almost twice as effective at attracting new participants compared to framing messages in terms of joining a community [32]. Although useful in identifying potential frames, the study did not adapt messages to specific motivations. This paper explores the possibility of certain frames being more effective for particular users relative to others, providing a first step in adapting slogan frames to the participant.

Studies have found that message framing is influenced by people's values and personality traits [22, 33]. A frame supporting a person's prior beliefs or values is often more effective than one contradicting those values. For example, in a study exploring the framing of the American gun debate, Democrats responded more strongly to a frame that blamed gun violence on weak gun safety laws compared to Independents or Republicans [22]. Frames can also be more or less effective depending on personality traits [33]. Levin et al. [33] showed that compared to the average participant, those scoring low on the Conscientiousness or high on the Agreeableness personality characteristics in the Big Five Personality Inventory [15] were less likely to prefer the risky choice for the negative frame in a study similar to Tversky and Kahneman [60]'s Asian Disease study.

In a similar vein, tailoring advertisements and slogans to specific user populations has been much more effective in a number of marketing scenarios [24, 25, 61]. For example, banner advertisements for cars that changed based on a user's inferred cognitive style (e.g., impulsive vs.

deliberate) led to an increase of user click-through rates and purchase likelihood compared to generic advertisements [61].

These personal differences in framing effects and marketing research can have interesting implications for the purpose of online experiments. In citizen science or advertising the makeup of users recruited is less important than in online experimentation, where the sample of participants recruited can lead to systematic sample biases if message framing indeed attracts participants differently. We explore this possibility, identifying useful frames for online study recruitment and measuring their effect on collecting diverse (or not so diverse) participant samples.

3 EXPERIMENTS

We conducted two studies on LabintheWild:

(1) **Slogan Comparison**: A study that collected participant slogan preferences for three existing LabintheWild studies.
(2) **Self-Selection Analysis**: An A/B testing study that measured the likelihood of participants selecting a study based on different slogans. We instrumented five LabintheWild studies' ending pages with suggestions of other LabintheWild studies the volunteer could try. The slogans used for each suggestion were randomly chosen from our slogan options in Study 1.

Study 1 allowed us to establish clear slogan preferences and their relation to motivations and demographics, since each participant had to rank all possible slogans. Study 2 strengthened the external validity of these findings, showing how these preferences do indeed lead to differences in participant recruitment for online studies.

3.1 Study 1: Slogan Comparison

3.1.1 Slogan Generation and Classification. To generate a diverse set of slogans for this study, we first decided on the following three LabintheWild studies, chosen to represent a variety of tasks:

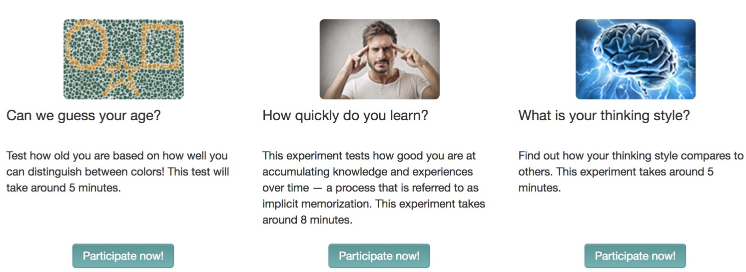

Fig. 1. The three studies for which we generated slogans.

(1) **Color Perception**, which explores how participants' demographics, specifically age and gender, influence their ability to distinguish colors. The current slogan for the study on the LabintheWild front page is, *"Can we guess your age?"*
(2) **Implicit Memory**, which tests how well participants acquire new knowledge from experiences over time [31]. The study asks participants to predict the weather (i.e., rainy or sunny) based on a set of icons presented. The current slogan for the study on the LabintheWild front page is, *"How quickly do you learn?"*
(3) **Thinking Style**, which analyzes how participants group new information. The study categorizes participants as either holistic or analytic thinkers based on a series of tasks that ask them

to group images and words together. The current slogan for the study on the LabintheWild front page is, *"What is your thinking style?"*

We recruited five participants (two female) to generate alternative slogans through a brainstorming session. We gave the participants an explanation of the concept of a slogan and the content of each of the three studies. We did not tell them the current slogan to prevent fixation on one idea or frame. We also did not tell participants about the motivations or ask them to create slogans in specific frames, seeking to place as few restrictions on slogan generation as possible. For each study, participants were given five minutes and instructed to generate as many and as diverse slogans as possible individually. They were then instructed to select the three they found most attractive. Afterwards, each member shared their selected slogans with the group. As a result, the participants generated 15 slogans for each study, for a total of 45 slogans.

Following the procedure in Lee et al. [32], where they based recruitment frames off of previously identified motivations to participate in citizen science projects, we categorized the slogans into frames matching user motivations for taking online experiments. The initial frames were based on the motivations found in Jun et al. [29]: supporting science, learning about oneself, comparing oneself to others, boredom, and fun. Two authors independently coded the slogans into these frames, giving multiple codes in order of importance in the case of disagreement, then met and resolved any differences in the coding. Due to comments from both coders on the difficulty and arbitrariness of differentiating the fun and bored frames, we grouped the fun and bored slogan frames together (now referred to as FUN *&* BORED). The final set of frames were: supporting science (SCIENCE), learning about oneself (SELF-LEARN), comparing oneself to others (COMPARE), and fun and boredom (FUN *&* BORED). After brainstorming and coding the slogans, one author contributed two additional slogans per study to increase the diversity of frames; these new slogans were then coded independently by the second author and resolved following the procedure above. Six slogans were selected for each study to test.

Before resolving differences, the inter-rater reliability using Cohen's Kappa over the entire set of slogans (including the two slogans added per study) was $\kappa = .63$. This somewhat low inter-rater agreement speaks to the difficulty of coding frames [11]. By resolving differences after independently coding all slogans, we sought to arrive at frames that we were confident would be consistent and intuitive. We also report on individual slogan performance. Table 1 identifies the slogans and their associated study and frames.

3.1.2 Procedure. After agreeing to the informed consent, participants filled out a demographics questionnaire including questions about their age and gender, as these demographic variables have been shown to affect volunteer engagement [29, 46]. The experiment then proceeded with 45 slogan pairs (15 pairs for each study) and asked participants to choose one of each in response to the question: *"Which study are you more likely to participate in?"* Slogan pairs were presented in random order. Each slogan was presented along with the logo and the same description of the study (see figure 2). To gather motivation data we followed the procedure in Jun et al. [29] and had participants rate their motivation to volunteer in a LabintheWild study using five 5-point Likert-type scales for the five motivations (fun, boredom, compare, self-learn, science) with text-anchors from 1 ('Not at all') to 5 ('Very much'). The five motivation scales were presented in a randomized order. We deployed this experiment on LabintheWild and shared the link via social media (i.e., Facebook and Twitter). We used two alternate slogans to advertise the study: *"See behind the scenes of LabintheWild!"* for one month and *"What will the next study be called on LabintheWild?"* for two months. We swapped slogans on the possibility that the slogans might attract participants with different motivations to participate.

Table 1. Slogans generated for each study, with the coded frames.

Slogan	Study	Frame
Are you a super color perceiver?	Color Perception	SELF-LEARN
What countries see more colorfully?	Color Perception	FUN & BORED
Discover the kaleidoscope in your eyes!	Color Perception	FUN & BORED
Help us discover how people see color differently!	Color Perception	SCIENCE
Do you see colors like others?	Color Perception	COMPARE
Shine a light on vision research!	Color Perception	SCIENCE
Can you predict the weather?	Implicit Memory	FUN & BORED
Help us build a roadmap of memory!	Implicit Memory	SCIENCE
Can you decode an alien language?	Implicit Memory	FUN & BORED
Test your subconscious!	Implicit Memory	SELF-LEARN
How does your memory stack up to others?	Implicit Memory	COMPARE
Contribute to our understanding of understanding!	Implicit Memory	SCIENCE
Do you think like others?	Thinking Style	COMPARE
Are you more like a scientist or a buddha?	Thinking Style	FUN & BORED
How is your brain wired?	Thinking Style	SELF-LEARN
How do people think differently?	Thinking Style	SCIENCE
Compare your thinking style!	Thinking Style	COMPARE
Support research on the way we think!	Thinking Style	SCIENCE

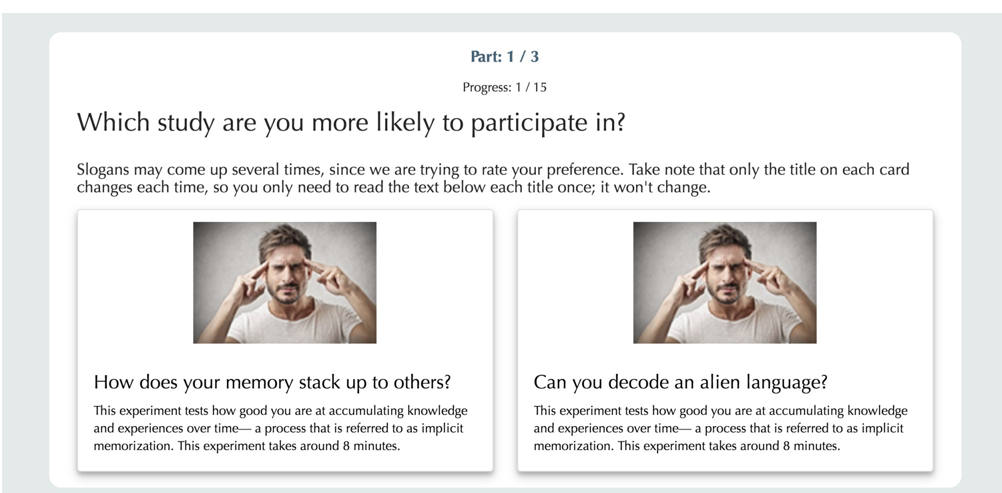

Fig. 2. Example of slogan choices participants saw as part of our slogan comparison study.

3.1.3 Participants. A total of 157 participants took part. After excluding participants who did not fully complete the study, were repeat testers, or indicated in the comments that they merely clicked through the study, our data set included 59 participants (33 female). Participants' mean age was 28.47 (sd = 15.57). Most participants came from the United States (35 participants) and had completed some college education (57 participants). Table 2 outlines participants' motivations.

Table 2. Average participant motivations for Study 1 and 2. Following the procedure in Jun et al. [29], participants were asked, *"To what extent are you participating in an experiment on LabintheWild for the following reasons?"* Motivations were ranked on five 5-point Likert-type scales (one for each motivation) from 1 ('Not at all') to 5 ('Very much').

Motivations	Mean (sd) - Study 1	Mean (sd) - Study 2
Comparison	3.37 (1.24)	3.38 (1.37)
Self-learn	3.86 (1.22)	4.29 (1.02)
Science	4.22 (0.87)	3.69 (1.25)
Fun	3.83 (1.19)	4.15 (1.03)
Boredom	2.78 (1.45)	3.25 (1.41)

3.1.4 Analysis. We first compared the motivation ratings we collected with the ratings found in Jun et al. [29], and found that the distribution of motivations in our sample were consistent with their findings. We used the Copeland Counting strategy [53] to rank slogan preference. Each slogan received a score calculated as the number of times it was selected. The slogan scores then vary from 0 (never selected) to 5 (always selected).

To analyze which slogan scores differed significantly from one another, we conducted a repeated measures ANOVA and Tukey post hoc tests across the scores of each slogan. The independent variable here was the slogan (e.g., *"Do you see colors like others?"*), while the dependent variable was the slogan score. We used a repeated measures ANOVA in order to control for the fact that the samples of each slogan's scores were not independent, as every participant saw every slogan. All post hoc analyses used the Benjamini-Hochberg correction [4] for multiple hypothesis testing. We augmented post hoc tests with Cohen's d effect size, which reports the magnitude of differences in means between samples, independent of sample size [57].

To answer how slogan frame preferences differ depending on participant motivations and demographics, we ran correlation tests for each of the participant motivations, age, and the Copeland count score for each slogan frame. To control for the possible influence of collinearity between motivations, we used partial correlations [17], which let us explore the relationship between the frame score and a single motivation while controlling for other motivations. We used Pearsons's r correlation coefficient for all correlation tests. Additionally, to see how gender influenced slogan preference, we compared the scores men and women gave to each slogan frame. In order to remain as close to prior work on these categories of participant motivations as possible, we follow the procedure in [29] and [40] to use parametric tests and mean reporting on the motivation ratings (see Pell [44] and Jamieson [27] for more information on this distinction).

All analyses were made in Python using the *statsmodels* and *SciPy* libraries [28, 52]. Our dataset and analysis scripts are publicly available at https://github.com/talaugust/slogan_testing.

3.1.5 Results.

What slogan frames are most preferred by participants in volunteer-based online experiments? Slogan scores differed significantly across slogans ($F_{17,59} = 11.49, p < .00001$). Table 3 summarizes the largest differences in scores for specific slogans in each study.

Over slogan frames, the top scoring frame was SELF-LEARN (m=3.19, sd = 1.43), then COMPARE (mean=2.97, sd=1.38), FUN & BORED (m=2.45, sd=1.61) and finally SCIENCE (m=1.90, sd=1.50). Given that there were ambiguities in coding the slogan frames, it is important to recognize the slogans themselves that tended to perform best. Slogans with terms like *"others," "test,"* and *"super,"* which

Table 3. Top and bottom slogan choices for the three studies averaged across all participants in Study 1. Scores can range between [0, 5], from never chosen to always preferred. Effect sizes are reported with Cohen's *d*.

Study	**Bottom choice (frame)**	**Top choice (frame) (d)**	**t-test**
Thinking Style	*Support research on the way we think!* (SCIENCE) (m=.95, sd=1.29)	*Do you think like others?* (COMPARE) (m=3.24, sd=1.23) (d=1.81)	p<.001
Color Perception	*Shine a light on vision research!* (SCIENCE) (m=1.75, sd=1.55)	*Are you a super color perceiver?* (COMPARE) (m=3.37, sd=1.36) (d=1.12)	p<.001
Implicit Memory	*Contribute to our understanding of understanding!* (SCIENCE) (m=1.61, sd=1.53)	*Test your subconscious!* (SELF-LEARN) (m=3.10, sd=1.53) (d=.98)	p<.001

were coded as COMPARE or SELF-LEARN, tended to outperform slogans with words like *"support"* and *"research"* (SCIENCE frame). Figure 3 plots the mean score for each slogan, grouped by framing.

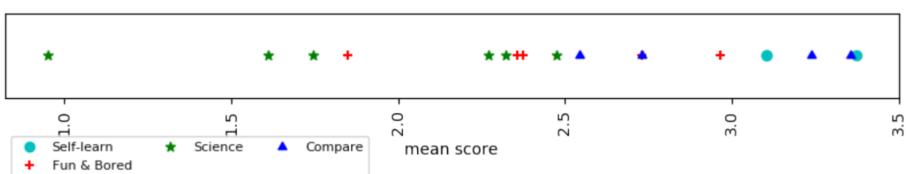

Fig. 3. Average slogan scores on a scale from 0 (never selected) to 5 (always selected) in Study 1 for all slogans grouped by frame.

Do slogan frame preferences differ depending on participant motivations and demographics? When controlling for motivations and age using partial correlations, the compare motivation was significantly correlated with the scores of the COMPARE frame ($\rho = .28$, $p < .05$). All other partial correlations between slogan scores and motivation and age ranged from $\rho = [-.17, .15]$ and were not significant. There were no significant differences in slogan scores between men and women.

The findings suggest that the SELF-LEARN and COMPARE slogan frames are most attractive to participants on average. Additionally, there is a slight relationship between participants' motivation and their preference for certain slogan frames (e.g., participants who rate highly their motivation to compare tend to also rate slogans in the COMPARE frame higher).

3.2 Study 2: Click-through rate analysis

Our second study was designed to find out whether preferences for specific slogan frames translate into participants selecting studies advertised with these frames.

3.2.1 Materials. LabintheWild studies present personalized feedback pages to participants after they complete experiments. These pages link to other LabintheWild studies, aiming to generate more traffic on the site. Currently, these links use static slogans to advertise studies. An example of links on the feedback page is shown in Figure 4.

As an instrumentation procedure for our participant self-selection analysis, we randomized the slogans presented to participants. In each case, two of the three LabintheWild studies from Study 1 were randomly presented to the participant on the feedback page. We only adapted the slogan

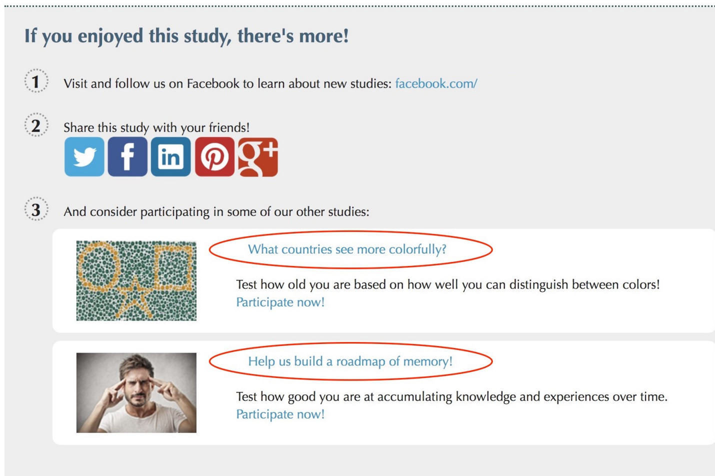

Fig. 4. Example of study links at the end of a study feedback page; A/B tested slogans are circled.

for the study (see Figure 4), while the thumbnail and description of the study stayed the same. We measured whether participants clicked on one of the two studies. The click-through rate was defined as the ratio of participants who clicked on the slogan to a specific study compared to the total number of participants who saw the same slogan.

We instrumented five LabintheWild studies:

- The same three studies as used in Study 1 (see study details in Section 3.1).
- **Website colorfulness and complexity**, a study that asked participants to rate the colorfulness and visual complexity (i.e., information clutter) of different websites.
- **Memory**, a study that asked participants to memorize and later recall sequences of numbers.

Each participant who finished any of these five studies was presented with two additional studies (out of the three for which slogans had been created in Study 1). The studies ranged in length from 5-10 minutes. The types of feedback pages presented to participants at the end of the studies were a diverse mixture of social comparisons, teaching participants more about science, and teaching participants more about themselves.

3.2.2 Procedure. After agreeing to the informed consent, participants in each of the five instrumented studies answered the same motivations questionnaire described in 3.1. Following the questionnaire, participants completed the study. At the end of the study, participants saw their feedback page, which contained at the bottom the two randomized slogans advertising additional studies. Due to the variable length of the feedback pages and the size of participants' screens, sometimes the participant needed to scroll down the page in order to see the randomized slogans. Each study also presented participants with a demographics form. Depending on the study, this form was either before the motivations questionnaire or before the feedback page at the end of the study. The demographics form was already integrated into each study, and thus we were unable to change its ordering to remain consistent across instrumented studies.

3.2.3 Participants. During 3 months in 2018, 3,055 participants completed the five studies. The participants came from 111 different countries, and education levels ranged from pre-high school to postdoctoral studies. For analysis, we excluded any participants who did not provide motivation

Table 4. Descriptive statistics for our second study sample.

	N	Mean age (sd)	% Female	% Male
Thinking Style	1232	26.88 (11.87)	54.87	44.13
Implicit Memory	440	24.49 (10.94)	53.18	46.82
Color Perception	785	26.27 (14.30)	49.55	50.44
Colorfulness & Complexity	104	28.79 (11.62)	51.92	48.08
Memory	109	24.47 (10.92)	48.62	51.38

ratings. We also excluded participants who declined to provide their age or gender and those who chose 'other' for gender because there were too few (< 3% of the overall sample) to provide a statistically viable sample. The final dataset totalled 2,670 participants. Table 4 presents more detail of participant composition across all five studies.

Overall 536 participants decided to take an additional study ("clicked through"); a click-through rate of 17.55%. After excluding participants who did not provide motivation or demographic information, we had a sample of 485 participants who clicked through.

3.2.4 Analysis. To find out how the use of different slogan frames affects online study recruitment, we ran chi-squared tests between the click-through rates of each slogan, as well as the average click-through of slogans grouped by frame. The chi-squared tests let us see if the click-through frequency for different slogans and frames significantly differed from a uniform click-through rate. When calculating click-through rate frequency, we used the original sample of 536 participants, since these analyses did not rely on participant motivations or demographics.

Finally, to analyze whether slogan frames recruit participants with different motivations or demographics, we conducted independent samples t-tests over the samples of participants each slogan frame recruited. Our goal was to see whether participants had significantly different motivation ratings or demographics, which would suggest that the slogan frame used did in fact recruit a biased sample of participants. We applied the Benjamini-Hochberg correction [4] for multiple hypothesis testing. We additionally augmented our tests with Cohen's d effect size [57]. For gender, we ran chi-squared tests between the sample of men and women recruited by each slogan frame. To augment our chi-squared tests, we used the ϕ coefficient, which is similar to Cohen's d as an effect size calculation for categorical variables [18].

3.2.5 Results.

How does using different slogan frames affect online study recruitment? Each slogan was presented an average of 336 times (sd = 63.38) and had an average click-through rate of 8.75% (sd = 3.21%). The rates of participants selecting the first or second study in the results footer (see Figure 4) were not significantly different (50.17% vs. 48.88%, respectively) ($\chi^2_2 = 0.09$, $p = .76$, $\phi = .01$).

Slogans had significantly different click-through rates ($\chi^2_{17} = 66.79$, $p < .05$). As in Study 1, slogans in the SELF-LEARN frame tended to perform best. Table 5 shows the slogans with the highest and lowest click-through rates for each study (thinking style, implicit memory, and color perception) and whether this difference was significant. Difference in slogan click-through rates are substantial: in some cases changing from a SCIENCE slogan to a COMPARE slogan (advertising the same study) can increase click-through rate by almost 9%.

SELF-LEARN performed best in terms of click-through rates (mean = 10.44%, sd = 3.0%), then FUN & BORED (mean= 9.38%, sd=4.14%), COMPARE (mean=8.64%, sd= 2.90%), and finally SCIENCE performed the worst (mean=7.45%, sd =2.91%). Looking at the top and bottom performing slogans, it is again clear that slogans with words like *"others"* and *"test"* (*"Test your subconscious!"* performed

Table 5. Top and bottom performing slogans for the three studies. Effect sizes are reported with ϕ coefficient for categorical variables.

Study	**Bottom choice (frame) (% click-through)**	**Top choice (frame) (% click-through) (ϕ)**	**Chi-squared**
Thinking Style	*Support research on the way we think!* (SCIENCE) (2.90%)	*Do you think like others?* (COMPARE) (11.51%) (ϕ = .58)	p<.05
Color Perception	*What countries see more colorfully?* (FUN & BORED) (8.68%)	*Discover the kaleidoscope in your eyes!* (FUN & BORED) (12.42%) (ϕ = .07)	n.s.
Implicit Memory	*Can you predict the weather?* (FUN & BORED) (4.15%)	*Can you decode an alien language?* (FUN & BORED) (14.52%) (ϕ = .56)	p<.05

second-best in the Implicit Memory Study with 13.3%), which were coded often as SELF-LEARN or COMPARE, led to better slogan performance compared to words like *"support"* (coded as SCIENCE). Figure 5 plots the mean score for each slogan, grouped by framing.

Slogan and frame performance are similar to Study 1, with one study having the exact same top and bottom slogans (i.e., *"Do you think like others?"* and *"Support research on the way we think!"*) and the same top and bottom performing frames overall (SELF-LEARN and SCIENCE, respectively).

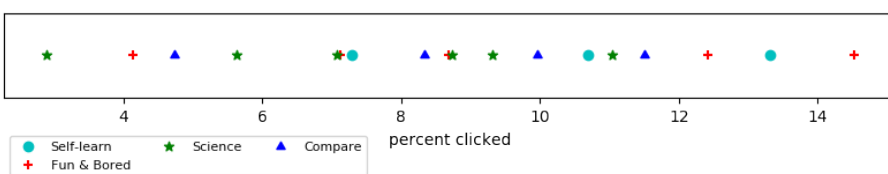

Fig. 5. Slogan click-through rates for Study 2.

Do slogan frames recruit participants with different motivations or demographics? Interestingly, the sample of participants who clicked through at all (i.e. selected to take an additional study, regardless of slogan) had significantly higher ratings of all motivations but boredom compared to participants who did not click through. Table 6 provides details between these differences in motivations.

The sample of participants recruited by the SELF-LEARN frame rated the bored motivation (mean=3.57, sd=1.35) significantly higher than the sample of participants recruited by the FUN & BORED frame (mean=3.14, sd=1.35) (t_{244} = 2.39, p< .05, d=.32) and the COMPARE frame (mean=3.09, sd=1.35) (t_{182} = 2.39, p< .05, d=.35). Figure 6 outlines the differences in average motivation of the samples recruited by each frame. Notice that, although not statistically significant, the SCIENCE frame also recruited participants who rated the matching science motivation highest. There were no significant differences in the age or gender of participants recruited by each frame.

These results, coupled with the slight but significant correlation in Study 1, suggest that there is a small effect of participant motivation on slogan preference, and that different slogan frames do lead to slight differences in the motivational makeup of participants recruited into a study.

4 DISCUSSION AND DESIGN IMPLICATIONS

This paper explored differences in participant slogan frame preferences for online experiments with volunteers. Are some slogan frames best at attracting volunteers? Do certain types of participants prefer specific frames more than others? If so, what sorts of slogans will attract whom?

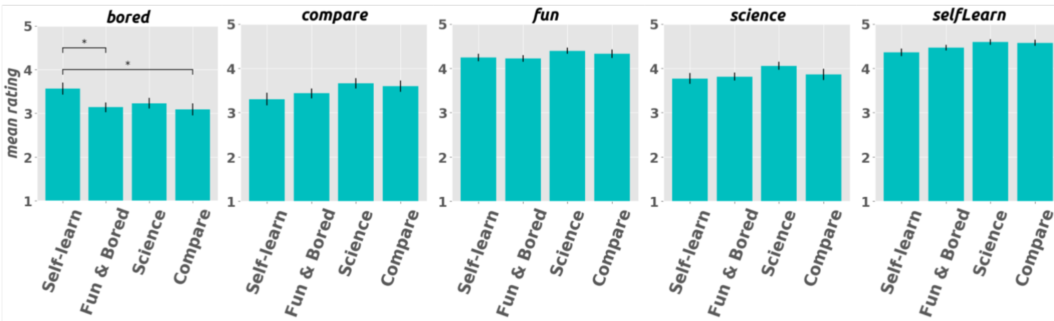

Fig. 6. Average motivation for participant samples recruited by each frame. Error bars are standard errors. * denotes $p < .05$ for independent samples t-test.

Table 6. Motivations for participants who did and did not click through. * ($p < .01$), ** ($p < .0001$))

Motivation	Click-through (N=485) (mean, sd)	No click-through (N=2185) (mean, sd)	Significance (effect size)
Comparison	mean=3.51, $sd = 1.37$	mean=3.35 $sd = 1.37$	* $t_{2668} = -2.27$, ($d = .11$)
Self-learn	mean=4.51, $sd = .77$	mean=4.24 $sd = 1.06$	** $t_{2668} = -5.11$, ($d = .27$)
Science	mean=3.89, $sd = 1.15$	mean=3.65 $sd = 1.27$	** $t_{2668} = -3.75$, ($d = .19$)
Fun	mean=4.30, $sd = .85$	mean=4.12 $sd = 1.06$	* $t_{2668} = -3.48$, ($d = .18$)
Boredom	mean=3.13, $sd = 1.44$	mean=3.22 $sd = 1.43$	$t_{2668} = .252$, ($d = .01$)

We found that participants prefer different slogans, and that these preferences are influenced by the participants' motivations, but not by their age or gender. Further, these preferences translate into differences into who self-selects into studies advertised with different slogans.

Some slogan frames generally recruited better than others: in the thinking style study, for example, using a slogan framed as comparing to others (COMPARE) increased recruitment by close to 9% compared to using a slogan framed as contributing to science (SCIENCE). Across all studies, using a slogan framed as learning more about oneself (SELF-LEARN) boosted recruitment an average of 3% over SCIENCE slogans. The findings suggest that altruistic recruitment is less successful than promising participants the possibility of learning about themselves or comparing to others.

The results also show that different slogan frames recruit participants with different motivations, leading to a sample bias in terms of participant motivation for taking online studies. The SELF-LEARN frame recruited participants who rated the bored motivation significantly higher than the sample of participants recruited by either the FUN & BORED or the COMPARE frame. One explanation for this finding is that the SELF-LEARN slogans tended to be sensationalized while still promising concrete feedback (e.g., *"Are you a super color perceiver?"* rather than something like *"How do you see color?"*). Considering that participants motivated by boredom are in general less likely to take additional studies, the SELF-LEARN slogans' combination of sensationalism and promised feedback appeared to be more effective at convincing these participants to take another study. This is in contrast to the COMPARE slogans, which tended to just promise feedback (e.g., *"Do you see colors like others?"*) and the FUN & BORED slogans, which tended to be more sensationalized with less concrete feedback (e.g., *"Discover the kaleidoscope in your eyes!"*). In fact, this combination of promised concrete feedback and excitement in the slogan language can also help explain why the SELF-LEARN slogans preformed best at recruiting participants in general.

The fact that participants with certain motivations were more likely to self-select into studies advertised with specific frames suggests that how a researcher advertises a study can lead to sample biases in terms of participant motivation. A participant's motivation for taking an online study can influence the data they provide [26, 29], suggesting that these types of sample biases can impact study results. Additionally, collecting a sample biased to those only motivated for a single reason can cause issues depending on what a researcher is studying. For example, if a researcher is running a study on boredom (e.g., [39]), then collecting a participant sample biased towards those who are all taking the study because they are bored could prove dangerous to the generalizability and reliability of the researcher's results.

In the following, we explore the design implications of our work. Researchers can use this new understanding of how frames affect study recruitment to both recruit more volunteers, as well as reduce sample biases and strengthen the validity of their online studies.

Use COMPARE or SELF-LEARN to boost recruitment. The most effective frame overall for recruitment was SELF-LEARN, while in certain studies COMPARE preformed best. These slogans often had words like *"test"* or *"others"* in them (e.g., *"Do you think like others?"*). Using either of these frames can increase recruitment substantially over the SCIENCE frame, whose slogans contained words such as *"support"* or *"research"* (e.g., *"Support research on the way we think!"*). This finding was consistent across both studies, suggesting that researchers interested in boosting recruitment can employ one of these two frames.

Change your frame to gather participants with diverse motivations. Where above we talked about using a SELF-LEARN frame to recruit participants faster, here we recognize that slogans in this frame might attract only certain types of participants. Using only a SELF-LEARN frame can recruit participants who are more motivated out of sheer boredom to take studies, which past research has shown can lead to increased drop-out rates and less attentive participants, threatening study results and validity [29]. Additionally, our own results show that participants motivated by boredom are less likely to take additional studies than participants with other motivations. This can be an issue for platform providers who seek to increase recruitment on a number of studies through participants taking additional studies.

When deciding on slogan frames to use in recruitment, researchers must recognize what sort of bias they might introduce into their study. One way a researcher can control the sample biases introduced by slogan frames is to cycle the frames they use. For example, a researcher who wants to gather participants with diverse motivations can cycle the slogan frame they use across SCIENCE, FUN & BORED, SELF-LEARN and COMPARE slogan frames. Each could be displayed for a set amount of time, possibly giving the SCIENCE frame longer exposure due to its slower recruitment rate. Each frame could attract different participants at different times, providing the researcher an overall sample that is more diverse than using one frame would have allowed.

5 LIMITATIONS AND FUTURE WORK

The current paper assumed that participant motivations were static, i.e., would not change during the course of a study. It is possible that taking a specific study, and seeing specific results, could prime [54] a participant to certain motivations. For example, an interesting personalized result might make a participant more motivated to learn about themselves (SELF-LEARN). It would be interesting to explore how taking an online study affects a participant's motivations by either surveying participants before and after studies or measuring behaviours to predict motivations.

Additionally, Study 1 recruited a participant pool of almost exclusively college-educated participants (57 out of 59). Study 1's skew might have explained some of the differences in Study 1 slogan ratings compared to Study 2 slogan performance.

The relatively low inter-rater agreement in categorizing slogan frames, as well as some clashes in matching motivations (e.g., the SELF-LEARN frame recruited participants more motivated by boredom than the FUN & BORED frame) suggest that the frames may have been too coarse-grained, hiding other framing factors that might affect slogan attractiveness. For example, we grouped all SCIENCE slogans into a single frame, yet there are other factors that can influence a participant to contribute to scientific research, such as perceived need, organization reputation, and social signals [3, 37, 42]. These difficulties and ambiguties reveal the importance of quantifying frames as clearly as possible when testing different slogans [11].

Work in Natural Language Processing (NLP) has developed code books and corpora for frames found in the media, mapping phrases and words to the frames that they invoke [11], as well as to more general semantic categories [51], yet currently these tools don't capture nuanced frame semantics in recruitment practices. Researchers using frames for online study recruitment could help build these tools with the NLP community by providing annotated corpora of slogans or recruitment messages, as well as explore the possibility of adapting general purpose framing tools (e.g., [51]) to the focus of online study recruitment. Exploring more nuanced, quantifiable frames by developing these tools could provide exciting insights into how to adjust a frame to be as effective as possible based on different participant motivations.

There are also a host of other possible dimensions that influence the attractiveness of a slogan or text, such as length, alliteration, or sentiment. These features are more about form than content, but have been shown to affect preference and attractiveness [1, 58]. Language influences our interpretation of the world in a myriad of ways; analyzing the biases this introduces into the online context is an exciting future direction of this research.

6 CONCLUSION

This is the first paper that examines how advertising online experiments, specifically through the use of slogan framing, attracts different populations. Our results suggest that slogans that incentivize participants with the benefit of learning something new about themselves, such as personalized feedback or social comparisons, can substantially increase recruitment compared to slogans appealing to participants' motivation to support science. Additionally, different frames attract participants with different motivations: slogans advertising studies as learning more about oneself attracted participants most motivated by boredom. Our results provide new guidelines for researchers attempting to recruit online volunteers while mitigating sample biases.

7 ACKNOWLEDGEMENTS

We greatly thank the LabintheWild participants who make this research possible. We also thank the reviewers for their time and input, and Alexis Hiniker for her comments on earlier drafts of the paper.

REFERENCES

[1] Adam L Alter and Daniel M Oppenheimer. 2009. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* 13, 3 (2009), 219–235. https://doi.org/10.1177/1088868309341564

[2] Judd Antin and Aaron Shaw. 2012. Social Desirability Bias and Self-reports of Motivation: A Study of Amazon Mechanical Turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2925–2934. https://doi.org/10.1145/2207676.2208699

[3] Timothy C. Bednall and Liliana L. Bove. 2011. Donating Blood: A Meta-Analytic Review of Self-Reported Motivators and Deterrents. *Transfusion Medicine Reviews* 25, 4 (Oct 2011), 317–334. https://doi.org/10.1016/J.TMRV.2011.04.005

[4] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1995), 289–300. https://doi.org/10.2307/2346101

[5] Jonah Berger and Katherine L Milkman. 2012. What Makes Online Content Viral? *Journal of Marketing Research* 49, 2 (apr 2012), 192–205. https://doi.org/10.1509/jmr.10.0353 arXiv:arXiv:1011.1669v3

[6] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis* 20, 3 (2012), 351–368. https://doi.org/10.1093/pan/mpr057

[7] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (sep 2012), 295–298. https://doi.org/10.1038/nature11421

[8] Amber E Boydstun, Dallas Card, Justin Gross, Philip Resnik, and Noah A Smith. 2014. Tracking the development of media frames within and across policy issues. (2014).

[9] Paul R Brewer. 2003. Values, Political Knowledge, and Public Opinion about Gay Rights. *Public Opinion Quaterly* 67 (2003), 173–201. https://doi.org/10.1086/374397

[10] Christopher J Bryan, Gregory M Walton, Todd Rogers, and Carol S Dweck. 2011. Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences* 108, 31 (August 2011), 12653–12656. https://doi.org/10.1073/pnas.1103343108

[11] Dallas Card, Amber E Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* 2, 438–444.

[12] Chun Tuan Chang and Yu Kang Lee. 2009. Framing charity advertising: Influences of message framing, image valence, and temporal framing on a charitable appeal. *Journal of Applied Social Psychology* 39, 12 (2009), 2910–2935. https://doi.org/10.1111/j.1559-1816.2009.00555.x

[13] Dennis Chong and James N Druckman. 2007. Framing Theory. *Annual Review of Political Science* 10 (2007), 103–26. https://doi.org/10.1146/annurev.polisci.10.072805.103054

[14] Vickie Curtis. 2015. Motivation to Participate in an Online Citizen Science Game: A Study of Foldit. *Science Communication* 37, 6 (Dec 2015), 723–746. https://doi.org/10.1177/1075547015609322

[15] J.M. Digman. 1990. Personality structure: The emergence of the five-factor model. *Annual Review of Psychology* 41 (1990), 417–440. https://doi.org/10.1177/017084068800900203 arXiv:arXiv:1011.1669v3

[16] Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43, 4 (1993), 51–58.

[17] Ronald Aylmer Fisher et al. 1924. The distribution of the partial correlation coefficient. *Metron* 3, 3-4 (1924), 329–332.

[18] Joseph L Fleiss. 1994. Measures of effect size for categorical data. *The handbook of research synthesis* (1994), 245–260.

[19] Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review* 19, 5 (2012), 847–857. https://doi.org/10.3758/s13423-012-0296-9

[20] Jorge Gonçalves, Vassilis Kostakos, Evangelos Karapanos, Mary Barreto, Tiago Camacho, Anthony Tomasic, and John Zimmerman. 2014. Citizen motivation on the go: The role of psychological empowerment. *Interacting with Computers* 26, 3 (may 2014), 196–207. https://doi.org/10.1093/iwc/iwt035

[21] Anja S Göritz. 2006. Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science* 1, 1 (2006), 58–70. http://www.ijis.net/ijis1_1/ijis1_1_goeritz.pdf

[22] Donald P. Haider-Markel and Mark R Joslyn. 2001. Gun policy, opinion, tragedy, and blame attribution: The conditional influence of issue frames. *Journal of Politics* 63, 2 (2001), 520–543. https://doi.org/10.1111/0022-3816.00077

[23] Joshua Hartshorne. 2018. Games With Words. https://www.gameswithwords.org/

[24] John R. Hauser, Glen L. Urban, Guilherme Liberali, and Michael Braun. 2009. Website Morphing. *Marketing Science* 28, 2 (mar 2009), 202–223. https://doi.org/10.1287/mksc.1080.0459 arXiv:1606.07988

[25] Jos Hornikx and Daniel J. O'Keefe. 2009. Adapting Consumer Advertising Appeals to Cultural Values A Meta-Analytic Review of Effects on Persuasiveness and Ad Liking. *Annals of the International Communication Association* 33, 1 (jan 2009), 39–71. https://doi.org/10.1080/23808985.2009.11679084

[26] Gary Hsieh and Rafał Kocielnik. 2016. You Get Who You Pay for: The Impact of Incentives on Participation Bias. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM Press, New York, New York, USA, 821–833. https://doi.org/10.1145/2818048.2819936

[27] Susan Jamieson. 2004. Likert scales: how to (ab)use them. *Medical Education* 38 (2004), 1212–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

[28] Eric Jones, Travis Oliphant, Pearu Peterson, et al. 2001–. SciPy: Open source scientific tools for Python. http://www.scipy.org/ Online.

[29] Eunice Jun, Gary Hsieh, and Katharina Reinecke. 2017. Types of Motivation Affect Study Selection , Attention , and Dropouts in Online Experiments. *Proceedings of the ACM on Human-Computer Interaction Archive* 1, 1 (2017), 1–15. https://doi.org/10.1145/3134691

[30] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. Worker Motivation in Crowdsourcing - A Study on Mechanical Turk. *Proceedings of the Seventeenth Americas Conference on Information Systems* 4, 2009 (2011), 1–11. https://doi.org/10.1145/1979742.1979593

[31] B J Knowlton, L R Squire, and Mark A Gluck. 1994. Probabilistic classification learning in amnesia. *Learning and Memory* 1, 2 (1994), 106–120. https://doi.org/10.1101/lm.1.2.106

[32] Tae Kyoung Lee, Kevin Crowston, Mahboobeh Harandi, Carsten Østerlund, and Grant Miller. 2018. Appealing to different motivations in a message to recruit citizen scientists: results of a field experiment. *JCOM, Journal of Science Communication* 17, 01 (2018), A02–2.

[33] Irwin P Levin, Gj Gaeth, Judy Schreiber, and Marco Lauriola. 2002. A New Look at Framing Effects: Distribution of Effect Sizes, Individual Differences, and Independence of Types of Effects. *Organizational Behavior and Human Decision Processes* 88, 1 (2002), 411–429. https://doi.org/10.1006/obhd.2001.2983

[34] Irwin P Levin and Gary J Gaeth. 1988. How Consumers are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Source Journal of Consumer Research* 15, 3 (1988), 374–378. http://www.jstor.org/stable/2489471

[35] Qisheng Li, Krzysztof Z Gajos, Katharina Reinecke, and Paul G Allen. 2018. Volunteer-Based Online Studies With Older Adults and People with Disabilities. *ACM Conference on Computers and Accessibility (ASSETS)* (2018). https://doi.org/10.1145/3234695.3236360

[36] Yang Liu, Roy Chen, Yan Chen, Qiaozhu Mei, and Suzy Salib. 2012. 'I loan because...': understanding motivations for pro-social lending. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. ACM Press, New York, New York, USA, 503. https://doi.org/10.1145/2124295.2124356

[37] Yong Liu, Denzil Ferreira, Jorge Goncalves, Simo Hosio, Pratyush Pandab, and Vassilis Kostakos. 2017. Donating context data to science: The effects of social signals and perceptions on action-taking. *Interacting with Computers* 29, 2 (2017), 132–146. https://doi.org/10.1093/iwc/iww013

[38] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (mar 2012), 1–23. https://doi.org/10.3758/s13428-011-0124-6 arXiv:http://ssrn.com/abstract=1691163

[39] Merrifield, Colleen. 2014. *Toward a Model of Boredom: Investigating the Psychophysiological, Cognitive, and Neural Correlates of Boredom*. Ph.D. Dissertation. http://hdl.handle.net/10012/8671

[40] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 15, 5 (dec 2010), 625–632. https://doi.org/10.1007/s10459-010-9222-y

[41] Oded Nov, Ofer Arazy, and David Anderson. 2011. Dusting for science. In *Proceedings of the 2011 iConference on - iConference '11*. ACM Press, New York, New York, USA, 68–74. https://doi.org/10.1145/1940761.1940771

[42] Oded Nov, Ofer Arazy, and David Anderson. 2014. Scientists@Home: What Drives the Quantity and Quality of Online Citizen Science Participation? *PLoS ONE* 9, 4 (apr 2014), e90375. https://doi.org/10.1371/journal.pone.0090375

[43] M. L. Ohmer. 2007. Citizen Participation in Neighborhood Organizations and Its Relationship to Volunteers' Self- and Collective Efficacy and Sense of Community. *Social Work Research* 31, 2 (jun 2007), 109–120. https://doi.org/10.1093/swr/31.2.109

[44] Godfrey Pell. 2005. Use and misuse of Likert scales. *Medical Education* 39, 9 (sep 2005), 970–970. https://doi.org/10.1111/j.1365-2929.2005.02237.x

[45] ProjectImplicit. 2011. ProjectImplicit. https://www.projectimplicit.net/index.html

[46] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2013. Galaxy Zoo: Motivations of citizen scientists. *arXiv preprint arXiv:1303.6886* (2013).

[47] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. *Proceedings of the 18th ACM conference on Computer Supported Cooperative Work & Social Computing* (2015), 1364–1378.

[48] Amber Marie Reinhart, Heather M Marshall, Thomas Hugh Feeley, and Frank Tutzauer. 2007. The Persuasive Effects of Message Framing in Organ Donation: The Mediating Role of Psychological Reactance. *Communication Monographs ISSN:* 74, 2 (2007), 229–255. https://doi.org/10.1080/03637750701397098

[49] UD Reips, B Batinic, W Bandilla, M Bosnjak, and L Graf. 1999. Financial incentives personal information and drop-out rate in online studies. In *Current Internet Science Trends, Techniques, Results*. Deutsche Gesellschaft für Online-Forschung.

[50] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2010. Who are the Turkers? Worker Demographics in Amazon Mechanical Turk. *Human Factors in Computing Systems (CHI)* (2010), 2863–2872. https://doi.org/10.1145/

1753846.1753873
[51] Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice.* Institut für Deutsche Sprache, Bibliothek.
[52] J.S. Seabold and J. Perktold. 2010. Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference.*
[53] Nihar B Shah and Martin J Wainwright. 2015. Simple, Robust and Optimal Ranking from Pairwise Comparisons. *arXiv* (2015), 1–17. arXiv:1512.08949 https://arxiv.org/pdf/1512.08949.pdfhttp://arxiv.org/abs/1512.08949
[54] Steven J. Sherman, Diane M. Mackie, and Denise M. Driscoll. 1990. Priming and the Differential Use of Dimensions in Evaluation. *Personality and Social Psychology Bulletin* 16, 3 (sep 1990), 405–418. https://doi.org/10.1177/0146167290163001 arXiv:0803973233
[55] P.M. Sniderman and S.M. Theriault. 2004. The Structure of Political Argument and the Logic of Issue Framing. *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change* (jan 2004), 133–164.
[56] Keiran Snyder. 2017. 1000 different people, the same words - Textio Word Nerd. https://textio.ai/1000-different-people-the-same-words-6149b5a1f351
[57] Gail M Sullivan and Richard Feinn. 2012. Using Effect Size-or Why the P Value Is Not Enough. *Journal of graduate medical education* 4, 3 (sep 2012), 279–82. https://doi.org/10.4300/JGME-D-12-00156.1
[58] Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic-and author-controlled natural experiments on Twitter. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14)* (2014).
[59] TestMyBrain. 2017. TestMyBrain. http://dx.plos.org/10.1371/journal.pone.0165100
[60] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458. https://doi.org/10.1126/science.7455683
[61] Glen L. Urban, Guilherme (Gui) Liberali, Erin MacDonald, Robert Bordley, and John R. Hauser. 2014. Morphing Banner Advertising. *Marketing Science* 33, 1 (jan 2014), 27–46. https://doi.org/10.1287/mksc.2013.0803
[62] VolunteerScience. 2018. Volunteer Science. https://volunteerscience.com/
[63] Heng Li Yang and Cheng Yu Lai. 2010. Motivations of Wikipedia content contributors. *Computers in Human Behavior* 26, 6 (nov 2010), 1377–1383. https://doi.org/10.1016/j.chb.2010.04.011
[64] Marc A. Zimmerman. 1995. Psychological empowerment: Issues and illustrations. *American Journal of Community Psychology* 23, 5 (oct 1995), 581–599. https://doi.org/10.1007/BF02506983

Received April 2018; revised July 2018; accepted September 2018