Should We Use an Abstract Comic Form to Persuade? Experiments with Online Charitable Donation

ANONYMOUS AUTHOR(S)

This paper examines the use of the abstract comic form over the plain text message for making online charitable donations. Persuading individuals to contribute to charitable causes online is hard and responses to appeals are typically low; charitable donations share the structure of public goods dilemmas where the rewards are distant and non-exclusive. In this paper, we examine if comics in abstract form are more persuasive than the plain text. Drawing on a rich literature on comics, we algorithmically synthesize a three panel abstract comic to create our appeal. We conducted a between-subject Amazon Mechanical Turk study with 307 participants on the use of abstract comic form to appeal for charitable donation. As part of our experimental procedure, we sought to persuade individuals to contribute to a real charity focused on Autism research. We compared the average donation to the charity under three conditions: the plain text message, an abstract comic that includes the plain text, and an abstract comic that additionally includes the social norm. We use Bayesian modeling to analyze the results, motivated by model transparency and their use in small-sized studies. Our experiments reveal that the message in abstract comic form elicits more donations than text (medium to large effect size=0.59). Incorporating social proof in the abstract comic message did not show a meaningful effect. Our studies have design implications: non-profits and governmental agencies interested in alleviating public goods dilemmas that share a similar structure to our experiment (single-shot task, distant, non-exclusive reward) ought to consider including messages in the abstract comic form as part of their online messaging.

Additional Key Words and Phrases: statistical facts, abstract comics, persuasion, information framing, social proof, hierarchical bayesian models

1 INTRODUCTION

Are persuasive text messages more effective when expressed in comic form? Text messages abound, asking us to act, either towards personal wellness goals (e.g. in an exercise app:"you need 2,000 more steps to reach 10,000 steps for the day"), reminders (e.g. push notifications from a task tracking app: "pick up prescription from the pharmacy at 6PM"), or appeals from charities online (e.g. from a Wikipedia campaign on the Wikipedia homepage: "If everyone reading this donated $5 our fundraiser would end today. Please donate to keep Wikipedia free."). Making these short text messages more effective has important implications in stimulating behavior, especially for non-profits and governmental agencies interested in alleviating public goods dilemmas.

There is a rich history of prior work in psychology that has examined how variations in the construction of the text message alter decisions. In a seminal paper, Tversky and Kahneman [55] examined information framing: can we elicit different decisions, by describing the outcome either as a loss or a gain, even though the two messages are indistinguishable in terms of expected utility? They found that individuals are risk-seeking when confronted with certain loss and risk-averse in the presence of gains. Cialdini examined the introduction of social norms in two different contexts: towel re-use in hotels [24], reduction of power consumption [53]. In the former study, replacing the standard text message in hotels asking the customer to re-use towels, with the social norm (e.g. "75% of the people in this hotel have re-used their towels"), increased towel re-use. The latter study had an intriguing result—while introducing the social norm in the message *did not* reduce overall energy consumption, adding emoticons to the message with the social norm made a difference. Specifically, they added a smiley face '☺' when the customer consumed less energy than average, and a frowny '☹' face when their energy consumption was worse than average. By adding emoticons average

2019. XXXX-XXXX/2019/4-ART $15.00
https://doi.org/

energy consumption decreased. The Schultz et al. [53] study motivates us to examine the role of the comic form, a highly expressive, affective medium, in communicating messages intended to stimulate a specific behavior.

McCloud [38] defines comics as "juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer." The simple and humorous nature of a comic makes it a unique medium for delivering informative and memorable messages; the use of metaphor in comics can make the underlying meaning vivid and more memorable than using a straightforward description [39, 38]. There has been significant prior work in educational contexts: illustration of scientific facts [39], educating end-users about computer-security [1] [57], communicating with multi-lingual audiences [12, 43].

Despite the importance of the comic form in contemporary culture, and the use of the comic form in educational settings, there is surprisingly limited work examining the effects of the comic form to stimulate behavior. In particular, we focus on the use of an abstract comic form. By an abstract representation, (see Figure 1 for an example) we mean that the comic de-emphasizes character detail (face, eyes, etc.) or details about the locale. As McCloud [38] points out, using abstract representations for the comic allows the reader to project themselves onto the comic character. The abstract form has an additional benefit: since the form is visually spare, it allows for comic panel algorithmic synthesis and for the personalization of messages. In this paper, we used a three-panel comic in abstract form.

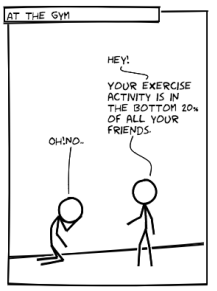


Fig. 1. An example single comic panel in abstract comic form.

How should we decide on the experimental context in which we could examine the effects of the abstract comic form in stimulating behavior? There are many different compelling behavioral contexts: personal wellness goals (e.g., diet, exercise), mundane tasks (e.g. "pick up dry cleaning"), as well as broader public-goods issues (e.g. "take the flu shot;" "donate to cure cancer")

Four design principles guide us, in our choice of the experimental behavioral context: nature of the reward; single shot tasks that preclude habit formation (a potential confound); an ecologically valid task; absence of specialized knowledge to perform the task. First, we would like the rewards to be distant, and non-exclusive, rather than proximal and exclusive so that individuals don't perform the task in anticipation of the immediate reward. Thus public goods dilemmas (e.g. "reducing carbon footprint;" "taking the flu shot"; "contributing to public knowledge") are all candidates. Second, while some longitudinal tasks (e.g., losing weight; eating healthy) have distant rewards (losing weight, or maintaining a diet takes time), and can positively affect the public good (with more healthy people, in the long-run, insurance rates will fall), these tasks are prone to habit formation, a potential confound. Furthermore, single-shot tasks such as "pick up yogurt at the grocery store today", often prompted by text reminders from our calendars or task-tracking apps, have an immediate, exclusive reward. Third, we would like to ensure that the experimental task is ecologically valid—a task that these individuals would be actually asked to perform in the wild, outside of the experimental context. Fourth, we would like the task not to require specialized knowledge (e.g. "asking doctors to make a decision"), so that other researchers could easily replicate and scale our experiment. Online charitable donation tasks satisfy these design principles as they are single-shot tasks, contribute to the public good with distant, non-exclusive rewards, occurs frequently, and easily tested for at scale. We chose a charity associated with Autism (Organization for Autism Research) for our experiment.

[1] Also used in introducing software: see https://www.google.com/googlebooks/chrome/big_00.html for an example.

Thus, in this work we answer two research questions in the context of the online charitable donation task:

RQ1: Does the use of the abstract comic form increase the level of donation over the plain text message? [2]

RQ2: What is the effect of introducing a social norm in the abstract comic, when compared to the comic without the social norm?

We would like to emphasize that it is unclear at the outset if a message in comic form ought to make any difference say towards a charitable donation. An argument in favor is that comics communicate affect well and are an important element of our visual culture; an argument against is that from rational actor theory in classical Economics, since the comic includes the text message used in the text-only condition and thus cannot alter the expected utility, the comic ought to serve as an irrelevant factor. To the best of our knowledge, this is the first study to examine the persuasive power of an abstract comic in an online public good setting with distant and non-exclusive rewards.

We ran the experiment using Amazon Mechanical Turk ($n = 307$; we paid each participant above minimum wage, at \$10.0/hr) where we randomly assigned each participant to one of the three conditions ('a plain text message', 'abstract comic', 'abstract comic with social norm'). The two comic conditions include the text used in the plain text condition. The social norm case adds one additional phrase to the comic that reveals the norm.

We performed a careful Bayesian analysis to analyze the data. We agree with the observations by Kay et al. [31], that beyond the impact on experimental replicability, shifting the question from "did it have an effect?" to "how strong was the effect?" is important to the HCI community. Two additional ideas—transparency, impact small-n studies—motivate our use of Bayesian analysis. First, with a Bayesian model, the researcher foregrounds all the aspects of the model; there are no modeling assumptions that need checking, not already foregrounded in the model description. Thus the model and the results are amenable to further scrutiny. Second, a Bayesian model is valid at *every* value of n (the number of observations); we do not have to wait for $n \geq 30$ to satisfy assumptions of Normality. For small n values, the result is of course affected by the choice of the prior; but by using weakly informative priors, we can ensure that the prior doesn't dominate inference. In a sense, as McElreath [40, Chapter 9] mentioned, Bayesian models make the *most conservative* inference given the data.

We would like to emphasize that Non-Bayesian statistics are powerful tools, and when used by an experienced statistician, they can dramatically simplify the process of inference. However, for researchers who wish to investigate their findings without access to a statistician, they need to be careful of the assumptions of the different tests: Normality (t-test); heteroscedasticity (e.g., ANOVA) and ensuring that the observations satisfy the assumptions required by the test; not doing so, can result in misleading inference.

Our findings: We show that using the comic has a meaningful increase in donations over the plain text conditions, with a medium to large effect size of 0.59. Thus we can answer **RQ1** affirmatively. To answer **RQ2**, we show that while the comic with social norm increases the donation level over the comic without the norm, the effect size is very small (0.11) and the increase is not meaningful. To summarize, the comic form meaningfully increases donations over the plain text, but the presence of the norm is not effective. We caution that the result holds for single-shot, public goods tasks. To fully understand the value of incorporating the social norm in the comic, for specific longitudinal tasks such as dieting, with distant, exclusive rewards and where habits can confound, we need future research.

[2]The abstract comic includes the same text from the plains text condition. Thus both conditions indicate the same expected utility.

Design Implications: The primary design implication of our findings: helping non-profits and governmental agencies with their online messaging strategies as they work to alleviate public goods dilemmas. In particular, public-goods dilemmas that require single-shot decisions (online charitable donations; taking the flu shot) are opportune candidates for intervention. We believe that these agencies can easily include the use of the comic form as part of their overall messaging strategy because the simplicity of the abstract comic form allows it to be easily synthesized (as we discuss towards of the end of this paper) and to additionally incorporate social norms.

The rest of this paper is organized as follows. In the next section, we discuss related work. Section 2 introduces two motivating ideas: the abstract comic form, and the idea of the social proof. Section 3 introduces our experimental method and in Section 4 we present our Bayesian analysis. In Section 5, we analyze the model for convergence and also discuss alternative models. Section 6 summarizes the main findings, identifies limitations and then discusses how we could algorithmically synthesize the comic panels. We present our conclusions in Section 7.

2 RELATED WORK

Now, we discuss related work in 1) Using Comics to Communicate Ideas, 2) Persuasion through Visual Stimuli, 3) Persuasion for Public Goods, and 4) Social Proof in Persuasive Messages.

2.1 Using Comics to Communicate Ideas

McCloud [38] defines comics as "juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer". Beyond obvious entertainment value, comics have been examined as an effective way of communicating abstract and complex ideas to broad audiences [39, 12, 38, 57]. On the one hand, the simple and humorous nature of comic makes comics becomes a unique media for delivering informative and memorable messages. By combining visual elements and texts, comics make the story more appealing. McDermott et al. [39] used comics to illustrate complex scientific facts. Bach et al. [4] explored graph comics as a medium to communicate dynamic networks changes. In education, comics have been used to reach populations with various backgrounds [39, 12, 38]. Zhang-Kennedy et al. [57] created "Secure Comics" to educate end-users with computer security knowledge. On the other hand, the use of metaphor in comics can make the underlying meaning vivid and more memorable than using a straightforward description [39, 38]. Moreover, comics can contain a personal story incredibly powerful for creating empathy, a key factor in persuasion, for readers [56]. Matsubara et al. [37] showed a link between the comic's content and the emotions felt by the readers. Thus complex messages can be easily interpreted and memorized through the use of the comic form. Although comics have shown strength in communicating complex and memorable ideas, their utility in persuasion has not yet been fully explored. To the best of our knowledge, this the first study to examine the use of abstract comics in persuading people to contribute in public goods dilemmas.

In our study, we chose abstract comics to persuade instead of other comic forms for the following reasons, First, as the comic becomes more abstract, readers will be more likely to project themselves onto the character and thus allowing the readers to empathizing with the character [38]. Second, comparing to other forms of the comic, the abstract comic contains the fewest visual elements; the simplicity also reduces reader's cognitive load to consume the message; important since persuadee's attention is scarce [29]. Moreover, the simplicity of abstract comics allows us to explore the idea of algorithmically synthesizing persuasive messages algorithmically into comic form. Therefore, in this study, we examined the persuasive power of messages in abstract comic form.

2.2 Persuasion Through Visual Stimuli

Beyond the realm of textual forms, prior research shows that using visual representations, e.g., graphics, video, and comics, are attractive in persuasion. Selker et al. [54] used motivational graphics or memes from 9GAG and Google to attract people's attention and persuade people for energy saving behaviors. Consolvo et al. [16] presents the user's exercise data as visual elements in the Ubifit Garden to persuade people to exercise. Sometimes, the use of visual can evoke strong emotions that makes the message more persuasive. Iyer and Oldmeadow [28] found that people who saw the images of the Kenneth Bigley kidnapping were more engaged in the later civic campaign than those who read about the kidnapping from the texts of the newspaper.Zhang-Kennedy et al. [58]'s visual rhetoric effectively persuade users to use up-to-date antivirus protection. Visual stimuli were more memorable as well [45]. The use of images makes advertisements more memorable and appealing. The visuals can leave a strong trace which may later on influence people's decision making, especially when people making judgments by the availability heuristics. Dey et al. [19] found the video in the crowdfunding campaign plays a vital role in persuading people to support. However, creating visual stimuli is often costly. The persuader needs to put time, often manual effort and resources to create persuasive visual stimuli. In our study, we looked into the abstract comic, simple visual stimuli that can be algorithmically synthesized and examined its persuasiveness in encouraging people to participate in online charitable donations.

2.3 Persuasion for Public Goods

Given the two characteristics of public goods, non-exclusive and non-rivalrous, individuals will receive no tangible benefits when acting in public goods dilemmas such as charitable donations [36, 27]. Therefore, external nudges play an important role in encouraging people to contribute. Researchers and policymakers have extensively studied who contributes and how to persuade people to contribute [46, 5, 2, 43, 11, 49, 11]. Midden et al. [42] reported strong persuasive power for environmental sustainable behavior when signaling personal goals in persuasive applications. Feiler et al. [21] found emphasizing altruistic reasons in donation requests can elicit more donations. Mankoff et al. [35] successfully used social technologies to leverage public commitment and competition in appealing energy-saving behaviors. Due to the distant or non-reward nature of individuals' public goods contribution, persuasive messages were the key.

Persuasive text messages are one of the most widely used methods to persuade individuals for voluntary charitable donations. They are easy to create and disseminate. Damgaard and Gravert [18] successfully used simple email reminders with the decision deadline to elicit charitable donation from the participants. With a simple sign like "Turn off the tap when not in us", people reduced water consumption and engaged in energy conservation behavior change [41]. Cotterill et al. [17] showed sending a pledge card with simple text "A list of everyone who donates a book will be displayed locally." encouraged 22% more households to donate books for Children's Book Week. However, comparing persuasive messages in other forms such as graphics, textual messages were harder to catch perusadee's attention, especially when persudaee's attention is limited. Moreover, when it comes to memorability, a key measure of persuasive effectiveness, studies found that in comparison to other media such as graphics and video, the text is most difficult to recall and recollect. Therefore, persuasive messages in other media form such as the abstract comic are worth investigating.

2.4 Social Proof in Persuasive Messages

By *social proof*, we refer to the idea that observing either their friends or people with whom they can relate adopted a behavior is persuasive for individuals to adopt the same behavior [14, 15].

The use of social proof is widely used in encouraging individuals to cooperate in collective action dilemmas [24, 53]. Goldstein et al. [24] conducted a famous experiment in a hotel on motivating environmental conservation. They found that descriptive norms (e.g., "the majority of guests reuse their towels") has more persuasive power than solely mentioning environmental protection. And this normative message gets more effective when the statement is about a provincial norm (e.g.,"the majority of guests in this room reuse their towels"). Amblee and Bui [1] studied social proof among online book reader communities and found that "electronic word of mouth " affects a book's quality, an author's reputation, and a book category's popularity which eventually influenced people's buying decision. Since the use of social proof is one of the most widely used influence weapons in creating persuasive messages, we want to see their effect when included in the persuasive messages in the abstract comic form.

3 METHOD

In this study, we examine if the abstract comic can impact a decision with monetary consequence on Amazon Mechanical Turk. The persuasion task is to ask participants to make a charitable donation decision to the Organization for Autism Research with the real money. As we mentioned in earlier sections, four design principles guide us to choose the online charitable donation task. First, the online charitable donation is a kind of collective dilemma where study participants will receive no external/distant rewards based on their decision, which signifies the impact from the persuasive message in study participants' decision-making process. Second, encouraging people to donate is a single shot task that requires no habit formation, which protects us from several confounding variables such as life events and resource constraints (e.g., time). Third, with the advance in social network and internet, although the task is still challenging, charities or organizations nowadays solicit donations online textual messages (e.g., Wikipedia's fundraising banner), which makes our task ecologically valid. Fourth, online charitable donation requires no specialized knowledge. In our study, we choose the Organization for Autism Research as Autism Spectrum Disorder (ASD), is a well-known developmental disorder that impairs communication and behavior [3]. In other words, ASD provides basic interest for the participants to support the related charitable organization. In this section, we will introduce the experiment design and describe our study participants recruiting process.

3.1 Experiment Design

Since the main goal for this study is to compare the power of a persuasive message in two forms, the abstract comic and the text, we first constructed two experimental conditions, comic condition (see Figure 2) and text condition. In the comic condition, participants will read a message asking if they are willing to support a charity in a three-panel abstract comic strip, whereas in the text condition, participants will receive the same message in pure textual form. To test the idea of social proof, we then added a third condition, comic with the social proof, a three-panel comic that has the same content as the comic condition but added one line text that indicates the normative behavior (see Figure 3).

The objectives of the persuasive message in all three conditions: persuading participants to donate to a charity from his/her own pocket. Similar to Lee and Hsieh [34], the money participants will use is part of their study compensation, a prospective bonus reward (10% chance of winning $5 bonus). We randomly assigned study participants to one of the three conditions; the participants are free to make a decision on the amount of donation, including not donating at all.

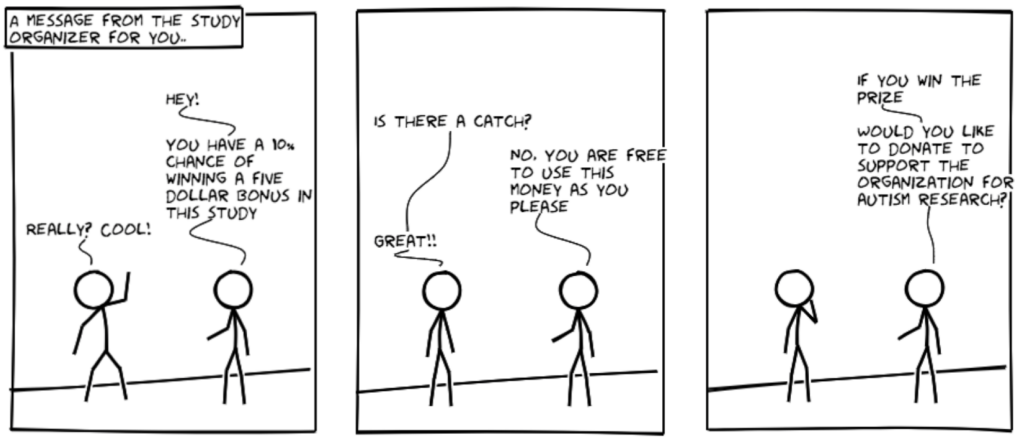


Fig. 2. Messages in the abstract comic form. Same as the text messages, the three-panel comic strip communicates three major objectives, 1) Participants will have 10 % of a chance winning $ 5 bonus upon the completion of the study (see the first panel). 2) Participants are free to use the money as they please (see the second panel). and 3) Participants can donate this bonus to the Organization for Autism Research (OAR) (see the third panel).

3.2 Generating Comics

In this study, we created a comic generator to generate the comic strips used in the study. The generator leveraged an open source comic library [26] to generate comic figures and used "rough.js" [51] to generate other elements such as text bubble and outlines. The generator impersonates the style found on 'XKCD' [44] comic. Our focus is not the XKCD style, but the fact that the generated comic is abstract. We believe the generator has the potential to be further developed as a general framework to automatically synthesize pure-textual persuasive messages into abstract comic forms (we discuss this aspect further in section Section 6).

Study Procedure. Once participants consented to join the study, we ask them if they are familiar with the Autism Spectrum Disorder (ASD). Then, each participant watched a short video produced by the Organization for Autism Research that promotes its fundraising activity "RUN FOR AUTISM". After watching the video, we asked participants to summarize the video using free text and ask them to provide their opinion about the effectiveness of the video. The recruiting message specifically mentioned this task of soliciting their opinion on video message's effectiveness. There are two main goals for this part of the study; First, we want to make sure prior familiarity with autism won't confound our study; Second, we want our main task less intrusive as soliciting charitable donation is not a common task on Amazon Mechanical Turk.

We then randomly assigned participants to each of the three conditions and then ask them to read the corresponding persuasive message (text, three-panel comic, three-panel comic with social proof). In the message, we provided the participant with a 10% chance of winning $5 additional compensation. We also provide them with the opportunity to donate to the Organization for Autism Research (OAR) which is the charity mentioned in the video they watch as part of the study.

To best demonstrate the persuasiveness of the message itself, we diffused the responsibility of donation amount among all participants. Similar to Lee and Hsieh [34], before the participants make their decision, they read "The total amount of money allocated to [the charity] by all the

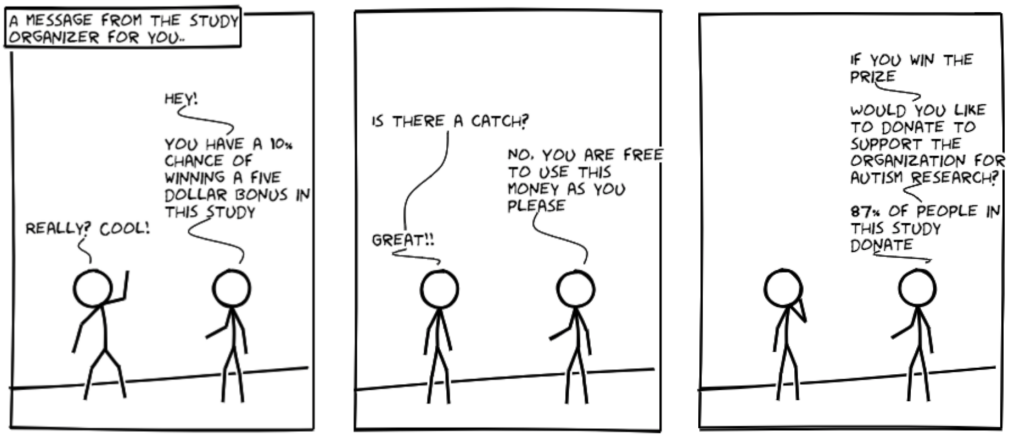


Fig. 3. Messages with social proof. Additional to the three major objectives, the comic with social proof communicates the idea of the social proof at the third comic panel. The figure "87%" comes from our pilot study

winning participants will be aggregated and donated at the end of the study." Then, we asked the participants to decide the amount of money they are willing to donate on a slider bar with \$0 and \$5 as two extreme ends. The default position of the slider bar is at the \$0 end.

Before leaving the study, we asked the participants to fill a demographic questionnaire about their gender, age, and education; the participants had the option of declining to state an answer for each question.

At the end of the study, we randomly chose 10% of the participants, donated to OAR based on the participants' decision, and rewarded each chosen participant the part of the bonus that they wished to keep.

To gather the basic statistics to create social proof, we first ran a pilot study with the first two conditions ($n = 60$) and used the donation statistics as part of the social proof message. In the pilot study, 87% of the participants donated a non-zero amount.

Organization for Autism Research (OAR). To increase the realism of our study, the donation decision is not hypothetical and participants are rewarded based on their decision and we as study organizers donate to the Organization for Autism Research (OAR). We chose OAR as the charity in our study for three reasons. First, Autism Spectrum Disorder (ASD), is a well-known, serious developmental disorder that impairs communication and behavior. In other words, ASD provides basic interest for the participants to support the related charitable organization. Second, the Organization for Autism Research is one of the most visible ASD related organizations that helps individuals with autism and provides assistance to parents, families, teachers, and caregivers. The goal of OAR is clear and reputable so participants won't question the authenticity of our message's motive. Finally, we wished to avoid a charity associated with a life-threatening condition such as cancer as it may create an experimental confound: we don't know if someone donates because their intrinsic desire to help with a life-threatening situation. While ASD can have serious consequences on the well being of those who have it, the public perception is that ASD is not-life threatening.

Persuasive Messages. The persuasive messages communicate three major objectives, 1) Participants will have 10 % of a chance winning $ 5 bonus upon the completion of the study. 2) Participants are free to use the money as they please. and 3) Participants can donate this bonus to the Organization for Autism Research (OAR). Therefore, in the text condition, study participants will read the following message,

> *You have a 10% chance of winning a five dollar bonus in this study. You are free to use this money as you please. If you win the prize, would you like to donate to support the Organization for Autism Research?*

In the two comic conditions, we created three-panel comic strips to communicate the *same* text message. The three-panel comic strip allows us to leverage one of the most fascinating aspects of comics—storytelling [38]. Consistent with "match on action" technique [38], we matched the panels on the gesture of the first character (the message recipient), while retaining a neutral gesture for the second character (who delivers the message).

In the comic with social proof condition, we created social-proof by adding one sentence on the last comic panel indicates the percentage of people in our study donated see Figure 3. The percentage (87%) corresponds to the number of people who donated a non-zero amount in the pilot study.

3.2.1 Participants Recruitment. In this study, we recruit our study participants from Amazon Mechanical Turk. Although Amazon Mechanical Turk is a crowdsourcing platform that has been widely used to gather human intelligence in AI research and social science experiments [47, 7, 10], we should be cautious when using such platform as the participant selection criteria is not transparent [33, 48]. In our study, we chose the Amazon Mechanical Turk for the following reasons. First, our persuasion task is about online charitable donation which targets internet users. Second, crowdsourcing platforms will help us reach a more diverse sample than using the researchers' own social network to attract participants [10]. Third, the main motive for Amazon Mechanical Turk workers is monetary rewards [7]. In our study design, people will more sensitive to the monetary reward they may get and carefully make the donation decision, which makes our persuasive message more crucial. Thus, we believe it is valid for our experiment to use Amazon Mechanical Turk as our sample pool.

Motivated by studies that show that populations on Amazon Mechanical Turk are diverse and mirror the US population [10, 6, 7], we recruit our participants from Amazon Mechanical Turk. However, we are aware of research that raises concerns in Amazon Mechanical Turk's sample representativeness [33, 48]. One potential solution is to use a panel company's population (e.g., Qualtrics). However, this method also has concerns in that the researcher can not directly cross-validate the sample's representativeness we have to trust the company's assertion.

We published our HITs on Amazon Mechanical Turk titled "A short survey about communicating autism campaign ads". The compensation was $10/hr, and the workers would get these rewards regardless of their performance. To ensure quality, the HIT is limited to English Speakers and people who have a 95% Approval Rate. On the HIT page, we instructed that repeated responses would be rejected. We told the participants that they would see a link to our experiment site.

4 STUDY ON PERSUASION: RESULTS

In this section, we will first present the raw data used in the analysis, then introduce the Bayesian Model we used for data analysis and our analysis result.

4.1 Raw Data

In total, we have 307 participants joined our study, 101 participants received the message in the text form, 102 participants received the same message in the abstract comic form, and 104 participants received the abstract comic message with social-proof. We ran an outlier analysis using Tukey Fence on participants' completion time and removed 30 participants from our dataset. The following analysis is based on a dataset with 277 participants, 91 of them is in pure-text condition, 97 of them is in comic condition, and 92 of them is in comic-social-proof condition. Of those 277 participants, 150 were self-identified as male, and 126 were self-identified as female, 1 participant chose not to disclose. 197 participants earned at least college degree. The median annual household income of our study participants is $50,000 - $59,999. All participants were familiar with the Autism Spectrum Disorder and acknowledged the importance of autism research.

Figure 4 compares the distributions of the charitable donations across the three conditions. Among all 307 participants, 253 (82.4%) participants donated non-zero amount to support the autism research; 74 (73.3%) participants from the text condition, 86 (84.3%) participants from the comic condition, and 93 (89.4%) participants from the comic with social proof condition.

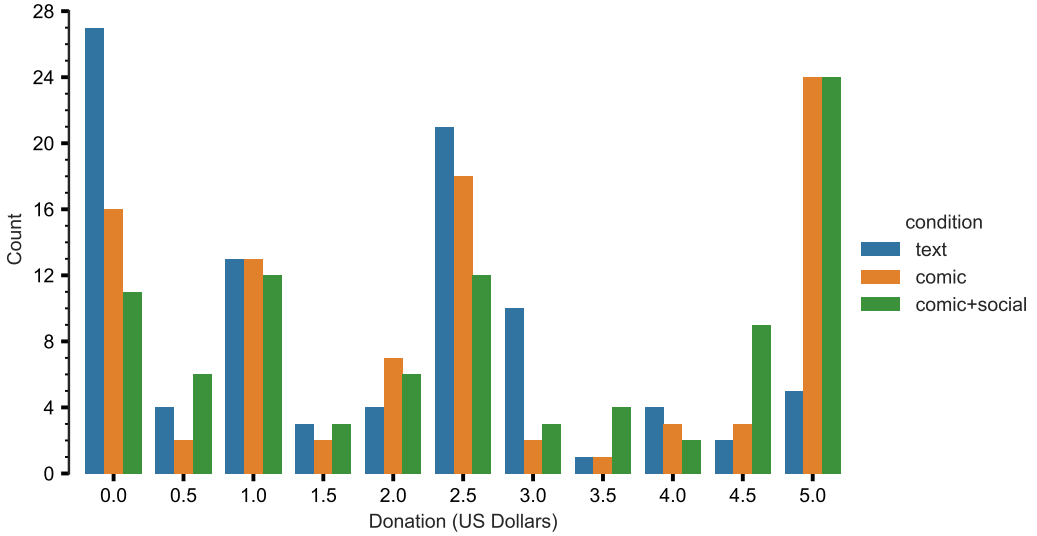


Fig. 4. The distribution of the amount of money participants decide to donate to the charity in each of the three conditions: text only; comic; comic with social proof. Among all 307 participants, 253 (82.4%) participants donated non-zero amount to support the autism research; 74 (73.3%) participants from the text condition, 86 (84.3%) participants from the comic condition, and 93 (89.4%) participants from the comic with social proof condition.

4.2 Bayesian Formulation

We use a Bayesian formulation of the problem of identifying suitable predictors for the messages in comic form. Kay et al. [31] provide a nice introduction on the appropriateness of Bayesian analysis for the HCI community. Bayesian analysis is attractive in our experiment due to two advantages: shifting the conversation from "did it work" to "how strong is the effect"; and benefits to small n studies.

Our experiment has three experimental groups: text, comic and comic with the social proof. One way to use classical statistical analysis for multiple groups would be to use ANOVA to see if the treatment (the use of comic) has any effect. Bayesian inference can sometimes differ from the standard ANOVA test widely used to compare treatment effects.

This is because ANOVA assumes equal variances within groups, and that the response within each group is drawn from a Normal distribution—in practice, both assumptions may be violated. It is straightforward in Bayesian Analysis to relax both assumptions: equal variances and Normality. The Normality assumption may be one reason why ANOVA and the t-test may be less sensitive to differences than Bayesian analysis [32, p. 470].

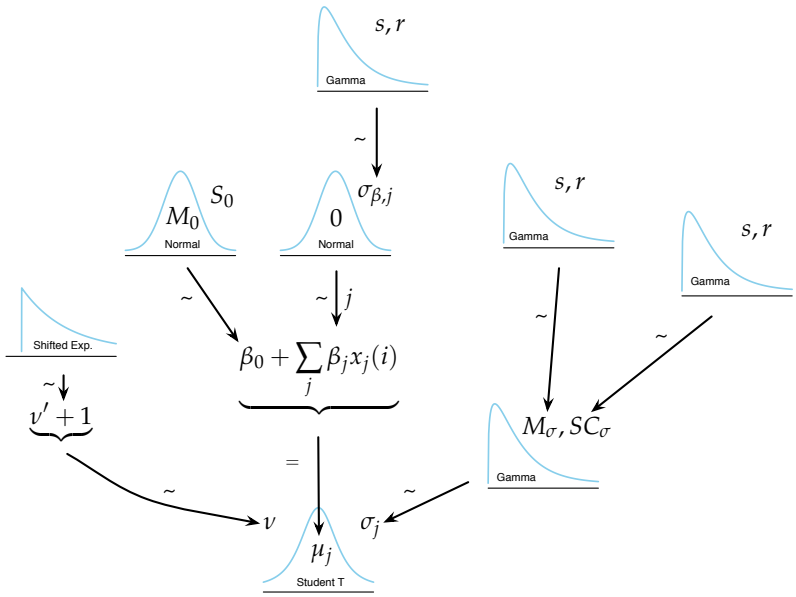


Fig. 5. Our proposed Hierarchical Bayesian model.

Now, we discuss our Bayesian formulation (see Figure 5 for a graphical view of the model). There is one outcome variable $y_{i|j}$, the amount of donation to the charity by each participant i, under experimental condition j: text, comic, and comic with social proof. Our Bayesian model:

$$
\begin{aligned}
y_{i|j} &\sim \text{StudentT}(\nu, \mu_j, \sigma_j), && (1)\\
\nu &\sim 1 + \exp(\lambda), && (2)\\
\mu_j &\sim \beta_0 + \sum_j \beta_j x_j(i), && (3)\\
\sigma_j &\sim \Gamma(M_\sigma, SC_\sigma), && (4)\\
M_\sigma &\sim \Gamma(s, r), && (5)\\
SC_\sigma &\sim \Gamma(s, r), && (6)\\
\beta_0 &\sim N(M_0, SD_0), && (7)\\
\beta_j &\sim N(0, \sigma_{\beta,j}), && (8)\\
\sigma_{\beta,j} &\sim \Gamma(s, r). && (9)
\end{aligned}
$$

Equation (1) says that the response $y_{i|j}$ of each group j is modeled as a StudentT distribution with mode μ_j, scale σ_j and with ν degrees of freedom; notice that this is a *drawing distribution*, not the t-test. The StudentT allows us to model a non-Normal outcome with a heavy tail; assuming $y_{i|j}$ to be Normally distributed is equivalent to setting $\nu = \infty$. The degrees of freedom ν is drawn from a shifted exponential distribution, to ensure $\nu \geq 1$; the mode μ_j corresponding to each group is drawn from a Normal distribution with $\mu_j = \beta_0 + \sum_j \beta_j x_j(i)$, where the group indicator variable $x_j(i) = 1 \iff$ subject i belongs to group j. The scale σ_j of the normal distribution is drawn from a Gamma distribution $\Gamma(M_\sigma, SC_\sigma)$, with mode M_σ and scale SC_σ. The mode M_σ and scale SC_σ are each drawn from two independent Gamma distributions $\Gamma(s, r)$ with shape parameter s and rate parameter r. The overall group response β_0 is modeled as a Normal distribution with mean μ_0 and variance σ_0. For each of the β_j for each group j in Equation (3), β_j is Normally distributed with mean $\mu = 0$ and $\sigma_{\beta,j}$ drawn from a Gamma distribution $\Gamma(s, r)$ with shape parameter s and rate parameter r. The random variables β_j are centered around $\mu = 0$ so that the group responses are modeled as deflections around the overall mean β_0.

Figure 5 indicates that we draw all the variances $\sigma_{\beta,j}$ from a Gamma $\Gamma(s, r)$ distribution, where s refers to the shape parameter and r refers to the rate parameter. We set the variables s, r to allows a wide range of values for $\sigma_{\beta,j}$. Notice that we draw the variances $\sigma_{\beta,j}$ of each group j independently from a *common* Gamma distribution, implying that the variances (equivalently, the extent of deflections from the mean) for each predictor β_j can be different. The main advantage of using a Gamma distribution is that we can specify a non-zero mode, important in controlling shrinkage in hierarchical models.

We draw the scale variables σ_j of the likelihood StudentT distribution from a common Gamma distribution, whose rate and shape parameters are also drawn from Gamma distributions. The advantage of this hierarchical approach is that the values of each element of σ_j inform the other elements. The "information sharing" among variables common to hierarchical Bayesian models and is an important reason why Bayesian models work so well with small datasets[3]. Finally, the constants M_0, S_0, s, r are set so that the priors are generous but weakly informative so that despite exploring all possible values, we ensure rapid MCMC convergence.

[3]The sharing of information causes the scale of each group σ_j to move towards the group variance, a phenomenon known as "shrinkage."

4.3 Analysis

We analyzed the data using PyMC3 [52], a popular framework for Bayesian inference. Computational techniques for Bayesian inference use a stochastic sampling technique called Markov Chain Monte Carlo (MCMC) that samples the posterior distribution $P(\theta|D)$, where we want to estimate the parameters θ given the observations D. In particular, we used the No-U Turn Sampler (NUTS) sampler.

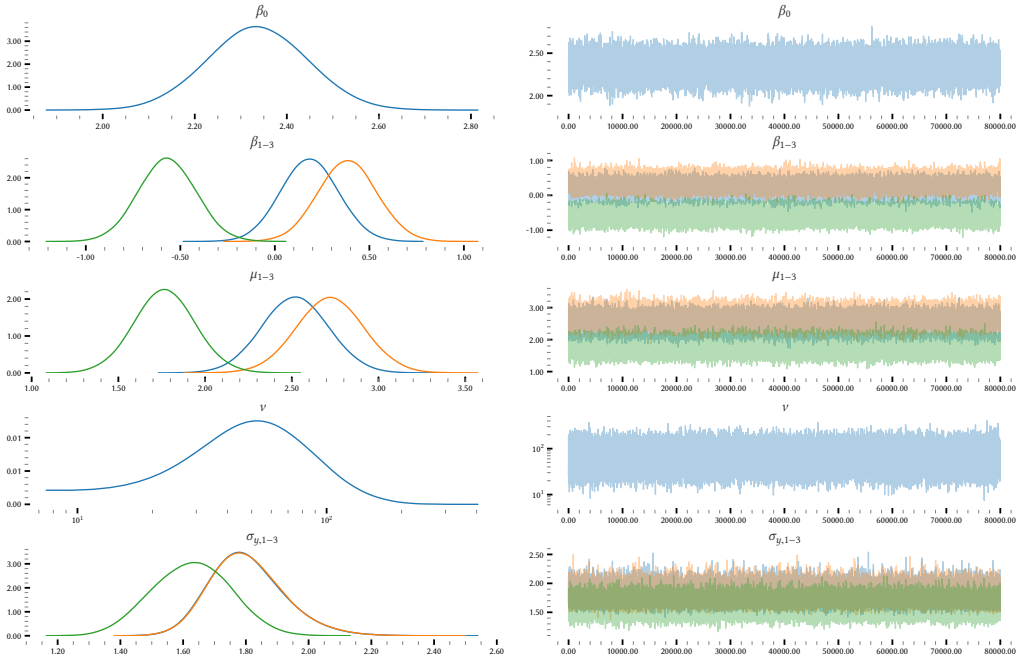


Fig. 6. Traceplot showing the results of the MCMC estimation. Left column shows the posterior distributions for $\beta_0, \beta_{[1-3]}, \mu_{[1-3]}, \nu, \sigma_{[1-3]}$, while the right column shows the corresponding traces. The posterior plots for ν, the parameter corresponding to degrees of freedom has a modal value of $\nu = 53$. While the ideal case of $\nu = \infty$ makes the StudentT distribution equivalent to the Normal distribution, in practice $\nu \geq 30$ is used to test for Normality. Thus the response distributions $y_{i|j}$ tend towards Normality. Notice that the mean of σ_2 (green curve, fourth panel, left column), variance of the group response in the case of text, is less than σ_1 or σ_3 (variances of the group responses for the comic, comic+social norm cases respectively), justifying our modeling assumptions of unequal variances across groups; the variances for the two comic conditions are nearly identical. The Gelman-Rubin statistic $\hat{R}$ was around 1, indicating that the different sampling chains converged. Furthermore, the effective sample size of all parameters was greater than 10,000.

Our analysis shows that the abstract comic form has a clear treatment effect over the corresponding text message. Figure 7 shows the contrast and the effect size of the contract for four cases. The first three columns of Figure 7 show the contrasts (top row) and the corresponding effect size (bottom row) between a pair of experimental conditions. The fourth column is a comparison between the text condition and the comic condition, by pooling together responses of the two comic conditions.

Let us examine one result ('text' v. 'comic') in detail; the analysis for the other cases follows a similar logic. The top left sub-figure shows that the contrast has a mode of 0.75 with the High

Posterior Density (HPD[4]) interval of [0.26, 1.27]. That is, subjects most frequently donate $0.75 more on in the comic condition than in the text condition, with 95% of the *increase in donations* lying between [$0.26, $1.27]. Since the HPD lies outside a meaningful ROPE (Region of Practical Equivalence)[5] of 0 ± 0.1, the result implies that there is a clear effect due to the treatment. Furthermore, the lower left sub-figure shows a moderate to medium effect size of 0.44; we consider an effect of 0.2 to be small, and an effect of 0.5 is a medium-sized effect. Since the distribution excludes the ROPE interval $[0 \pm 0.1]$, half of the small-sized effect value of 0.2, the discovered effect size is meaningful.

The second column contrasts the comic with social proof case ('comic+social') with the plain text ('text'). We find that the contrast has a mode of 0.95 with the High Posterior Density (HPD) interval of [0.44, 1.47]. The modal effect size is 0.55, with an HPD interval of [0.25, 0.86] which excludes a ROPE of 0 ± 0.1 indicating a meaningful, slightly larger than a medium-sized effect.

The third column contrasts the donations in the comic ('comic') condition against the comic with the social proof ('comic+social'). While the contrast is positive with a mode of 0.19, notice that the HPD interval $[-0.31, 0.72]$ overlaps a ROPE of 0 ± 0.1 implying that the observed differences are not meaningful. The corresponding modal effect size of this contrast is 0.11 with an HPD of $[-0.17, 0.39]$ with about 21.9% of the HPD lying to the left of 0, implying that the observed effect size is not meaningful.

The fourth column of Figure 7 compares the text condition ('text') with the comic condition by pooling the responses in the two comic conditions. The pooling of the 'comic' and the 'comic+social' data is straightforward in Bayesian analysis. Since the MCMC *jointly* estimates the posteriors of all variables, we can create a new variable that averages the values of the posteriors of the 'comic' and 'comic+social' cases for every step of the MCMC, to represent an averaged 'comic' condition. We find that the contrast has a mode of 0.86 with the High Posterior Density (HPD) interval of [0.42, 1.30]. Notice that the HPD interval is narrower than the corresponding HPD intervals comparing ('text' v. 'comic') and ('text' v. 'comic+social') The modal effect size is 0.59, with an HPD interval of [0.28, 0.89] which excludes a ROPE of 0 ± 0.1 indicating a meaningful, a medium to large sized effect, where the large effect size is 0.8.

In this section, we presented a Bayesian model to analyze the overall effect of using an abstract-comic to persuade people making charitable donation decisions, and understand the effect of adopting persuasive techniques in the abstract comic form. The results show that the use of abstract-comic produces a meaningful, medium to large modal effect (0.59) in persuading participants to donate. Although abstract-comic with social proof ('comic+social') can produce a larger effect that the case without social proof ('comic'), the contrast between abstract-comic and abstract-comic with social proof is not meaningful. The code and data will be released after publication. Next, we discuss the findings, design implications and limitations.

[4]The HPD interval is the location of 95% of the posterior density. This is similar to, but different from the idea of the confidence interval used in non-Bayesian Statistics. In non-Bayesian Statistics, a confidence interval of say 95% is informally interpreted as "with 95% probability the parameter of interest lies in a specific interval; the tails are of equal width (i.e. 2.5%)"; the HPD on the other hand, is the *densest* interval covering 95% of the posterior. The HPD is guaranteed to include the most likely value, but this is not always true for confidence intervals; see McElreath [40, p. 57] for a simple example. For a more careful definition of the confidence interval, see Hoekstra et al. [25], which also includes a discussion on the misinterpretation of confidence intervals.

[5]Unlike non-Bayesian Statistics, where one can ask for example, if the two means for two treatments are different $P(\mu_1 \neq \mu_2)$, in Bayesian statistics, one asks if the HPD interval of the distribution $P(\mu_1 - \mu_2)$, that is, the distribution of the difference of the means of the two treatments, excludes an interval where we can consider the two treatments equivalent. This equivalence interval is domain dependent.

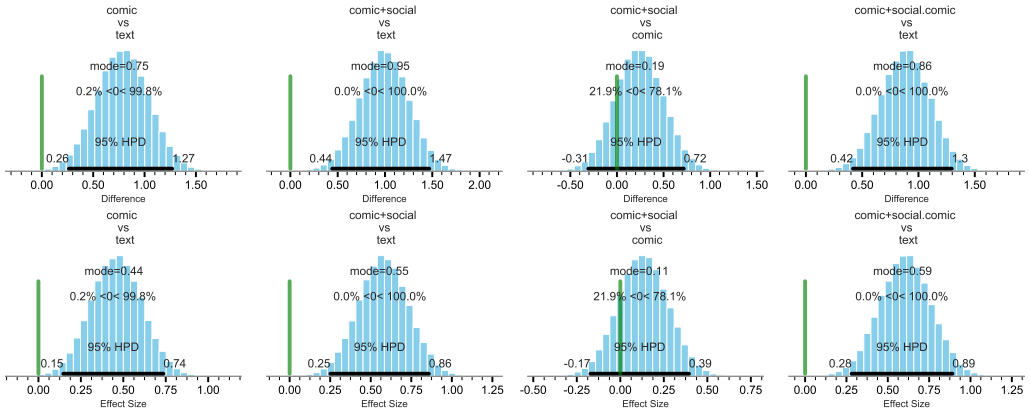


Fig. 7. The figure shows contrasts between the three experimental conditions ('text', 'comic', 'comic+social'). We also show in the fourth column contrasts between the text condition and pooled case of the two comic conditions to highlight the effect of using the abstract comic. Since we are highlighting contrasts, each sub-figure also shows a green line located at 0.The main finding: the comic treatment creates a medium to large sized effect (mode=0.59, pooled case, fourth column, bottom) when compared to the 'text' condition. There is a slight increase in effect size when using the comic with social proof ('comic+social', second column, bottom) in comparison to the case when there is a comic without the social proof ('comic', first column, bottom). However, these differences are not meaningful, since the about 21.2% of the HPD of the contrast between the 'comic' and the 'comic+social' case lies to the left of 0.0 (second column).

5 MODEL CRITICISM

In this section, we examine alternatives to the proposed model, a straightforward task in Bayesian analysis. We begin with a short section explaining our choice to use Bayesian analysis. Then in 5.2, we examine how well the proposed models explain the observed data, including model convergence. Finally, in 5.3, we examine alternative model formulations.

5.1 Why Bayesian?

In a recent paper, Kay et al. [31], make a persuasive argument that Bayesian methods are better suited to the HCI community, including making the case that Bayesian methods allow for replicating the results and improving the strength of the conclusions by using previous outcomes as priors. We would add two reasons, in addition to those by Kay et al. [31] to explain our decision to use Bayesian models.

Transparency: With a Bayesian model, the researcher foregrounds all the aspects of the model; there are no modeling assumptions that need checking, not already foregrounded in the model description. Non-Bayesian statistics are powerful tools, and when used by an experienced statistician, they can dramatically reduce the process of inference. However, for researchers who wish to investigate their findings without access to a statistician, they need to be careful of the assumptions of the different tests: Normality (t-test); heteroscedasticity (e.g., ANOVA) and ensuring that the data satisfy the assumptions. Omitting the right sequence of analysis can lead to inferences not supported by the data.

Small n studies: A Bayesian model is valid at *every* value of n; we do not have to wait for $n \geq 30$ to satisfy assumptions of say Normality. For small n values, the result is of course affected by the choice of the prior; but by using weakly informative priors, we can ensure that

the prior doesn't dominate inference. Furthermore, when Bayesian models use maximum entropy likelihood functions (e.g., members of the exponential family, that include the Normal distribution and the gamma distribution), we make the *most conservative* inference given the data. See McElreath [40, Chapter 9] for an excellent description of the use of maximum entropy models in Bayesian analysis.

5.2 Posterior Checks, Convergence and Normality

In this section, we examine the model fits by examining how well the model makes predicts the observed data, model convergence and our assumptions about Normality.

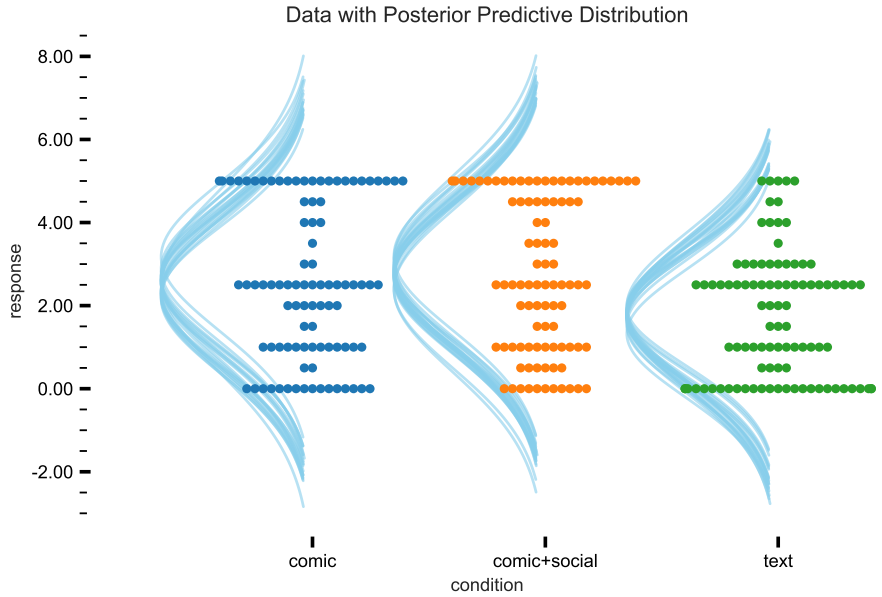


Fig. 8. Posterior prediction check: traces from the posterior distributions for each condition are super-imposed onto the observational data. The x-axis shows the three conditions, while the y-axis shows the contribution (each dot corresponds to a response by a single subject). The t-distribution captures the heavy tails, and we also see that the scale of the t-distribution in the case of the text condition is smaller than the other two cases.

One of the advantages of a Bayesian model is that one can use the model to make predictions. Since the probabilistic model is generative, we can simply draw observations from the model. When we superimpose the posterior distribution traces onto the observational data in Figure 8 we can observe three things. First, notice that the mean contribution for the 'text' condition is less than the 'comic' or the 'comic+social norm' conditions, which seems reasonable given the increased number of \$0.0 contributions for the 'text' condition in comparison to the other two cases. Second, the use of the heavy-tailed t-distribution captures well the tails of the distribution at \$0 and at \$5.0. Third, the scale of the distribution for the text condition is smaller than the scales for the two comic conditions, justifying our use of different scale parameters σ_j for the three experimental conditions.

The model shows good convergence, as evidenced by the traceplot in Figure 6. The Gelman-Rubin statistic $\hat{R}$ was around 1, indicating that the different sampling chains converged. Furthermore, the effective sample size of all parameters was greater than 10,000.

We used the t-distribution to model likelihood thinking that data may be heavy-tailed implying that a heavy-tailed distribution may be a better likelihood function than a Normal distribution. Let us examine the posterior distribution for ν, the degrees of freedom of the t-distribution. As a reminder, the t-distribution is equivalent to the Normal distribution when $\nu = \infty$. We can see that while the 95% HPD lies between [17.37, 147.19], less than 7% of the posterior lies below $\nu = 30$, the traditional rule-of-thumb in non-Bayesian statistics for use of the Normal distribution.

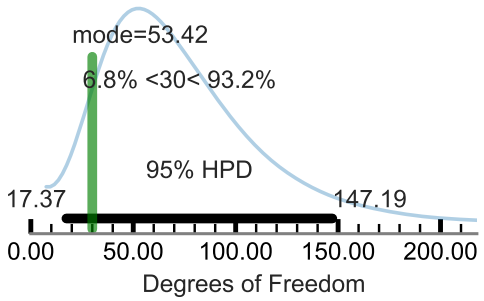


Fig. 9. The posterior distribution for ν, the normality parameter of the t-distribution. When $\nu = \infty$ the t-distribution is identical to the Normal distribution. The posterior distribution shows a vertical green bar for $\nu = 30$, the traditional cut-off condition on degrees of freedom, for using a Normal distribution. The mode $\nu = 53.42$, and the posterior probability distribution implies that only about 7% of the posterior lies to the left of the green line (i.e. $P(\nu \leq 30) \approx 0.07$), implying that the assuming that the likelihood function to be Normally distributed will give similar inference.

Having discussed posterior-prediction checks, model convergence, and normality, next, we discuss alternatives to the model.

5.3 Alternative Models

As a first instance, consider a similar model, except that the scale parameter of the likelihood function is *not nested* like our current model. Instead, we consider the equal variance case, where all the variances are equal (similar to ANOVA), and that the variance is drawn from a uniform distribution. In other words, $\sigma_j = \sigma \sim U(L, H)$, where $L > 0$ and where H is a large constant. The main effect of the equal variance assumption is that there is no information sharing among the groups as would be the case when each σ_j is drawn from the same distribution, whose parameters have hyper-priors; the latter is our current model.

The main effect of constraining our simplified model is that we are slightly poorer in predicting the observed data since all the variances are guaranteed by the model to be equal, whereas we can see from Figure 6 that the mean scale (or equivalently variance) in the text condition is lower. We are skipping the traceplot and the contrast plots in this case, as they are similar, to Figure 6 and Figure 7, except that the effect size for the combined case is slightly lower due to the equal variance assumption.

Instead, we compare the two models using WAIC (Widely Applicable Information Criterion), a principled way to compare models when they have identical likelihood functions [23]. WAIC uses

the predictive loss to compare two models with different parameters. First, WAIC computes the average log likelihood of each training data point (over the posterior distribution) less the variance of the log likelihood for the same data point; and then it computes the sum over all data points. That is, WAIC for a model:

$$\text{WAIC} = -2\left(\sum_{i}^{N} \log \Pr(y_i) - V(y_i)\right)$$

where, N is the total number of training points, y_i is the observation, $\Pr(y_i)$ is the averaged data likelihood over the posterior, and where $V(y_i)$ is the variance of the data likelihood over the posterior. When we compare the two models, one with unequal variances, and one with equal variances. We show our results in Table 1:

Table 1. WAIC comparison between the model with equal variances against the case when the variances are not constrained to be equal (i.e. we use a hierarchical model). Both cases assume the Student-t likelihood function, the function used in this paper. The columns show respectively, WAIC, pWAIC (the effective number of parameters; also: pWAIC = $\sum_i V(y_i)$), dWAIC (the difference between the WAIC scores of the other models with the best model), weight (the relative probability that the model explains the data) SE, the standard error of the WAIC estimate, dSE is the standard error of the difference of the current model against the top model. The table shows that the hierarchical model with unequal variances better explains the observations.

Model	WAIC	pWAIC	dWAIC	weight	SE	dSE
Unconstrained variances, hierarchical	1102.48	3.95	0.00	1.00	14.61	0.00
Constrained, equal variances	1104.17	3.42	1.69	0.00	14.25	1.54

The results in Table 1 say that model with the unconstrained variances is better at explaining the data than the model with constrained variances; the relative probability that the hierarchical model with unconstrained variances better explains the observation is 1.0 (refer to the weight column in Table 1)

How useful was the choice of the Student-t drawing distribution Equation (1), instead of assuming that the outcomes are drawn from a Normal distribution? Our analysis of the posterior distribution of the degrees of freedom parameter ν shows that there is only a small probability ($P(\nu \leq 30) \approx 0.07$) that $\nu \leq 30$. Thus, we may use the Normal likelihood function, without meaningfully affecting the conclusions. Indeed, in our experiments, when we do model the observations with a Normal likelihood, assume equal variances, we find no meaningful differences in the contrasts or the effect sizes, with the Student-t model with equal variances (results omitted due to space constraints).

Since the charitable donations are bounded to lie between \$0.0 and \$5.0, might we benefit from using bounded likelihood functions like the Beta distribution $Y \sim Beta(\alpha_j, \beta_j)$ to represent the charitable donations Y_j under the different conditions j? While a $Beta(\alpha, \beta)$ distribution lies between $[0, 1]$, we can scale down the contributions to lie in $[0, 1]$ to use with the $Beta(\alpha, \beta)$ distribution. But notice from Figure 4 that in each experimental condition, there is a central lobe, and heavy tails at each extreme, notably at \$0.0 and at \$5.0.

Our view of models motivated by McElreath [40] is that they represent an *epistemological* claim, not an *ontological* claim (i.e. a physical assumption about the world). Since our goal is to understand the average tendency to give to charity under the different conditions, and not to make predictions (as might be the case if we were trying to model donation with age as a predictor), the fact that the t-distribution is not bounded is less relevant here.

To summarize, we discussed transparency and utility in small-n studies as motivation for our use of Bayesian modeling. Our posterior predictive check shows that the t-distribution models well the heavy tails, and shows that the variances for the comic conditions are different from the text condition. The model shows good convergence with the Gelman-Rubin statistic $\hat{R}$ was virtually identical to 1.0 for all parameters; the effective sample size was greater than 10,000 for all parameters. The modal value of ν, the degrees of freedom parameter was around 53, suggesting that we could also use the Normal likelihood function. Comparison between the hierarchical model with unconstrained variances and the model with equal variances shows that the hierarchical model is better at explaining the observations.

6 DISCUSSION

We first summarize our findings. Then, we will propose a framework for algorithmically synthesizing persuasive messages into the abstract comic form, discuss design implications, and identify limitations.

From asking individuals to act to appealing for charitable donations, text messages have been widely used in simulating behaviors. Scholars from psychology have shown how variations in the construction of text messages alter decisions. We explored and examined the role of the abstract comic form, a highly expressive, affective medium in communicating persuasive messages. To test the effectiveness of abstract comic persuasive messages, we persuaded individuals to make online charitable donation decisions, a common public good dilemma. In the dilemma, due to the non-exclusive and non-rivalrous nature of public goods, persuading individuals to contribute is hard and crucial. Also, online charitable donation task not only avoids confounding factors such as habit formation which exists in other persuasion tasks such as exercise and healthy diet but also assures the ecological validity of our study as charity and organizations often solicit donations online. In our study, we compared the persuasive power of text messages, comic messages, and comic messages with social proof in asking charitable donations to public health research (e.g., the Organization for Autism Research).

Our results show that study participants prefer persuasive messages in comic form over plain text in collective action decisions with the real cost. When making the charitable donation between \$0 to \$5, study participants donated \$0.86 more if they read the persuasive message in an abstract comic form. The results demonstrate the persuasive power of abstract comic in stimulating behaviors in pro-social decisions. Our findings are consistent with prior research on visual stimulus in persuasion and the benefits of comics in communication. One potential explanation is that study participants were more attracted by the comic strip and projected themselves onto the character. When the projection happens, the persuadee may be able to digest the information better which stimulated them to donate more to the Organization for Autism Research. However, when comparing between the comic condition and comic with the social proof condition, although study participants who read the comic with 'social proof' donated more, our analysis result did not imply meaningful effect size. Although the use of social proof showed a significant impact on other forms of persuasive messages (e.g., pure textual), in the abstract comic persuasive messages, the effect is not substantial. One of the potential explanation is that the design of our comic strip did not signify the idea of social proof other than in the text bubble. It is worth to explore other comic designs, e.g., adding other donators as comic characters, that can make the social proof more salient.

Another natural question is whether the maximum donation amount of \$5 in our study has any influence on our findings (e.g., will people make the same donation decision if they can donate more?). Prior studies in behavioral economics suggest that small-stake experiment in developed countries can be replicated in developing countries where the stake becomes a significant portion

of study participant's weekly income [8, 9, 30]. Additionally, Post et al. [50] showed that the results from small-stake experiments would hold with higher stakes in developed countries as well. Those results seem to suggest that our findings will hold with higher donation amount. However, the best way to confirm is through actual experiments.

Although in our study, the persuasive goal is making charitable donation decisions, we believe our findings suggest future research on longitudinal behavior change (e.g., health) via the abstract comic.

6.1 Framework for Algorithmically Synthesized Abstract-Comic Persuasive Messages

Our study showed the persuasive power of abstract comics in encouraging people to make pro-social decisions. However, one drawback of using visual stimulus in persuasion is cost during the creation process. Although compared to creating other persuasive visual stimulus such as videos or complex graphical illustration, the simplicity of abstract requires less effort, the creation process is not easy. In this section, we propose a framework that allows full/semi- automatic generation of abstract persuasive messages and identify crucial features that need to be addressed in future work. We believe such a tool will lower the barrier for the persuader to take advantage of the abstract comic (as demonstrated in our study) in encouraging individuals to act in public good dilemmas.

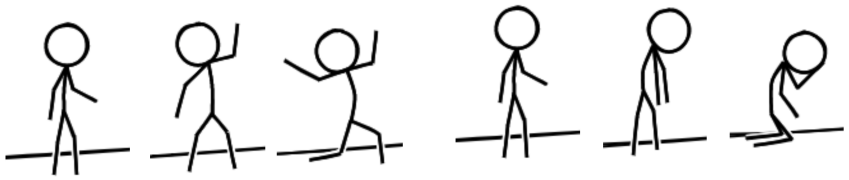
In our study, we created a comic generator with existing packages including "cmx.io" [26] and "rough.js" [51] to generate the three-panel abstract comic strip. With several pre-defined character gestures (see Figure 10), the generator only requests the text input from the persuader to create the comic message.

We believe the generator we built can be further developed as a framework for algorithmic synthesis. Now, we identify crucial features that need to be addressed in future research. First, the framework should be able to automatically select the character gesture that best fit the persuasive message's context. For example, when the message receiver was told good news, his/her gesture should that conflict. The appropriate mapping will create a natural and coherent comic message which is the key for an expressive message. We need future research to build such a model. Second, the framework should be able to use other abstract comic elements, such as inter-character distance and shading, to create persuasive comics. To achieve this feature, we need to understand how different elements affect the persuasiveness of the abstract comic messages. Third, the framework should be able to individualize the persuasive comics by using persuadee's behavioral data to inform comic contents. We could derive the statistics for the social norm from behaviors of friends and the data from a person's own activity for information framing, and map the data to the character gesture, thus reinforcing the framing.

6.2 Design Implications

Our main design implications are two folded. First, when encouraging individuals to act in collective actions or contribute to public goods, non-profit organizations and governmental agencies ought to consider to use abstract comic in their online messaging strategies to persuade. From our results, using abstract comics can, in particular, persuade people to act in public-goods dilemmas that require single-shot decisions (online charitable donations; taking the flu shot). Additionally, when constructing persuasive comics, it is worthwhile to consider to incorporate social norms.

Second, our results also inform the development of a framework that can algorithmically synthesis textual persuasive messages into abstract comic form and uses data-driven methods to personalize the message to further increase messages' persuasive power. With such a framework, agencies can create their abstract comic persuasive messages easily and incorporate them to alleviate real-world public goods dilemmas.



(a) Gestures for positive framed messages (b) Gesture for negative framed messages

Fig. 10. Different character gestures to communicate various levels of emotional intensity. The left figure shows gestures from neutral to the happiest. The right figure shows gestures from neutral to the most frustrated.

6.3 Limitations & Future Work

Distant and Non-exclusive Task: Although our study asked participants to make the decision with a real cost, the decision domain is limited to only one scenario, online charitable giving. In this task, the participant's reward is distant and non-exclusive (e.g., participant's donation decision won't bring any immediate reward and won't exclude the participant from the research outcome). Although both characteristics help us clearly test the persuasive benefits of abstract comics, they also limited our generalizability. We need future research to understand the pros and cons of using abstract comic messages in persuasive tasks with different reward characteristics. For example, for persuasive goals in the quantified self movement [20, 13] such as exercise and dieting, the reward is distant (e.g., people's health won't be improved immediately after exercise) but exclusive (e.g., healthy life situation mainly benefit the individual him/herself). Moreover, our persuasion task is for an online charitable donation. Future work is needed to extend our result to offline charitable donation solicitation where persuasive text messages, such as mailers, also often used.

Comic Message Construction: In our study, we constructed the comic messages in the XKCD comic style. Although the XKCD style is easily recognizable, there are many different ways to create abstract comics. Comic, as a creative art form, has rich semantics and vocabularies to communicate ideas [38]. Out of necessity, we limited the complexity our comics strip (e.g., number of characters, gestures, and the number of panels). Future research is needed to explore and understand how other elements in the comic, for example, inter-character distance and character's gesture, affect the persuasiveness of the comic. Furthermore, while simplicity grants us the possibility to automate the generation process, future technology may allow us to generate more complicated persuasive abstract comics.

7 CONCLUSION

Inspired by a rich history in persuasive test messages construction and benefits of abstract comic in communication, this paper examined if the abstract comic form was more persuasive than the corresponding plain text in the context of encouraging participants to donate for charitable causes. We conducted a field study on Amazon Mechanical Turk with 307 participants. In the study, participants received one of the three persuasive messages designed to ask for a donation to the Organization of Autism Research, a persuasive text message, a three-panel comic strip, and a three-panel comic strip incorporating the idea of social proof, and made donation decisions with real money. We analyzed the results using a hierarchical Bayesian framework that allows for

understanding effect sizes, as well are transparent and helpful in small-n studies. The results shows convincingly that the three-panel abstract comic is more persuasive than the text (a medium to large effect size = 0.59). We also show that while the comic with social norm increases the donation level over the comic without the norm, the effect size is very small (0.11) and the increase is not meaningful. To summarize, the comic form meaningfully increases donations over the plain text, but the presence of the norm is not effective. We caution that the result holds for single-shot, public goods tasks; the value of the social norm in the comic, for exclusive tasks with distant rewards such as exercise, or dieting needs future research. The main implication of our work is that non-profits and governmental agencies ought to consider using abstract comic in their online campaign as they work to alleviate public goods dilemmas. Consider the low conversion (0.3 %) rate reported in Wikipedia's online fundraising campaign [22], even small increase in the conversion rate will make a huge difference. We believe that these agencies can easily include the use of the comic form as part of their overall messaging strategy because the simplicity of the abstract comic form allows it to be easily synthesized and to additionally incorporate social norms.

As next steps, we plan to develop an algorithmic framework that automatically synthesizes text persuasive messages and maps a person's behavioral data (e.g. amount walked this week) to a three-panel persuasive comic. We also plan to conduct longitudinal field experiments with an emphasis on storytelling where individuals receive three-panel comics over time, and comics are connected with a storyline.

REFERENCES

[1] Naveen Amblee and Tung Bui. 2011. Harnessing the influence of social proof in online shopping: the effect of electronic word of mouth on sales of digital microproducts. *International Journal of Electronic Commerce*, 16, 2, 91–114.

[2] James Andreoni. 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *The economic journal*, 100, 401, 464–477.

[3] American Psychiatric Association et al. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

[4] Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. 2016. Telling stories about dynamic networks with graph comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, San Jose, California, USA, 3670–3682.

[5] Gary S Becker. 1974. A theory of social interactions. *Journal of political economy*, 82, 6, 1063–1093.

[6] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods*, 43, 3, 800.

[7] 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, 351–368.

[8] Hans P Binswanger. 1980. Attitudes toward risk: experimental measurement in rural india. *American journal of agricultural economics*, 62, 3, 395–407.

[9] Hans P Binswanger et al. 1981. Attitudes toward risk: theoretical implications of an experiment in rural india. *Economic Journal*, 91, 364, 867–890.

[10] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6, 1, 3–5.

[11] John J Burnett. 1981. Psychographic and demographic characteristics of blood donors. *Journal of Consumer Research*, 8, 1, 62–66.

[12] Stephen Cary. 2004. *Going graphic: Comics at work in the multilingual classroom*. Heinemann Portsmouth, NH.

[13] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (CHI '14). ACM, Toronto, Ontario, Canada, 1143–1152.

[14] Robert B Cialdini. 1993. Influence: the psychology of persuasion. *New York, Morrow*.

[15] Robert B. Cialdini and Noah J. Goldstein. 2004. Social influence: compliance and conformity. *Annual Review of Psychology*, 55, 1, 591–621. PMID: 14744228.

[16] Sunny Consolvo et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1797–1806.

[17] S Cotterill, P John, and L Richardson. 2010. The impact of a pledge campaign and the promise of publicity on charitable giving: a randomised controlled trial of a book donation campaign. In *Randomised Controlled Trials in the Social Sciences Conference, York, September 2010*.

[18] Mette Trier Damgaard and Christina Gravert. 2017. Now or never! the effect of deadlines on charitable giving: evidence from two natural field experiments. *Journal of Behavioral and Experimental Economics*, 66, 78–87.

[19] Sanorita Dey, Brittany Duff, Karrie Karahalios, and Wai-Tat Fu. 2017. The art and science of persuasion: not all crowdfunding campaign videos are the same. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 755–769.

[20] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (DIS '14). ACM, Vancouver, BC, Canada, 667–676.

[21] Daniel C Feiler, Leigh P Tost, and Adam M Grant. 2012. Mixed reasons, missed givings: the costs of blending egoistic and altruistic reasons in donation requests. *Journal of Experimental Social Psychology*, 48, 6, 1322–1328.

[22] [n. d.] Fundraising 2010/banner testing. ().

[23] Andrew Gelman, Jessica Hwang, and Aki Vehtari. 2014. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24, 6, (Nov. 2014), 997–1016.

[24] Noah J Goldstein, Robert B Cialdini, and Vladas Griskevicius. 2008. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, 35, 3, 472–482.

[25] Rink Hoekstra, Richard D Morey, Jeffrey N Rouder, and Eric-Jan Wagenmakers. 2014. Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review*, 21, 5, 1157–1164.

[26] Comix I/O. [n. d.] Comix i/o. ().

[27] R Mark Isaac, Kenneth F McCue, and Charles R Plott. 1982. Public goods provision in an experimental environment.

[28] Aarti Iyer and Julian Oldmeadow. 2006. Picture this: emotional and political responses to photographs of the kenneth bigley kidnapping. *European Journal of Social Psychology*, 36, 5, 635–647.

[29] Marco A. Janssen, Allen Lee, and Hari Sundaram. 2016. Stimulating contributions to public goods through information feedback: some experimental results. *PLoS ONE*, 11, 7, (July 2016), 1–16.

[30] Steven J Kachelmeier and Mohamed Shehata. 1992. Examining risk preferences under high monetary incentives: experimental evidence from the people's republic of china. *The American Economic Review*, 1120–1141.

[31] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-centered design of statistics: why bayesian statistics better fit the culture and incentives of hci. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, Santa Clara, California, USA, 4521–4532.

[32] John Kruschke. 2014. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

[33] Richard N Landers and Tara S Behrend. 2015. An inconvenient truth: arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial and Organizational Psychology*, 8, 2, 142–164.

[34] Yu-Hao Lee and Gary Hsieh. 2013. Does slacktivism hurt activism?: the effects of moral balancing and consistency in online activism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 811–820.

[35] Jennifer Mankoff et al. 2010. Stepgreen. org: increasing energy saving behaviors via social networks. In *Fourth International AAAI Conference on Weblogs and Social Media*.

[36] Gerald Marwell and Ruth E Ames. 1981. Economists free ride, does anyone else?: experiments on the provision of public goods, iv. *Journal of public economics*, 15, 3, 295–310.

[37] Mizuki Matsubara, Olivier Augereau, Charles Lima Sanches, and Koichi Kise. 2016. Emotional arousal estimation while reading comics based on physiological signal analysis. In *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding*. ACM, 7.

[38] Scott McCloud. 1993. *Understanding comics*. William Morrow Paperbacks.

[39] Jason E. McDermott, Matthew Partridge, and Yana Bromberg. 2018. Ten simple rules for drawing scientific comics. *PLoS Computational Biology*, 14, 1.

[40] Richard McElreath. 2015. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

[41] Doug McKenzie-Mohr. 2011. *Fostering sustainable behavior: An introduction to community-based social marketing*. New society publishers.

[42] Cees Midden, Teddy McCalley, Jaap Ham, and Ruud Zaalberg. 2008. Using persuasive technology to encourage sustainable behavior. *Sustainability WS Pervasive*, 113.

[43] Edward Miguel and Mary Kay Gugerty. 2005. Ethnic diversity, social sanctions, and public goods in kenya. *Journal of public Economics*, 89, 11-12, 2325–2368.

[44] Randall Munroe. 2009. Xkcd: a webcomic of romance, sarcasm, math, and language. *Creative Commons Attribution-Noncommercial*, 2.

[45] Richard E Nisbett and Lee Ross. 1980. Human inference: strategies and shortcomings of social judgment.

[46] Mancur Olson. 2009. *The logic of collective action*. Vol. 124. Harvard University Press.

[47] Gabriele Paolacci and Jesse Chandler. 2014. Inside the turk: understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23, 3, 184–188.

[48] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk.

[49] Edgar A Pessemier, Albert C Bemmaor, and Dominique M Hanssens. 1977. Willingness to supply human body parts: some empirical results. *Journal of Consumer Research*, 4, 3, 131–140.

[50] Thierry Post, Martijn J Van den Assem, Guido Baltussen, and Richard H Thaler. 2008. Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review*, 98, 1, 38–71.

[51] Rough.js. [n. d.] Rough.js. ().

[52] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. 2016. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2, (Apr. 2016), e55.

[53] P Wesley Schultz, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18, 5, 429–434.

[54] Ted Selker, Shih-Yuan Yu, Che-Wei Liang, and Jane Hsu. 2015. Sweetbuildinggreeter: a demonstration of persuasive technology for public space. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 475–486.

[55] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, 211, 4481, 453–458.

[56] Marcus B Weaver-Hightower. 2017. Losing thomas & ella: a father's story (a research comic). *Journal of Medical Humanities*, 38, 3, 215–230.

[57] Leah Zhang-Kennedy, Robert Biddle, and Sonia Chiasson. 2017. Secure comics: an interactive comic series for improving cyber security and privacy. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference* (HCI '17) Article 65. BCS Learning & Development Ltd., Sunderland, UK, 65:1–65:3.

[58] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. 2014. Stop clicking on "update later": persuading users they need up-to-date antivirus protection. In *International Conference on Persuasive Technology*. Springer, 302–322.